Build an Extensible Data Warehouse Foundation

Establish a well-architected core model with just enough oversight and governance.
Table of contents

1. Build an Extensible Data Warehouse Foundation
2. Executive Brief
3. Execute the Project/DIY Guide

Phase 1. Prepare for the Data Warehouse Foundation Project
   1.1 Structure the Data Warehouse Project
   1.2 Review Data Management Impacts

Phase 2: Establish a Data Warehousing Strategy and Business Drivers
   2.1. Identify Essential Business Processes to Support
   2.2. Create the Core Data Model
   2.3. Create an Architectural Strategy
   2.4. Explore Technology Options

Phase 3. Plan for Data Warehouse Governance
   3.1. Plan for the Data Warehouse Center of Excellence
   3.2. Define Operating Procedures

3. Summary/Conclusion
4. Appendices
Build an Extensible Data Warehouse Foundation

Establish a well-architected core model with just enough oversight and governance.
Don’t just build a data warehouse, build it smartly.

Data warehouse projects are some of the most challenging projects for IT because of their complexity. That being said, the data warehouse is still a technology that can help you gain a significant competitive advantage.

Start by architecting a data warehouse correctly and build it one subject area at a time.

Similar to building a house, the key to success is creating manageable pieces and integrating those pieces based on a well-architected plan and foundation with just enough oversight and governance. With this approach, you are able to develop a program around the data warehouse that will be able to empower your business over time.

Daniel Ko,
Manager, Information Management
Info-Tech Research Group
Our understanding of the problem

This Research Is Designed For:

✓ CIOs who want to build a data warehouse (DW) that helps business users make informed decisions.
✓ Business executives who want to work with IT to leverage a data warehouse to improve business processes and access to correct data.

This Research Will Help You:

✓ Establish an overall vision and identify key business processes for data warehousing.
✓ Create the minimal level of data warehouse governance to maintain your data warehouse.
✓ Understand business needs to pick the most suitable technological option for the data warehouse.

This Research Will Also Assist:

✓ Project managers and business analysts in accelerating a data warehousing project by following a best-practice data warehousing approach that leverages emerging data modeling patterns and data technologies.
✓ Architects who want to make the most logical data warehouse decisions.

This Research Will Help Them:

✓ Identify which architectural components suit your needs.
✓ Populate source data into the data warehouse using the Rosetta Stone approach.
✓ Detail the essential data elements that will feed tables in the core of the data warehouse.
Executive summary

**Situation**

- Organizations are fully aware that data is a critical asset and an indispensable element in successful decision making.
- The data and information landscapes have drastically expanded and diversified; combined with the increase in the speed of business, decision makers are in need of a system that provides the right information at the right time.

**Complication**

- Data warehouse implementation is a costly and complex undertaking, and can end up not serving the business needs appropriately.
- Too heavy a focus on technology creates a data warehouse that isn’t sustainable and ends up with poor adoption.
- Emerging data sources and technologies add complexity to how the appropriate data is made available to business users.

**Resolution**

- Leverage an approach that focuses on constructing a data warehouse foundation that is able to address a combination of operational, tactical, and ad hoc business needs.
- Invest time and effort to put together pre-project governance to inform and provide guidance to your data warehouse implementation with minimal ongoing operations requirements to keep the data warehouse healthy.
- Develop the “Rosetta Stone” views of your data assets to facilitate data modeling and the order of building your data warehouse environment.
- Select the most suitable architecture pattern to ensure the data warehouse is “built right” at the very beginning.

**Info-Tech Insight**

1. A data warehouse is a project; but successful data warehousing is a program. An effective data warehouse requires planning beyond the initial technology implementation.

2. Governance, not technology needs to be the core support system for enabling a data warehouse program.

3. Understand business processes at the operational, tactical, and ad hoc levels to ensure a fit-for-purpose data warehouse is built.
The speed of business is showing no signs of slowing down.

Globalization of commerce and the shift towards creating customer experiences are significant drivers in the quickening of business users’ information needs.

66% of business users are unsatisfied with the speed of information delivery.

57% don’t feel they have access to data relevant to decision making.

55% of business users reported a “shortening of their decision-making window.”

Source: Aberdeen Group, 2014
Enterprise-wide views of data drive more holistic insights

By creating a data warehouse environment layer that focuses on integrating real-time and historical data from multiple data sources and can be dedicated to reporting, your organization is in a better position for information sharing and enterprise-wide collaborative decision making that is efficient, effective, and timely.

Application-level reporting is ideal for short-term, operational decision making, and time-based reporting for in-application data. However, this same data becomes difficult to access and costly to leverage when trying to combine with other datasets to make enterprise strategic decisions if there is no data warehouse that has already defined the true sources of data for just such purposes.

Leading Barriers to Analytics Use

<table>
<thead>
<tr>
<th>Issue</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>No centralized approach to capturing and analyzing data for our company’s use</td>
<td>32%</td>
</tr>
<tr>
<td>Our company lacks proper technology and infrastructure to capture the data</td>
<td>23%</td>
</tr>
<tr>
<td>Leadership does not understand or support the use of analytics</td>
<td>12%</td>
</tr>
<tr>
<td>We lack the correct talent to use our data</td>
<td>10%</td>
</tr>
<tr>
<td>Our industry hasn’t really employed analytics in a significant way</td>
<td>8%</td>
</tr>
<tr>
<td>Privacy issues</td>
<td>3%</td>
</tr>
<tr>
<td>Other</td>
<td>12%</td>
</tr>
</tbody>
</table>

Source: Deloitte, 2013

Analytics Is a Key Factor in Decision Making

- Better decision making based on data: 49%
- Better enablement of key strategic initiatives: 16%
- Better relationships with customers and business partners: 10%
- Better sense of our risk and better ability to react to changes in the economic environment: 9%
- Better financial performance of the organization: 9%
- Better capability to respond to buying trends in the marketplace: 5%
- Identification and creation of new product and service revenue streams: 1%

Source: Deloitte, 2013
An integrated, centralized decision-support database and the related software programs used to collect, cleanse, transform, and store data from a variety of operational sources to support business intelligence. A data warehouse may also include dependent data marts (DAMA-DMBOK Guide, 2009).

Business intelligence is becoming a table stakes activity; all organizations need hard evidence to support decision making.

Data warehouses (and associated architecture) bridge the gap between mass accumulation of data in source systems and comprehensive views of information, en route to creating shared insight and institutional knowledge at the top of the data value chain.
Industry Spotlight: Enterprise data warehouses are driving new value in the US healthcare sector

As healthcare organizations today are motivated more than ever to provide better patient outcomes per dollar spent, the use of business intelligence tools to facilitate data-driven decision making is imperative. Data comes from multiple sources, such as electronic health records (EHRs), financial systems, patient satisfaction surveys, and external laboratory testing databases – all may be used to improve services delivered.

At Texas Children’s Hospital, reports sourced from the enterprise data warehouse cost 70% less to build than reports created directly from EHRs; one well-planned and flexible data visualization was able to replace ten static reports.

Using the enterprise data warehouse as a single source of truth, North Memorial Healthcare was able to coordinate multiple groups of healthcare professionals to reduce the number of pre-39 week deliveries by 75%.

Key Benefits of the Healthcare Data Warehouse:

- Enabling a More Efficient, Scalable Reporting Process
- Ensuring Consistent Data That Everyone Can Trust
- Enabling Meaningful, Targeted Quality Improvement

Source: Health Catalyst, 2014
Industry Spotlight: Streamlining data availability to improve reporting in government finance

Tanzanian Revenue Authority (TRA) is responsible for several financial functions of the government, including the collection of tax revenues and conducting research to advise changes to tax policies. The Research, Policy and Planning department of the TRA ran into significant challenges with reporting and analysis as a result of having no centralized point of access for data residing in multiple siloes.

TRA established an enterprise data warehouse that served as a hub to link internal and external data sources across departments. In addition to the storage infrastructure, the enterprise DW was designed for easy linkage to business intelligence tools and incorporated strict data validation and security procedures.

**DW Adoption Benefits:**

1. Automated Data Integration
2. Automation of Business Process
3. Third-Party System Integration
4. Centralized Database Management System

Source: TechnoBrain, 2015
Data warehousing solutions are viewed as essential, but major challenges remain

A data warehouse is a powerful enabler in converting data to a usable asset, but can quickly turn into a nightmare when issues beyond the technology arise.

Successful data warehousing is reliant on several data management enablers in Info-Tech’s data management framework, namely data governance and data quality management.

- Dealing with data quality and consistency issues: 58%
- Integrating data from disparate sources: 53%
- Getting funding/executive support for new projects: 33%
- Managing multiple data warehouses or data marts: 28%
- Integrating DWs with “big data” systems/unstructured data stores: 27%
- Managing growing volumes of data in data warehouse: 25%
- Evaluating and selecting data warehouse technology: 25%
- Maintaining system performance and availability: 21%
- Implementing additional data warehouses: 15%
- Scaling systems to support more users/queries: 14%

Source: Tech Target, 2013
Big data technology might be revolutionizing data warehousing, but the core isn’t going anywhere

With the excitement around big data and associated technologies that allow for ingestion and processing of larger volumes and varieties of data, some have been quick to say that the current state of data warehousing will be a thing of the past.

However,

96% of organizations do not see big data taking the place of the existing data warehouse infrastructure.

What technologies does your organization plan to use to support its big data environment?

Mainstream relational databases/data warehouses 55%
Analytics databases (columnar, MPP, etc.) 52%
Data warehouse appliances 46%
Hadoop/MapReduce 41%
Data virtualization software 28%
Complex event processing/real-time data integration tools 26%
NoSQL databases 21%
Other 3%

Hadoop, NoSQL, and similar platforms are attractive for providing new capabilities for combining and storing data sets, but they do not replace a data warehouse; they COMPLEMENT a data warehouse.

Source: Dimensional Research, 2015
Info-Tech emphasizes that a successful data warehouse should be framed as a program, not a project.

Due to the high implementation costs and intensive architectural planning up-front, there is a tendency to assume that the data warehouse will simply run itself once the “project” is completed.

Data warehousing is a long-term commitment to creating an environment that facilitates data delivery and information consolidation for business users.

“The biggest issue is that data warehouses are funded as a one-time project; they’re a living organism and need to have dedicated resources.”

– Chris Debo
Senior Manager, Schneider Downs & Co., Inc.

Investing in defined processes and governance up-front is arguably just as integral to data warehouse sustainability as architecture or the technology platform.
Embed **extensibility** within the foundation of your data warehouse program

Industry-specific innovation will continue to change the standard for how products and services are delivered to consumers, forcing organizations to find new ways to accommodate these demands.

Rapid advances in technology are creating new opportunities for data-driven insight, primarily through the harnessing of **big data**.

**And still…**

55% of organizations are currently **unimpressed** with the data warehouse’s ability to handle unstructured data ([The Information Difference, 2015](#)).

Pre-empt user dissatisfaction by building a core that can support a wider variety of business needs.
Use a three-piece framework to support your foundation

Inspired by architecture, a strong data warehouse foundation can be constructed using a **triangular structure** that is made of 1) foundational information needs, 2) foundational data architecture, and 3) foundational technology stack.

Future iterations will be built on top of this triangular structure. **Foundational understanding of information needs** has to be established to provide context for the data warehouse. **High-level data modeling** has to be completed to handle the common business entities of your organization. The **right technologies have to be selected** so that the stack is flexible yet solid for future development.
Follow Info-Tech’s approach to creating a robust data warehouse foundation

Business needs are going to constantly be evolving, with data and information needs of users following suit. Without creating a base from which you can adapt to these changes, IT runs the risk of alienating business users and hindering your company’s growth.

<table>
<thead>
<tr>
<th>Challenge</th>
<th>Solution</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 Identification of business processes</td>
<td><strong>Solution:</strong> Our blueprint uses in-depth brainstorming sessions as a starting point to get to the root of business opportunities and frustrations to provide greater insight into how IT can construct the right environment.</td>
</tr>
<tr>
<td>2 Standardizing data modeling guidelines</td>
<td><strong>Solution:</strong> By having a thorough understanding of the business objectives, selecting the right pieces to your data warehousing puzzle is much more straightforward. Thorough descriptions of the different architectural pieces will help your data warehouse project team pick the best technologies for your business environment.</td>
</tr>
<tr>
<td>3 Right-sizing the architecture</td>
<td><strong>Solution:</strong> By discussing a wide variety of processes to consider before jumping into the technology, your project team will have a set of flexible guidelines to ensure the integrity of the data warehouse as the program grows.</td>
</tr>
<tr>
<td>4 Creating sustainable governance</td>
<td><strong>Solution:</strong> By discussing a wide variety of processes to consider before jumping into the technology, your project team will have a set of flexible guidelines to ensure the integrity of the data warehouse as the program grows.</td>
</tr>
</tbody>
</table>
Keep track of your data warehouse success as your program develops

Your current environment is too heavily reliant on point-to-point sourcing of data that produces inconsistent reporting. Without a proper understanding of what your users want and what the underlying data model should be, business intelligence capabilities are impaired and business satisfaction decreases.

This blueprint focuses on building a foundation that allows for later iterations to be layered on top by detailing business processes that will inform data model and architecture design. Combined with creating governance bodies and policies, this is the beginning of your data warehouse program.

By establishing a program that places emphasis on governance as well as technology, your organization’s data warehouse environment becomes more sustainable, extensible, and capable of providing critical value to your business users.
Blueprint overview – Build an Extensible Data Warehouse Foundation

Prepare for the Data Warehouse Foundation Project
- Structure the Data Warehouse Project
  - Inputs: Current state and business context
  - Outputs: Understanding of relevant roles and responsibilities, Detailed project risks
- Data Warehouse Foundation Project Plan
- Position Descriptions
- Data Warehouse Foundation Work Breakdown Structure
- Data Warehouse Steering Committee Charter
- Data Warehouse Program Charter

Establish the Business Drivers and Data Warehouse Strategy
- Identify Essential Business Processes to Support Data Warehouse Strategy
  - Inputs: Detailed business objectives and processes
  - Outputs: Prioritized list of processes to guide data warehouse design
- Technology Capability Map

Plan for Data Warehouse Governance
- Plan for the Data Warehouse Centre of Excellence
  - Inputs: Data warehouse foundation scope
  - Outputs: Plan for establishing a data warehouse center of excellence, Roadmap for data warehouse foundation activities

Create the Core Data Model
- Inputs: Identified data elements that support business processes
- Outputs: Core “Rosetta Stone” documents that will inform core data warehouse tables
- Business Data Glossary
- Data Classification Inventory Tool
- Master Data Mapping Tool

Create an Architectural Strategy
- Inputs: Key processes
- Outputs: High-level data warehouse architecture model
- Data Warehouse Architecture Planning Tool

Define Operating Procedures
- Inputs: Data warehouse foundation scope
- Outputs: Documented procedures for data handling/manipulation, Service-level agreements
- Data Warehouse SOP Workbook
- Data Warehouse SLA
Use these icons to help direct you as you navigate this research

Use these icons to help guide you through each step of the blueprint and direct you to content related to the recommended activities.

This icon denotes a slide where a supporting Info-Tech tool or template will help you perform the activity or step associated with the slide. Refer to the supporting tool or template to get the best results and proceed to the next step of the project.

This icon denotes a slide with an associated activity. The activity can be performed either as part of your project or with the support of Info-Tech team members, who will come onsite to facilitate a workshop for your organization.
Info-Tech offers various levels of support to best suit your needs

**DIY Toolkit**
“Our team has already made this critical project a priority, and we have the time and capability, but some guidance along the way would be helpful.”

**Guided Implementation**
“Our team knows that we need to fix a process, but we need assistance to determine where to focus. Some check-ins along the way would help keep us on track.”

**Workshop**
“We need to hit the ground running and get this project kicked off immediately. Our team has the ability to take this over once we get a framework and strategy in place.”

**Consulting**
“Our team does not have the time or the knowledge to take this project on. We need assistance through the entirety of this project.”

Diagnostics and consistent frameworks used throughout all four options
## Build an Extensible Data Warehouse Foundation – project overview

<table>
<thead>
<tr>
<th>Prepare for the Data Warehouse Foundation Project</th>
<th>Establish the Business Drivers and Data Warehouse Strategy</th>
<th>Plan for Data Warehouse Governance</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.1 Structure the data warehouse project</td>
<td>2.1 Identify essential business processes to support</td>
<td>3.1 Plan for the data warehouse centre of excellence</td>
</tr>
<tr>
<td>1.2 Review data management impacts</td>
<td>2.2 Create the core data model</td>
<td>3.2 Refine operating procedures</td>
</tr>
<tr>
<td></td>
<td>2.3 Create an architectural strategy</td>
<td></td>
</tr>
<tr>
<td></td>
<td>2.4 Explore technology options</td>
<td></td>
</tr>
<tr>
<td>Discuss structuring your project team, defining success metrics and risks, and organizing a steering committee.</td>
<td>Walk through how to characterize operational, tactical, and ad hoc business processes that will guide the data warehouse.</td>
<td>Plan for the formation of a data warehouse center of excellence.</td>
</tr>
<tr>
<td>Discuss the impacts of other data management practices on your data warehouse foundation project.</td>
<td>Discuss the four “Rosetta Stones” for data modeling.</td>
<td>Discuss defining standard operation procedures and service-level agreements for the data warehouse.</td>
</tr>
<tr>
<td></td>
<td>Develop an architecture strategy based on business and data needs, and review data warehouse vendor landscape.</td>
<td></td>
</tr>
</tbody>
</table>

### Guided Implementations

**Module 1:** Prepare for the Data Warehouse Foundation Project

- Prepare for the Data Warehouse Foundation Project

**Module 2:** Establish the Business Drivers and Data Warehouse Strategy

- Establish the Business Drivers and Data Warehouse Strategy

**Module 3:** Plan for Data Warehouse Operations

- Plan for Data Warehouse Operations

---

*Info-Tech Research Group*
# Workshop overview

Contact your account representative or email [Workshops@InfoTech.com](mailto:Workshops@InfoTech.com) for more information.

|------------|-------------|---------------------------------|----------------------------------|-------------------------------------------------|--------------------------------------------------|
|            | Pre-Workshop | 1.1 Identify foundation project team and create a RACI chart.  
1.2 Develop timelines for foundation project completion.  
1.3 Define critical success factors, key performance metrics, and project risks.  
1.4 Define the current and future states for key data management practices. | 2.1 Understand the most fundamental needs of the business.  
2.2 Define the data warehouse vision, mission, purpose, and goals.  
2.3 Detail the most important operational, tactical, and ad hoc activities the data warehouse should support.  
2.4 Link the processes that will be central to the data warehouse foundation. | 3.1 Walk through the four-column model and business entity modeling as a starting point for data modeling.  
3.2 Create data models using the business data glossary and data classification.  
3.3 Identify master data elements to define dimensions.  
3.4 Design lookup tables based on reference data.  
3.5 Create a fit-for-purpose data warehousing model. | 4.1 Develop a technology capability map to visualize your desired state.  
4.2 Establish a data warehouse center of excellence.  
4.3 Create a data warehouse foundation roadmap.  
4.4 Define data warehouse service-level agreements.  
4.5 Create standard operating procedures. |
|            |             |                                               | |                                                       |                                                   |
| Deliverables | 1. Workshop Itinerary  
2. Workshop Participant List | 1. Job Descriptions and RACI  
2. Data Warehouse Foundation Project Plan  
3. Work Breakdown Structure  
4. Data Warehouse Steering Committee Charter | 1. Data Warehouse Vision and Mission  
2. Documentation of Business Processes  
3. Data Warehouse Program Charter | 1. Business Entity Map  
2. Business Data Glossary  
3. Data Classification Scheme  
4. Data Warehouse Architecture Model | 1. Technology Capability Map  
2. Project Roadmap  
3. Service-Level Agreement  
4. Data Warehouse Standard Operating Procedure Workbook |
PHASE 1

Prepare for the Data Warehouse Foundation Project

Build an Extensible Data Warehouse Foundation
## Phase 1 outline

Complete these steps on your own, or call us to complete a guided implementation. A guided implementation is a series of 2-3 advisory calls that help you execute each phase of a project. They are included in most advisory memberships.

### Guided Implementation 1: Prepare for the Data Warehouse Foundation Project

**Proposed Time to Completion (in weeks): 2-3**

<table>
<thead>
<tr>
<th>Step 1.1: Structure the Data Warehouse Project</th>
<th>Step 1.2: Review Data Management Impacts</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Start with an analyst kick off call:</strong></td>
<td><strong>Review findings with analyst:</strong></td>
</tr>
<tr>
<td>• Discuss the structure and capabilities of the current data management team.</td>
<td>• Review identified project risks, as well as challenges in defining the data warehouse foundation team.</td>
</tr>
<tr>
<td>• Review the necessary roles and responsibilities required for a data warehouse foundation project.</td>
<td>• Discuss the data management practices that are most critical to successful data warehousing.</td>
</tr>
<tr>
<td>• Discuss potential risks for the project.</td>
<td><strong>Then complete these activities...</strong></td>
</tr>
<tr>
<td><strong>Then complete these activities...</strong></td>
<td>• Gather project team to discuss current data management practices.</td>
</tr>
<tr>
<td>• Define the project team roles and responsibilities.</td>
<td>• Define current and future state for data management practices.</td>
</tr>
<tr>
<td>• Create rough timelines for project completion.</td>
<td><strong>With these tools &amp; templates:</strong></td>
</tr>
<tr>
<td>• Define critical success factors, project metrics, and risks.</td>
<td>Data Management Assessment and Planning Tool</td>
</tr>
<tr>
<td><strong>With these tools &amp; templates:</strong></td>
<td>Data Warehouse Program Charter Template</td>
</tr>
<tr>
<td>Data Warehouse Architect, BI Specialist and DI Specialist Job Descriptions</td>
<td></td>
</tr>
<tr>
<td>Data Warehouse Foundation Project Plan Template</td>
<td></td>
</tr>
<tr>
<td>Data Warehouse Work Breakdown Structure Template</td>
<td></td>
</tr>
</tbody>
</table>

### Phase 1 Results & Insights:

- Having a good team is better than having an ideal plan. A good team can build an ideal plan but not vice versa.
- You ran a bit but you fell. Go back and learn how to walk before you can run. Revisit essential data management practices and make sure you have the minimal maturity before you build a data warehouse.

Call 1-888-670-8889 or email GuidedImplementations@InfoTech.com for more information.
**Phase 1, Step 1: Structure the Data Warehouse Foundation Project**

**This step will walk you through the following activities:**

1.1.1 Create a RACI chart for the foundation project team.
1.1.2 Understand what a data warehouse can and cannot enable.
1.1.3 Define critical success factors and key performance metrics for data warehouse foundation.
1.1.4 Identify and prioritize project risks.
1.1.5 Develop timelines for foundation project completion.

**This step involves the following participants:**
- Project sponsor
- Project manager
- IT specialists
- External experts (as needed)

**Outcomes of this step**

- Assign roles and responsibilities for the project.
- Develop a comprehensive and actionable data warehouse foundation project plan.
The foundation project team goes beyond the data warehouse architect role

The data warehouse is intended to be a centralized point that facilitates the flow of data from source systems and applications to a usable product for informed business decision making. The foundation team needs to reflect the **end-to-end nature of data delivery** and have roles that are responsible, accountable, consulted, and informed on what activities are occurring at each stage in the process.
Establish the data warehouse foundation project team

The core project team oversees the management and governance of the project and interacts with the overall data warehouse sponsors and champions to ensure the data warehouse truly creates value for the enterprise and people are adopting it.

**Goals**

- Ensure a strong DW foundation is built.
- Make sure the foundation is strong yet flexible for future additions.
- Develop processes around execution of future iterations.
- Integrate future additions to the base foundation.
- Make key decisions on the DW foundation project.

**Tasks**

- Monitor and control the overall scope.
- Align data warehouse to strategic business needs.
- Maintain foundational information needs.
- Create a business glossary and high-level entity modeling.
- Select and implement technological components of the technology stack.

**Info-Tech Insight**

Personnel in the core project team should be *experienced* both in the data warehousing field and in your organization. **External advisors** might be brought in to provide an unbiased opinion on your project. The roles are *permanent long-term positions* so that institutional data warehousing experiences and memory can be maintained.
Staffing can make or break the data warehouse foundation project – manage your people resources effectively

The most optimal data warehouses function in an environment where all members of the project team have an understanding of what needs to occur at each stage in the data flow. This is often unattainable in most organizations, given the constraints on time, physical location of team members, resources dedicated to the project, etc.

More than 20 IT staff

Large organizations with a sizeable IT team will be able to assign the main facets of the project (data integration, data architecture and modeling, solution delivery, user testing) to specific individuals or sub-teams.

Over the course of the project, the groups in these areas should have training sessions to facilitate knowledge transfer and provide a more seamless solution.

Smaller organizations will need team members to wear multiple hats and have responsibilities that span the solution end to end.

Roles such as the data architect will need specialized knowledge that goes beyond the depth required of the whole team. Some BAs would also be involved in the development of BI content.

Fewer than 20 IT staff
Ensure you have **data warehouse architect** capabilities on your team

A **data warehouse architect** is an indispensable role in the creation of a data warehouse foundation. They need to have a thorough understanding of how business processes and objectives are translated into technical requirements to be implemented by database admin and developers.

Their activities revolve around:

- Strategy and Planning
- Logical and Physical Data Models
- Review and Approve Architectural Changes

Leverage Info-Tech’s **Data Warehouse Architect Job Description Template** to source for candidates externally, or identify a current member of staff with the right capabilities.
Consolidate your data sources seamlessly with a data integration specialist role

A data integration specialist needs to be present to coordinate the organization’s data integration (DI) patterns with appropriate DI architectures, such as ETL or other transformation/staging layers. The DI specialist must ensure that adoption of DI solutions are in compliance with legal and regulatory standards and align with overarching business strategies.

Leverage Info-Tech’s **Data Integration Specialist Job Description Template** to source for candidates externally, or identify a current member of staff with the right capabilities.
Bring a **BI specialist** on board to translate the data warehouse assets into actionable resources for business users

The **business intelligence specialist** role is to strategically design and implement BI software and systems, including integration with databases and data warehouses. This includes selecting, blueprinting, gathering requirements (in conjunction with business analysts), designing, and rolling out BI solutions to end users.

The business intelligence specialist is also responsible for ensuring high levels of BI availability through support functions and in-depth testing.

Leverage Info-Tech’s **BI Specialist Job Description Template** to source for candidates externally, or identify a current member of staff with the right capabilities.
Assemble the core project team and assign responsibility with a RACI chart

- Establish data warehouse project governance.
- Create data warehouse project plan and roadmap.
- Manage project sponsors and champions.
- Identify and assign subsequent iteration teams.
- Track data warehouse progress and monitor data warehouse metrics.

### Mandate

<table>
<thead>
<tr>
<th>Presents</th>
<th>Note</th>
<th>Recommendations</th>
</tr>
</thead>
</table>
| Role     | Distinguish between roles and the persons in the role, e.g. use BAs, not names of individual BAs like David, John, and Tori. | Key roles for the core team include:  
- Project manager, project sponsor, data warehouse architect, core BAs, data integration specialist, BI specialist |
| Involvement | R = responsible (perform the responsibility)  
A = approve (make decisions)  
C = consulted (provide opinions)  
I = informed (keep in the loop) |  
- Inform as many project team members as possible.  
- Consult people with experience. Consult externally to get unbiased opinions. Formulate your own position after consulting.  
- If multiple persons are covering the same role, define responsibilities in detail to avoid duplicate efforts or a lack of coverage. |
| Responsibility | Distinguish between responsibilities and tasks, e.g. project management is a responsibility vs. the creation of a status report is a task. | Key responsibilities of the core team include:  
- Project governance, stakeholder management, stakeholder education, approval of later iteration scope, review of iteration outcomes, and monitoring project success. |
Use a RACI chart to specify who is responsible, accountable, consulted, and informed about a task. The chart below is a sample.

### Output
- Defined involvement for foundation project team

### Materials
- Sample RACI
- *Data Warehouse Foundation Project Plan Template*

### Participants
- Data warehouse foundation team

<table>
<thead>
<tr>
<th></th>
<th>CIO</th>
<th>Project Sponsor</th>
<th>Project Manager</th>
<th>Business Analyst(s)</th>
<th>Data Warehouse Architect</th>
<th>Data Integration/BI Specialists</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Project Scoping</strong></td>
<td>I</td>
<td>C</td>
<td>A</td>
<td>R</td>
<td>C</td>
<td>C</td>
</tr>
<tr>
<td><strong>Requirements Gathering</strong></td>
<td>I</td>
<td></td>
<td></td>
<td></td>
<td>C</td>
<td>I</td>
</tr>
<tr>
<td><strong>Solution Building</strong></td>
<td>I</td>
<td></td>
<td></td>
<td></td>
<td>R</td>
<td>R</td>
</tr>
<tr>
<td><strong>Solution Testing</strong></td>
<td>I</td>
<td></td>
<td></td>
<td></td>
<td>R</td>
<td>I</td>
</tr>
<tr>
<td><strong>Results Gathering</strong></td>
<td>I</td>
<td>C</td>
<td>A</td>
<td>R</td>
<td>I</td>
<td>I</td>
</tr>
<tr>
<td><strong>Results Communication</strong></td>
<td>R</td>
<td>R</td>
<td>A</td>
<td>I</td>
<td>I</td>
<td>I</td>
</tr>
<tr>
<td><strong>Project Closure</strong></td>
<td>R</td>
<td>A</td>
<td>R</td>
<td>I</td>
<td>I</td>
<td>I</td>
</tr>
</tbody>
</table>
Understand what a data warehouse is and isn’t

1.1.2

To properly build the right architecture, the data warehouse foundation team and relevant stakeholders need to be aware of what a data warehouse is meant to enable and what activities are better suited to other data technologies.

1. Gather the data warehouse IT team (data architects, integration specialists, etc.) as well as senior IT and business stakeholders (ideally members who would be on the steering committee).

2. Have each person, using sticky notes, write down three statements about what they believe a data warehouse is meant to do.

3. Use another three notes to write down statements about what functions a data warehouse cannot or should not perform.

4. With the technical staff facilitating, consolidate the overlapping statements to build consensus in the expectations. Discuss any discrepancies in what people expect of the data warehouse, particularly if there are contrary statements.
   - Have participants defend their rationales for what the data warehouse should or should not be able to do.

Ensure that several key business stakeholders participate (or at the very least sit in on) in this activity. Increasing the business-user understanding of data technologies will help manage expectations and ideally reduce the chances of hasty platform purchases.
Establish critical success factors and metrics to measure the success of the data warehouse

One explanation for the high failure rate for data projects is that **critical success factors (CSFs)** are not defined. CSFs are necessary and essential for a data warehouse project to achieve its mission.

**Key performance indicators (KPIs)** are also needed to measure the magnitude of project success. Collectively, CSFs and KPIs measure 1) if the project is successful and 2) how successful it is.

\[
\text{CSFs (Indicate successful or not)} \times \text{KPIs (Indicate how successful or unsuccessful)} = \text{Success registered and Success quantified}
\]

1. Get the core project team and some other senior IT stakeholders in a room.
2. If applicable, review any KPIs and metrics currently in use to evaluate IT’s activities, particularly around data flow, architecture, and delivery.
3. Discuss any recent interactions with or feedback provided by business users, positive or negative, with respect to IT meeting information needs. Look to identify any recurring themes or concerns as metrics/success factors that should be given priority.
4. Develop a preliminary list of CSFs and KPIs.
   - Once the high-level driving statements are articulated in phase 2 (vision, mission, objectives, etc.), have the project team reconvene to validate the current list of KPIs and CSFs or make additions prior to the start of the project.
Establish critical success factors and metrics to measure the success of the data warehouse (continued)

1.1.3

30-45 minutes

Project Metrics

Goals:
1. Accelerate data warehouse construction
2. Higher build quality

Metrics:
• Time to complete iteration projects
• Number of projects dedicated to re-work
• Amount of time spent on re-work

Sample CSFs

• Data in the data warehouse must be named in a business-friendly fashion as outlined in the naming convention guidelines.
• Historical data must be tracked in the data warehouse. Both granular and summarized forms of historical data should be presented.
• Data has to be transformed to create YTD, YTD for the previous year, YTD for the last five years.
• Customer and product data should be integrated so that users can analyze with the two dimensions.

Sample KPIs

Efficient
• Time to information

Effective
• Data quality score

Affordable
• Percent of self-serve users

Comprehensive
• Number of integrated data sources
Mitigate common risks when building the data warehouse foundation

Addressing risks associated with initial data warehouse efforts is essential as you want to build a solid foundation that can be extended by subsequent iterations. Neglecting to properly identify these risks will jeopardize the reliability and sustainability of the data warehouse.

Risks are typically found in the below logical components of the data warehouse foundation:

<table>
<thead>
<tr>
<th>Information Needs</th>
<th>Data</th>
<th>Technology</th>
</tr>
</thead>
<tbody>
<tr>
<td>Users have too many information needs whereas the supply is limited.</td>
<td>Different data architects have different views on data modeling.</td>
<td>Indecisiveness in determining which platforms/tools need to be retired to accommodate the data warehouse.</td>
</tr>
<tr>
<td><strong>Recommendation:</strong> Document all information needs. Later on in the project you will prioritize them and execute on the most impactful ones.</td>
<td><strong>Recommendation:</strong> Encourage different views on the data model. Compare and discuss the views and develop a framework to select the most suitable data modeling methodology.</td>
<td><strong>Recommendation:</strong> Communicate and engage the impacted stakeholders. Outline plans to migrate the users from the legacy analytics tools to the new data warehousing platform.</td>
</tr>
<tr>
<td>Different departments have conflicting information needs.</td>
<td>Data granularity is not well defined.</td>
<td>Determining the technology stack based on “we are a vendor X shop.”</td>
</tr>
<tr>
<td><strong>Recommendation:</strong> Encourage a consensus of the conflicting needs. A consensus will ease modeling and implementation. If not, conflicting needs can be handled by developing different solutions.</td>
<td><strong>Recommendation:</strong> Educate stakeholders on data granularity. Discuss data granularity in the warehouse using the detailed information needs as a tool for scoping.</td>
<td><strong>Recommendation:</strong> The choice of the technology stack should not be relying entirely on the notion of “vendor X shop.” Data warehousing stack is modular in nature and its selection should be rationalized.</td>
</tr>
</tbody>
</table>
Develop a risk matrix to identify, prioritize, and mitigate data warehouse project risks

Given the identified business processes and desired architecture, prioritize the risks associated with your data warehouse project. Use a prioritization scheme to determine which risks need to be folded into the earliest stages of the implementation plan.

<table>
<thead>
<tr>
<th>Risk Color</th>
<th>Action</th>
</tr>
</thead>
<tbody>
<tr>
<td>Red</td>
<td>First Priority Risks – Mitigate immediately without delay.</td>
</tr>
<tr>
<td>Yellow</td>
<td>Second Priority Risks – Mitigate in days. Try to reduce the likelihood or impact.</td>
</tr>
<tr>
<td>Green</td>
<td>Third Priority Risks – Mitigate in the near future. Monitor them proactively.</td>
</tr>
<tr>
<td>Light Green</td>
<td>Fourth Priority Risks – Mitigation is not needed. Deal with the risks when they arise.</td>
</tr>
</tbody>
</table>

**Info-Tech Insight**

Organizational culture presents a significant risk to data warehouse projects as business processes will likely need to be adjusted to support data integrity in the warehouse. Revisit past efforts around improving data-driven decision making and get executive involvement to communicate the importance of this program.
Avoid falling victim to common warehouse project risks

Sample risk prioritization chart:

<table>
<thead>
<tr>
<th></th>
<th>Low Impact</th>
<th>Medium Impact</th>
<th>High Impact</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>High Probability</strong></td>
<td>Users revert back to Microsoft Excel to analyze data</td>
<td>The data warehouse does not satisfy the business need</td>
<td>Fail to address data issues: quality, integration, definition</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Data warehouse focuses too much on operational needs</td>
<td>Inadequate communication with stakeholders throughout the project</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Poor documentation creates confusion and reduces user adoption</td>
<td>Users find the BI interface to be confusing</td>
</tr>
<tr>
<td><strong>Medium Probability</strong></td>
<td></td>
<td>Failure to define and monitor KPIs</td>
<td>Lack of support from the sponsors</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Poor training results in low user adoption</td>
<td>No centralized governance</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Organization culture is resistant to the change</td>
<td>Poor training results in misinformed users</td>
</tr>
<tr>
<td><strong>Low Probability</strong></td>
<td></td>
<td></td>
<td>Business units independently invest in information silos</td>
</tr>
</tbody>
</table>


Establish a **data warehouse steering committee** to solidify your data warehousing program

> If you have a data warehouse group that is solely constructing and building according to requirements, you’re going to fall short and end up **building a silo** without someone at the top that is implementing some level of overarching governance.

— Anonymous
Senior Data Architect, Consulting Firm

A **data warehouse steering committee** serves to:
- Discuss data warehouse initiatives and prioritize them according to importance.
- Make major decisions around the data warehouse.
- Provide a touchpoint to share project updates, socialize success stories, and facilitate insight sharing.

### Data Warehouse Steering Committee

<table>
<thead>
<tr>
<th>Department</th>
<th>LOB 1</th>
<th>LOB 2</th>
<th>LOB 3</th>
<th>IT</th>
</tr>
</thead>
<tbody>
<tr>
<td>Senior executives</td>
<td>VP</td>
<td>SVP</td>
<td>VP 1, VP 2</td>
<td>CIO, CDO, CISO</td>
</tr>
<tr>
<td>Middle management</td>
<td>Manager 1, manager 2</td>
<td>Manager 1, manager 2</td>
<td>Manager 1, manager 2</td>
<td>Architects, DBAs, data stewards, BI specialists, integrators</td>
</tr>
<tr>
<td>Frontline workers</td>
<td>Representative 1,</td>
<td>Representative 1,</td>
<td>Representative 1,</td>
<td>Developers, analysts</td>
</tr>
<tr>
<td></td>
<td>representative 2</td>
<td>representative 2</td>
<td>representative 2</td>
<td></td>
</tr>
</tbody>
</table>

**Info-Tech Tip**

Limit the number of representatives to **20** (regardless of the size of your organization) to balance efficiency and the level of representation. All major decisions have to be passed by representatives in IT and relevant business groups.
Leverage our *Data Warehouse Steering Committee Charter* as a starting point for coordinating data warehouse stakeholders

The creation of a governance structure like a steering committee can provide advice and guidance for organizations implementing a data warehouse.

The *Data Warehouse Steering Committee Charter* is a customizable template for delineating responsibilities among the members of the committee to further an organization’s data warehouse program. This template will help to:

- Clearly state the objectives of the committee.
- Define granular responsibilities and protocols.
- Make the committee accountable for data warehouse decisions.
- Establish the data warehousing governance committee.
- Define actions and next steps after the committee is formed.

*Info-Tech Insight*

Smaller organizations may opt to fold in the function and responsibilities of the data warehouse steering committee into an existing data management group, such as the **data governance steering committee**.

---

**Data Warehouse Steering Committee Charter Template**

**Introduction: How to Use This Template**

A data warehouse steering committee is an excellent tool for making both IT and the business accountable for data governance decisions. A charter is an essential document for defining the scope and purpose of the committee. It should also clarify the objectives of data warehousing and BI initiatives, as well as define how this committee will operate, including its role within the organization and the responsibilities of the committee.

To use this template, simply customize any text below in dark gray to fit the needs of your enterprise.

This document is intended for use as a guide, and should be used in accordance with your enterprise’s legal and compliance environment.

**Purpose**

The Data Warehouse Steering Committee (DWSC) oversees the data warehouse program initiatives for [insert company name]. Members of the DWSC are appointed by [insert role], and are accountable to the senior executives within IT and the business. The core objectives of a data warehouse governance program are to:

1. Strengthen data-driven decisions.
2. Ensure data warehouse development projects are consistently defined and well-structured.
3. Establish data warehouses as an enterprise data clearinghouse.
4. Improve the consistency of data use across the enterprise.
5. Encourage sharing of data and information.

The DWSC is a strategic-level, interdisciplinary group that is tasked with the responsibility of overseeing organizational data warehousing, including building and maintaining the existing data warehouse, policies, and standards, and change management.

The DWSC has the authority to make decisions on all aspects of data warehousing for [insert company name].

The DWSC will identify the following:

1. A mandate for creating and implementing policies and procedures to ensure the fulfillment of data warehouse data and metadata.
2. The creation of a formal decision-making structure for data warehouse approval.
3. The provision of a two-way method for communicating (e.g., a reporting structure and/or communication plan) about data warehouse initiatives across the organizations.
4. Formal procedures for involving business and technical groups for data warehouse projects.
5. The defining and enforcement of best practices around data sharing.

The DWSC is not responsible in any way for operating budget or expenditures. Day-to-day operations of both IT and the business units are expected to be influenced by decisions made by the DWSC, but not to be directly carried out by the committee. The DWSC will advise business unit staff to make any decisions or take any action which violates enterprise policies or regulations.
Keep track of your data warehouse initiatives with status updates

Given the significant investment required to build a data warehouse, it is critical that the moving parts in the foundation are actively monitored to prevent project rework and wasted resources.

Info-Tech’s Project Status Report Template allows you to provide an efficient summary of:

• Purpose and scope
• Objectives (and link to strategic objectives)
• Planned/unplanned and completed/uncompleted activities
• Initiative participation
• Project milestones and timelines

Info-Tech Tip

Streamline your data warehouse steering committee meetings by using these templates as the standard for project/initiative presentations and circulating them in advance of each meeting.
Construct a **work breakdown structure** for your data warehouse project

A **work breakdown structure (WBS)** is a useful project management tool that aids in clearly parsing overall objectives and deliverables for a project into manageable and measureable tasks.

**VALUE:**

- Assists with project scoping by better estimating the number of tasks required to deliver an overall component.
- When combined with a RACI chart, task responsibilities and accountabilities are less ambiguous.
- Clear definition of tasks breeds commitment to the project outcomes.
- Outlining project tasks in detail can reduce re-work/backtracking once the project is underway.

With a data warehousing project, timely delivery of results is crucial, particularly in **winning over skeptics in senior management**.
Leverage Info-Tech’s *Data Warehouse Work Breakdown Structure Template* to detail tasks

**Overview**

Use this tool to create a detailed summary of the steps and activities required to carry out your Agile data warehouse initiative.

**Recommended Participants**

- Data Warehouse Project Team, Project Sponsor
- Data Warehouse Architects, IT Specialists, Business Representatives *(must involve)*
- Additional DM Professionals, Relevant SMEs *(nice to involve)*

---

**Data Warehouse Work Breakdown Structure**

<table>
<thead>
<tr>
<th>WBS</th>
<th>Task</th>
<th>Duration</th>
<th>Predecessors</th>
<th>Resources</th>
<th>Involvement</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Project Kick-off</td>
<td>1 day</td>
<td>Test</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1.1</td>
<td>Define objectives</td>
<td>1 day</td>
<td>Test</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1.2</td>
<td>Create scope</td>
<td>1 day</td>
<td>Test</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1.3</td>
<td>Define requirements</td>
<td>1 day</td>
<td>Test</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1.4</td>
<td>Design architecture</td>
<td>1 day</td>
<td>Test</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>Requirement Gathering</td>
<td>3 days</td>
<td>PM, Project Sponsor, Business Units</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2.1</td>
<td>Identify key business goals and objectives</td>
<td>3 days</td>
<td>PM, Project Sponsor, Business Units</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2.2</td>
<td>Identify key business processes and data sources</td>
<td>3 days</td>
<td>PM, Project Sponsor, Business Units</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2.3</td>
<td>Identify data warehousing user cases</td>
<td>3 days</td>
<td>PM, Project Sponsor, Business Units</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2.4</td>
<td>Define data warehouse requirements, data and processes</td>
<td>3 days</td>
<td>PM, Project Sponsor, Business Units</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2.5</td>
<td>Define data warehouse vision, goals, and objectives</td>
<td>3 days</td>
<td>PM, Project Sponsor, Business Units</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2.6</td>
<td>Define data warehouse architecture and processes</td>
<td>3 days</td>
<td>PM, Project Sponsor, Business Units</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2.7</td>
<td>Define data warehouse design specifications</td>
<td>3 days</td>
<td>PM, Project Sponsor, Business Units</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

---

**Tool Guideline**

<table>
<thead>
<tr>
<th>Tab 1</th>
<th>Instructions</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Use this tab to understand how the tool works.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Tab 2</th>
<th>WBS</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Use this tab to characterize your Agile data warehouse project tasks by duration, resources, and involvement.</td>
</tr>
</tbody>
</table>

Download the *Data Warehouse Work Breakdown Structure Template*.  

---
Set time expectations for data warehouse foundation completion

<table>
<thead>
<tr>
<th>1.1.5</th>
<th>30-45 minutes</th>
</tr>
</thead>
</table>

### Data Quality
- What is the data quality at the level of the source systems?
- Have data cleansing activities been undertaken?
- Are procedures to correct data errors in place?

### Data Complexity
- How different are the data sets being integrated?
- Is data ownership well defined, or does it span multiple business units?
- Have fields been standardized within each data source?

### People
- How many staff can be dedicated to the project?
- What are the internal skillsets that can be leveraged? Will external consultants/contractors be needed?

### Technology
- How complex is your organization’s tool landscape?
- Are solutions shared or siloed amongst business units?

### User Base
- How many users is the data warehouse intended to support?
- What is the demand for analytics with respect to volume and complexity?

### INPUT
- Current data environment

### OUTPUT
- Expected timeline for completion of foundation project

### Materials
- Data Warehouse Foundation Project Plan Template

### Participants
- Data warehouse foundation project team

---

**Info-Tech Insight**

Delivering quick wins on projects is important for full-scale adoption of the solution, but do not compromise the integrity of data modeling and quality by rushing the planning and governance stages of the foundation project.
Consolidate the outputs of the exercises incorporated in this blueprint in the *Data Warehouse Foundation Project Plan*

The documentation for creating the data warehouse foundation needs to be centralized such that a clear line can be drawn from the business needs and processes through to the architecture planning and technology.

This template is intended to be a **living document** that should be revisited and updated as business needs change and new data warehouse extension projects are undertaken.

Use Info-Tech’s *Data Warehouse Foundation Project Plan Template* to capture all the project outputs.
Phase 1, Step 2: Review Data Management Impacts

This step will walk you through the following activities:
1.2.1 & 1.2.2 Discuss the state of current data management practices related to data warehousing.
1.2.3 Define the current and future states for key data management practices.

This step involves the following participants:
- Data Warehouse Foundation Team (Project Manager, Data Warehouse Architects, Business Analysts, Business Representatives, etc.)

Outcomes of this step
- Gain a holistic understanding of how other data management practices support an effective data warehouse.
- Identify opportunities for data management practice improvement.
Revisit existing **data management practices** when beginning the data warehouse project

As part of establishing robust governance at the outset of the data warehouse foundation project, other data management practices need to be appropriately evaluated to identify activities **beyond** the data warehouse implementation that will enable success.

Fifty-eight percent of organizations identify dealing with **data quality/consistency** issues as a problem when implementing a data warehouse.

Source: TechTarget, 2013

Info-Tech has identified four of DAMA DMBOK’s data management practices that should be evaluated at this stage of the project:

- Data Quality
- Data Governance
- Metadata
- Reference & Master Data

**Info-Tech Insight**

If business intelligence activities supplied by the enterprise data warehouse is the **destination**, then investing in foundational data management practices is the **journey**.
Address **data quality** practices to ensure business users can have confidence in the data warehouse

### Guiding Principles and Value

**Monitor** the fitness of data in the environment.

**Improve** the quality of data through ongoing cleansing efforts and correction plans.

### Overview

Data quality encompasses planning, implementation, and control activities that apply quality management techniques to measure, assess, improve, and ensure the fitness of data for use (Data Management Book of Knowledge, 2009).

### Data Warehouse Relevance

The value of a data warehouse platform is questionable if the data flowing in has not been cleansed according to business information needs. Bad data left unattended jeopardizes the effectiveness of any downstream business intelligence and analytics.

### Value of Data Quality Management

Data quality and its management is integral to the performance and perception of the entire data management program. Data quality management:

- Aligns integrity and fitness of data to the business’ requirements.
- Gauges the quality of data and the performance of the data management practices that support its delivery.

### Objectives of Data Quality Management

- Ensure **usable** and **trustworthy** data is available to the business and its processes.
- Analyze and recommend improvements to data management components in order to support improved data quality.
- Coordinate efforts between **IT** and the **business** that manage and maintain the quality of data assets on an ongoing basis.
- Investigate data quality issues, determining **root causes** and designing **corrective plans** that enable the maintained integrity of data assets.

### Critical Success Factors for Effective Data Quality Management

- Activities of data quality are recognized as a business function that must be supported by executives and executed tactically by business and IT staff.
- Continual cleansing of data assets (automated and manual).
- Periodical reviews and improvements of the management practices for data.
Data governance preserves integrity by dictating how data assets can and should be handled

Guiding Principles and Value

Governs and directs the data management practices supporting the organization’s treatment of data assets.

Overview

Data governance is the exercise of authority, control, and shared decision making in terms of planning, monitoring, and enforcement over the management of data assets. It can also be reviewed as “high-level, executive data stewardship” (Data Management Book of Knowledge, 2009).

The value of data governance comes from its ability to set policies and standards, and ensure their adherence in both business and IT environments. It governs:

- How the business consumes and stewards data assets.
- How data management professionals execute components of data management.

Key Objective

Governance of data assets will become embedded in the processes and perspectives of the business.

How to get there:

- Self-governing processes, assisted by tools and widely adopted governance principles.
- A data-oriented culture where the business owns the data and takes responsibility for its management.

Data Warehouse Relevance

While the enterprise data warehouse can pave the way for greater information sharing between business units, a lack of governance increases the likelihood that the wrong people are handling data inappropriately.

Defined policies and procedures need to exist around where data can be integrated and transformed that will grant business users enough freedom to generate insights, but also preserve the integrity of internal data assets.

Separated roles and responsibilities help to create a system of checks and balances that ensures policy alignment to internal and external requirements for data management.
Discuss data management practices that enable the data warehouse

1.2.1

Having reviewed the fundamental aspects of fundamental data management practices, organize a roundtable session with the foundation project team and senior IT stakeholders.

To facilitate an open dialogue, pose the prompting questions below to the meeting participants.

When there is consensus on data management components that are lacking, document the gap in the Data Warehouse Foundation Project Plan Template.

Data Governance Considerations:
- Does a central governance body exist?
- Are data management policies and procedures defined, socialized, and enforced?
- Are there identified data stewards/data owners? Have their expectations and accountabilities been clearly articulated?
- Are changes in procedures and practices aligned with organizational change management activities?

Data Quality Management Considerations:
- Do we have procedures for validating the trustworthiness and quality of external data sources?
- Are there data quality checks at each stage of the data lifecycle (ingestion, storage, usage)?
- Is there an organizational culture around maintaining data quality as an asset?

Input
- Current state of data management practices

Output
- Pre-implementation improvement activities

Materials
- Data Management Assessment and Planning Tool
- Data Warehouse Foundation Project Plan Template

Participants
- CIO
- SVP data management
- The data warehouse foundation team
Make your data warehouse environment easy to navigate with appropriate metadata management

Overview
Metadata is structured information that describes, explains, locates, or otherwise makes it easier to retrieve, use, or manage an information resource. Metadata is often called data about data or information about information (NISO).

Metadata management is the function that manages and maintains the technology and processes that creates, processes, and stores metadata created by business processes and data.

Value of Effective Metadata Management
- Supports the traceability of data through an environment.
- Creates standards and logging that enable information and data to be searchable and catalogued.
- Metadata schemas enable easier transferring and distribution of data across different environments.

Data Warehouse Relevance
Investing in a metadata management practice increases the findability of your data assets, improving the efficiency and speed with which reports can be created. It also makes data more intelligible for business users, paving the way for richer requirements gathering conversations with IT, as well as self-service business intelligence.

Metadata supports the use of:
- Big Data (unstructured data)
- Content and Documents (unstructured and semi-structured data)
- Structured data (master, reference, etc.)

Critical Success Factors of Metadata Management
- Consistent and documented data standards and definitions.
- Architectural planning for metadata.
- Incorporation of metadata into system design and the processing of data.
- Technology to support metadata creation, collection, storage, and reviews (meta-data repository, meta marts, etc.).

Source: IDG via Oracle, 2015

The projected amount of the unstructured data in the digital universe and an organization’s environment by 2020.

As data becomes more unstructured, complex, and manipulated, the importance and value of metadata will grow exponentially and support improved:
- Data consumption
- Quality management
- Risk management

Source: IDG via Oracle, 2015
Reference and master data management creates consistency in the interpretation of your most valuable data assets

**Definitions**

**Master Data**
An organization’s data about:
- Parties (customers, employees, etc.)
- Products
- Finances
- Locations

**Reference Data**
Data with standardized formats and restricted value options (e.g. zip codes or country codes).

**Master Data Management**
The control over master data values to enable consistent, shared, and contextual use across systems of the most accurate, timely, and relevant version of truth about essential business entities (DAMA, 2009).

**Critical Success Factors for MDM**
- Business buy-in and leadership in executing an MDM project.
- A deep understanding of how master and reference data elements are created and consumed by business processes.
- Relationships and architectural planning is planned, documented, and applied to processes and IT systems.
- A successful balance of processes, resourcing, and technology that supports an MDM program.
- An interoperable environment that successfully integrates systems and propagates master files when necessary.

**Objectives of MDM**
- Improve the quality and accessibility of the organization’s most valuable and widely used data assets.
- Identify singular versions of the truth (a singular authoritative master file) for data elements on an enterprise-wide level.

**Data Warehouse Relevance**

Master data elements are essential in creating the fact tables and establishing the dimensions that will be the core of the data warehouse.

Building robust policies, detailed mapping of data architecture, and clear data definitions will aid in preserving the integrity of master data while allowing more volatile data sources to combine for analytical purposes.
Discuss data management practices that enable the data warehouse

1.2.2

1. Having reviewed the fundamental aspects of fundamental data management practices, organize a roundtable session with the foundation project team and senior IT stakeholders.

2. To facilitate an open dialogue, pose the prompting questions below to the meeting participants.

3. When there is consensus on data management components that are lacking, document the gap in the Data Warehouse Foundation Project Plan Template.

Metadata Management:
- Is metadata currently captured for key data assets?
- Does a comprehensive business data glossary exist?
- Are there defined processes in place for changes and updates to metadata?
- Do we have standardized taxonomies and nomenclature for both internal and external data assets of varying structure types?

Reference and Master Data Management:
- Does documentation of universal master domains exist? Do our governance policies reflect the importance of these master data sets?
- Are reference and master data accurately replicated and distributed across the organization?
- Can we solve some issues with consistent reference data by purchasing an external data feed?

Input
- Current state of data management practices

Output
- Pre-implementation improvement activities

Materials
- Data Management Assessment and Planning Tool
- Data Warehouse Foundation Project Plan Template

Participants
- CIO
- SVP data management
- The data warehouse foundation team
Define current and future states for data warehouse-enabling practices

1. With the data warehouse foundation team and individuals responsible for data management (i.e. data stewards) in one room, complete tabs 4 and 6 of the Data Management Assessment and Planning Tool for the relevant data management practices:
   - Data Quality Management
   - Data Governance
   - Reference and Master Data Management
   - Metadata Management

Use the following legend to guide your scoring of each applicable capability:

Assessment Legend (Based on CMMI)

- 1 = Initial/Ad hoc
- 2 = Developing
- 3 = Defined
- 4 = Managed and Measureable
- 5 = Optimized

2. Analyze and debrief the results:
   Review the individual capability findings and use the average maturity level results to guide your team in having a comprehensive understanding of the current state of your data management practice.

Download the Data Management Assessment and Planning Tool.

30-45 minutes

INPUT
- Current state of data management practices

OUTPUT
- Pre-implementation improvement activities

Materials
- Data Management Assessment and Planning Tool
- Data Warehouse Foundation Project Plan Template

Participants
- CIO
- SVP data management
- The data warehouse foundation team
Create the optimal environment for your data warehouse program by establishing a **data management practice**

Data warehousing and business intelligence focus on making your data assets available to your business users – this is the **end stage** in the data management journey.

Before your organization can effectively drive decision making with data-derived insight, you need to determine how the business and IT will work together to establish a **data management practice** that enables that objective.

Download our [Create a Plan for Establishing a Business-Aligned Data Management Practice](#) blueprint.
Your data warehouse program starts with a well-articulated charter

Data warehousing needs to be treated as a continuous program to be successful. Articulating and documenting overarching statements for what needs to be accomplished will serve as the guidepost for subsequent project efforts.

“If there is no leadership that is continually talking about the change and promoting it, the project will fall by the wayside. There also needs to be a commitment to the change in practice that is exhibited by all the leaders in the organization.”

– Liselle Ramcharan
Project Manager, TD Insurance
If you want additional support, have our analysts guide you through this phase as part of an Info-Tech workshop.

Book a workshop with our Info-Tech analysts:

- To accelerate this project, engage your IT team in an Info-Tech workshop with an Info-Tech analyst team.
- Info-Tech analysts will join you and your team onsite at your location or welcome you to Info-Tech’s historic Toronto office to participate in an innovative onsite workshop.
- Contact your account manager (www.infotech.com/account), or email Workshops@InfoTech.com for more information.

The following are sample activities that will be conducted by Info-Tech analysts with your team:

1.1.1-1.1.4

Structure the data warehouse foundation project
Our analysts will work with your project team to appropriately define the roles and responsibilities for the individuals involved, as well as clearly articulate what success looks like in your environment and the risks associated with this initiative.

1.2.1-1.2.3

Review the impacts of data management practices
The analysts will facilitate a discussion that will appraise the current state of your overall data management practices to pinpoint gaps that can derail a data warehouse project.
PHASE 2

Establish the Business Drivers and Data Warehouse Strategy

Build an Extensible Data Warehouse Foundation
Phase 2 outline

Complete these steps on your own, or call us to complete a guided implementation. A guided implementation is a series of 2-3 advisory calls that help you execute each phase of a project. They are included in most advisory memberships.

Guided Implementation 2: Establish the Business Drivers and Data Warehouse Strategy
Proposed Time to Completion: 4-5 weeks

**Step 2.1: Identify Essential Business Processes to Support**

Start with an analyst kick off call:
- Discuss your vision and overall goals for the data warehouse program.
- Discuss the levels of business activities that the data warehouse can enable.
- Discuss which business units would be ideal to pilot the foundation project with.

Then complete these activities…
- Elicit the high-level business needs.
- Create data warehouse vision, mission, goals, and objectives.
- Document the three types of business activities that will be the focus of the foundation project.

With these tools & templates:
- Data Warehouse Program Charter Template
- Data Warehouse Foundation Project Plan Template

**Step 2.2: Create the Core Data Model**

Review findings with analyst:
- Review the target business processes.
- Review Info-Tech’s four-column model for connecting business processes to data elements.
- Discuss the important of the four “Rosetta Stone” elements for data modeling.

Then complete these activities…
- Create the foundation for your data model using the four Rosetta Stone views.
- Define the needs for history tracking in the data warehouse.

With these tools & templates:
- Business Data Glossary
- Data Classification Inventory Tool
- Master Data Mapping Tool

Phase 2 Results & Insights:
- Don’t jump ahead to requirements; develop a deep understanding of business processes at the operational, tactical, and ad hoc levels.
- Prevent your data warehouse from becoming a simple dumping ground for corporate data by using the “Rosetta Stones” to create a robust data model.
## Phase 2 outline (continued)

Complete these steps on your own, or call us to complete a guided implementation. A guided implementation is a series of 2-3 advisory calls that help you execute each phase of a project. They are included in most advisory memberships.

### Guided Implementation 2: Establish the Business Drivers and Data Warehouse Strategy

**Proposed Time to Completion: 4-5 weeks**

<table>
<thead>
<tr>
<th>Step 2.3: Create an Architectural Strategy</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Start with an analyst kick off call:</strong></td>
</tr>
<tr>
<td>• Review the business processes that are targets for the foundation project.</td>
</tr>
<tr>
<td>• Discuss the different models for data delivery, outside of data warehouse technology.</td>
</tr>
<tr>
<td><strong>Then complete these activities...</strong></td>
</tr>
<tr>
<td>• Gather project team and stakeholders to discuss what a data warehouse is and is not.</td>
</tr>
<tr>
<td>• Distinguish between ETL and ESB needs.</td>
</tr>
<tr>
<td>• Complete the questions in the <em>Data Warehouse Architecture Planning Tool</em>.</td>
</tr>
<tr>
<td>• Create a fit-for-purpose architecture model.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Step 2.4: Explore Technology Options</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Review findings with analyst:</strong></td>
</tr>
<tr>
<td>• Review identified project risks, as well as challenges in defining the data warehouse foundation team.</td>
</tr>
<tr>
<td>• Discuss the data management practices that are most critical to successful data warehousing.</td>
</tr>
<tr>
<td><strong>Then complete these activities...</strong></td>
</tr>
<tr>
<td>• Gather the project team to discuss current data management practices.</td>
</tr>
<tr>
<td>• Define current and future state for data management practices.</td>
</tr>
</tbody>
</table>

**With these tools & templates:**

- Data Warehouse Architecture Planning Tool
- Data Warehouse Foundation Project Plan Template

Call 1-888-670-8889 or email GuidedImplementations@InfoTech.com for more information.
Phase 2, Step 1: Identify Essential Business Processes to Support

This step will walk you through the following activities:

2.1.1 Understand the most fundamental needs of the business.
2.1.2 Identify the data warehouse vision, mission, purpose, and goals.
2.1.3-2.1.6 Detail the most important operational, tactical, and ad hoc activities the data warehouse should support.
2.1.7 Link the processes that will be central to the data warehouse foundation.

This step involves the following participants:

- Business stakeholders
- IT representatives

Outcomes of this step:

- Articulate the high-level strategic goals as a guidepost for subsequent planning.
- In-depth understanding of overall business objectives and the most important processes that the data warehouse should enable.
Realize data warehouse benefits by making the right decisions quicker with holistic, high-quality data

Functionalities provided by a data warehouse help the business to make better, timely, and comprehensive decisions based on shared data to improve customer service and the bottom line. The data warehouse also helps IT to better manage information and serve the business.

<table>
<thead>
<tr>
<th>What a Data Warehouse Brings</th>
<th>Business Impacts Created</th>
</tr>
</thead>
<tbody>
<tr>
<td>Efficient</td>
<td></td>
</tr>
<tr>
<td>Reduce time to information</td>
<td>Respond to business needs faster to enable quicker decisions, creating a more agile business operation.</td>
</tr>
<tr>
<td>Reduce time to run analytical queries/data mining</td>
<td>Users get their information at the right time, driving user adoption.</td>
</tr>
<tr>
<td>Effective</td>
<td></td>
</tr>
<tr>
<td>Trusted queries and reports to decision makers, reducing reliance on spreadsheets</td>
<td>Data warehouse serves as a centralized data repository, reducing the proliferation of shadow IT.</td>
</tr>
<tr>
<td>Analytics using high-quality data</td>
<td>Accurate decisions are made with clean and updated data, driving decision accuracy and relevance.</td>
</tr>
<tr>
<td>Comprehensive</td>
<td></td>
</tr>
<tr>
<td>Analytics with integrated data</td>
<td>Comprehensive customer profiling, improving relevant customer engagement and generating opportunities for cross- and upselling.</td>
</tr>
<tr>
<td>Analytics with historical data</td>
<td>Historical data can be mined to find patterns and trends to enhance future planning and forecasting, identifying new business opportunities.</td>
</tr>
</tbody>
</table>
Building a data warehouse program should be aligned with addressing **frustrations, concerns, and opportunities**

Moving a data warehouse from a technology-focused IT project to a sustainable program relies on business users who are genuinely engaged. Business drivers should be the root of the program and inform each step of the process.

**Three ways to approach the business:**

Asking questions with these frames in mind will elicit an emotional response from users that will ultimately get the project team closer to what business users *really* want.

Vendors are knocking on your door to get business from the business units. They will sell and implement department-focused applications for the business units without much enterprise planning in mind. Change the game by working with the business to understand their needs and address them with a one-stop data store – the data warehouse.
Understand the fundamental needs of the business

2.1.1

With each business representative in the room, have them answer each of the following questions on a sticky note:

1. a) What is the biggest opportunity that you want to embrace?
   b) What is the most significant risk that you want to alleviate or avoid?
   c) What is the most persistent frustration that you want to eliminate?

2. Have the activity facilitator read out the opportunities, risks, and frustrations; look for overlap in the suggestions as well as implicit agreement from the participants. This will help to hone in on a mission-critical objective that the data warehouse needs to support.

3. Select one or two points from each group of sticky notes and denote them as top priorities.

4. Document these outputs using the Data Warehouse Program Charter Template.

INPUT

- Current business user environment

OUTPUT

- Consolidation of business hopes, frustrations, and concerns

Materials

- Whiteboard
- Sticky notes
- Data Warehouse Program Charter Template

Participants

- Business stakeholders
Use these vision-statement starting ideas to craft your own

Having a **vision** at the front and center of your data warehouse program helps to communicate the critical importance of data warehousing. A clear vision also establishes a unified business-IT understanding of the 3-5 year high-level plan and the objectives of data warehousing in the organization.

**Some starting points include:**

- The organization is dedicated to creating a data warehouse program that consists of an enabling structure that helps the organization get the *right information*, to the *right people*, at the *right time*.

- To accurately interpret and analyze data from different systems with confidence that data attributes and definitions conform to enterprise standards established through data governance policies and procedures.

> **“All companies are now **data companies**, especially in the financial sector. Anything data-related that you do now should be crafted to fit into an **overall vision.**”**

– Randy Piscione
Enterprise Data Architect, BMO Financial Group
Use these mission-statement starting ideas to craft your own

Having a mission statement for your data warehouse program outlines the “whats” of data warehousing: what it will accomplish, what it does for the business, and what it will do to improve the organization’s strategic use of its data.

Some starting examples include:

• Identify the decision-making power for policies and procedures that will affect business data.
• Align data initiatives with corporate strategies to promote consistent organizational goals.
• Effectively manage and maintain data resources, ensuring the integrity, reliability, availability, and compliance of organizational data and information.
• Decrease the risk of data misuse through the implementation of policies and procedures.
• Increase adherence to regulatory requirements and improve the overall security of organizational data as it stored, shared, and used throughout the organization.
• Provide ongoing structure to ensure adherence to data governance.
• Turn data into information that supports the organization’s…
Identify the data warehouse program vision, mission, purpose, and goals

To avoid any ambiguity about what the data warehouse program is meant to entail, a detailed discussion on what needs to be accomplished and socialized is critical. This is the purpose of the charter – to clearly state and socialize with the business the vision, mission, purpose, and goals of the program.

<table>
<thead>
<tr>
<th>Talk</th>
<th>Walk</th>
</tr>
</thead>
<tbody>
<tr>
<td>Vision</td>
<td>3-5 year view</td>
</tr>
<tr>
<td>Mission</td>
<td>What</td>
</tr>
<tr>
<td>Values</td>
<td>How</td>
</tr>
</tbody>
</table>

Don’t develop these crucial statements and goals in isolation. Ideally, you would collaborate with the business to come up with and update the statements during multiple rounds of review.

**INPUT**
- Business input and drivers

**OUTPUT**
- Clear vision, mission, purpose, and goals of the program

**Materials**
- Data Warehouse Program Charter Template

**Participants**
- IT representatives
- Business leaders
Create **guiding principles** as a pillar of your data warehouse program

Data modeling principles can be developed to provide a framework for future data architecture works. The principles are a good tradeoff between no data architecture governance and too much control with very defined data modeling workflows and approvals.

Use an industry-renowned enterprise architecture framework such as [TOGAF, which has enterprise architecture principles](https://www.info-techresearchgroup.com) that are easily transferrable to the data warehouse environment.

<table>
<thead>
<tr>
<th>Category</th>
<th>Principle</th>
<th>Manifestation</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Corporate</strong></td>
<td>Data is a corporate asset</td>
<td>All employees are stakeholders in producing and maintaining good data.</td>
</tr>
<tr>
<td></td>
<td>Data is shared</td>
<td>Create an enterprise view of data instead of a siloed view.</td>
</tr>
<tr>
<td><strong>Data Quality</strong></td>
<td>Right data</td>
<td>Definition of “good data” is informed by business processes.</td>
</tr>
<tr>
<td></td>
<td>Assess, improve, monitor</td>
<td>Data quality is on-going. Iterate through assess, improve, and monitor.</td>
</tr>
<tr>
<td><strong>Data Governance</strong></td>
<td>Data ownership</td>
<td>There is traceable accountability for data that is generated and consumed.</td>
</tr>
<tr>
<td></td>
<td>Data stewardship</td>
<td>Identified individuals are responsible for maintaining the overall integrity of data assets in the data warehouse.</td>
</tr>
<tr>
<td><strong>Master/Reference</strong></td>
<td>One version of the truth</td>
<td>Data siloes need to be federated and business definitions need to be documented in order to create a true enterprise-wide view.</td>
</tr>
<tr>
<td><strong>Data Management</strong></td>
<td>Compliance with law</td>
<td>Data policies are reflective of changes in compliance requirements.</td>
</tr>
<tr>
<td></td>
<td>Balance between accessibility and restriction</td>
<td>Data classification tables are well defined to ensure data integrity in the data warehouse while still enabling the business.</td>
</tr>
</tbody>
</table>
Define and refine guiding principles for the data warehouse

2.1.3

30 minutes

Having some high-level guiding principles is essential to the data warehouse program. Those guiding principles should be consulted before major decisions are made.

1

The DW center of excellence (CoE) should direct these activities. Brainstorm some guiding principles based on these TOGAF categories. Develop additional principle categories if needed.

<table>
<thead>
<tr>
<th>TOGAF Guiding Principle Categories</th>
</tr>
</thead>
<tbody>
<tr>
<td>Corporate</td>
</tr>
<tr>
<td>Data Quality</td>
</tr>
<tr>
<td>Data Governance</td>
</tr>
<tr>
<td>Master/Reference Data Management</td>
</tr>
<tr>
<td>Data Security</td>
</tr>
</tbody>
</table>

2

Document the data warehouse guiding principles in the Data Warehouse Program Charter Template. Create posters or handouts so that involved professionals can make use of the guiding principles when the time comes.

Info-Tech Insight

Take a lean approach to guiding principle development. Leave no room for interpretation of those guiding principles – be clear, be realistic, and support the business.
Ensure your data delivery model can support **three categories** of business activities

A well-architected data warehouse will have components that facilitate **drilling down** and **rolling up** to provide granular or holistic views of information.

<table>
<thead>
<tr>
<th>Operational Activities</th>
<th>Tactical Activities</th>
<th>Ad Hoc Activities</th>
</tr>
</thead>
<tbody>
<tr>
<td>Are consistent day-to-day activities occurring on the frontlines of the business.</td>
<td>Take place to enable more informed decision making on a consistent basis, in line with corporate mandate.</td>
<td>Take place with varying regularity; usually drive provisional decision making or creating future strategies.</td>
</tr>
</tbody>
</table>

These processes enable one or more of Harvard Business Review’s “**value-added disciplines**.”

**Info-Tech Insight**

A data warehouse should simplify data delivery, but can only do so if a thorough understanding of the business processes exists. Creating a comprehensive list of the different activities will help to identify common data needs.
Capture your key operational and tactical processes

Operational processes are the everyday activities that make up the core of your business.

At the operational level, the focus is on how data is used in your organization within the applications around the following functions:

Tactical processes bridge the gap between the day-to-day activities and the overarching strategic goals.

With the rationale and motivations provided by the strategy, tactical planning focuses on putting things into action and producing tangible results.
Develop descriptions of your operational processes

2.1.4

1. Think of two to three day-to-day transactions or events that drive the operations of your organization.

2. Identify the systems that support those transactions.

3. Name the operational processes that are occurring within these systems.
   - Is the process **manual** or **automated**?

4. List the data that is linked to these systems.
   - Cultural dynamics (with respect to data ownership, data trust/usage, data sharing, etc.).
   - Specific policies associated with operational data use.
   - Any audit and compliance requirements associated with the data (i.e. re-creation of data lineages).

5. Document the outputs in the *Data Warehouse Foundation Project Plan Template*.

**60 minutes**

**INPUT**

- Day-to-day operational activities

**OUTPUT**

- Description of activities and improved understanding of current data assets

**Materials**

- *Data Warehouse Foundation Project Plan Template*

**Participants**

- Business stakeholders/process owners
- Project manager
- CIO
Develop descriptions of your operational processes (continued)

### 2.1.4

<table>
<thead>
<tr>
<th>Transaction/Event</th>
<th>Systems</th>
<th>Operational Process</th>
<th>Data Attributes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Customer makes a purchase from an online store</td>
<td>Inventory/Order Mgmt.</td>
<td>• Validate item is in stock (automated)</td>
<td>• SKU/InventoryID • ProductName</td>
</tr>
<tr>
<td></td>
<td>CRM</td>
<td>• Update inventory count (automated)</td>
<td>• CustomerID • CustomerName • Customer_Email • TransactionID • TransactionDate</td>
</tr>
<tr>
<td></td>
<td>Finance/Billing</td>
<td>• Submit approved purchase for pick and delivery</td>
<td>• TransactionID • TransactionDate • TransactionSubtotal</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Create new customer account (automated)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Update customer purchasing history (automated)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Update accounts receivable (automated)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Process payment</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Generate e-receipt (automated)</td>
<td></td>
</tr>
</tbody>
</table>

Time: 60 minutes
Outline your tactical activities

2.1.5

Tactical activities often involve **system-to-system integration** that is fulfilled through an automated or manual business process in order to satisfy data requirements that **go beyond a single source application**. Identifying these activities will begin to inform the design for a better solution that meets data needs.

1. Before listing tactical activities, revisit your current business strategy and focus. What are your current and repeated decision-making needs?

2. Review the standard reports that are currently sourced from operational activities and data, as well as reports that you are looking to create with better access to data.

<table>
<thead>
<tr>
<th>Customer purchasing patterns</th>
<th>Monitoring product performance over time</th>
</tr>
</thead>
<tbody>
<tr>
<td>Customer retention</td>
<td>Profitability breakdown (identifying key revenue streams)</td>
</tr>
<tr>
<td>Cost reduction opportunities (optimizing purchasing/suppliers)</td>
<td>Marketing campaign effectiveness</td>
</tr>
<tr>
<td>Market segmentation (by customer demographics)</td>
<td>Time-based transaction volume analysis</td>
</tr>
</tbody>
</table>

3. Review the process or processes for how access to multiple data sources to support an activity is approved or enabled.
   - Consider at what stages and in what capacity data owners, data stewards, and information security teams are involved.

4. Document the outputs in the *Data Warehouse Foundation Project Plan Template*. 

**INPUT**
- Current tactical activities
- Desired tactical reports

**OUTPUT**
- Description of tactical activities for later reference

**Materials**
- Data Warehouse Foundation Project Plan Template

**Participants**
- Business stakeholders/process owners
- Project manager
- CIO
Understand and address ad hoc processes to improve data warehouse delivery

Ad hoc processes encompass the activities that don’t have the regularity of operational or tactical processes.

Depending on the granularity and scope of the information involved, ad hoc processes can be conducted from front-line/operational or enterprise-wide/strategic perspectives.

Business intelligence/analytics, predictive modeling, and big data blending are considered ad hoc processes. As the frequency of reporting tends to be lower, more time can be dedicated to slicing and dicing, aggregating, or extrapolating data.

The insights that come out of the ability to manipulate data in these ways can influence what becomes the standard for business operations.

In organizations that are continually evolving, ad hoc processes that are consistently codified and repeated are converted into tactical or operational activities.

Info-Tech Insight

Ad hoc analyses, self-service BI, and data blending are closely related. In many situations, ad hoc analyses are fulfilled by the business using some kind of self-service BI. In other situations, the BI team may use a sandbox to blend datasets to fill an ad hoc request.
Detail the most important ad hoc requirements the data warehouse should support

This discussion is focused on the **data-related wants and opportunities** business users are hoping to embrace on the back of an improved data delivery model. In addition, ad hoc reporting can be **triggered in response to external events** that require the business to adapt quickly.

1. Gather business users from relevant business units.

2. Starting with current ad hoc activities, answer the following questions:
   - What is the activity and what is its frequency?
   - What is the purpose for it?
   - Is it mission critical, and why?

3. After detailing the existing ad hoc processes, answer the same questions for reporting that does not currently exist but would be useful in the future.
   - Consider improvements in the areas of predictive modeling, business intelligence/business analytics, and blending internal data with big data.

4. Document these processes and the accompanying descriptions in the **Data Warehouse Foundation Project Plan Template**.
Detail the most important ad hoc requirements the data warehouse should support (continued)

2.1.6

### Ad hoc process examples

<table>
<thead>
<tr>
<th>Process</th>
</tr>
</thead>
<tbody>
<tr>
<td>Overlaying customer sentiment data with internal product defect/failure reports.</td>
</tr>
<tr>
<td>Analyzing changes in product/service consumption patterns in response to geopolitical events or natural disasters.</td>
</tr>
<tr>
<td>Federation of all raw and derived data plus analytic algorithms and big data into a governed data warehouse environment.</td>
</tr>
<tr>
<td>Analyzing customer sentiment data in preparation for investing in a new product line.</td>
</tr>
</tbody>
</table>

### Info-Tech Tip

In addition to thinking about the “art of the possible” with respect to new data sources and reporting/analytics, consider whether a federated or self-service model for data delivery would work best in your organization.
Link and prioritize the processes that will be central to the data warehouse foundation

### 2.1.7

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
</tbody>
</table>

A data warehouse program isn’t built in one go. The foundation needs to have a defined scope to address the most business-critical processes, and later iterations will expand the versatility of services provided.

1. Gather the operational, tactical, and ad hoc processes ideated in previous activities.

2. Align the processes at each level based on the overall goals and objectives that they support.
   - Be sure to connect them to the goals and objectives created in the previous activities.

3. Identify a set of processes that links well to an overall objective and touches on important activities in a key business unit or incorporates a number of units.

4. Prioritize based on business value, effort, and urgency.

Document the other process flows as projects to be completed after the foundation is established using the *Data Warehouse Foundation Project Plan Template*.

**INPUT**
- Operational, tactical, and ad hoc activities previously defined

**OUTPUT**
- Consolidated list of business activities

**Materials**
- Whiteboard
- Sticky notes
- *Data Warehouse Foundation Project Plan Template*

**Participants**
- Business stakeholders/process owners
- Project manager
- CIO
Align business activities for the data warehouse

Document these groupings in the *Data Warehouse Foundation Project Plan Template.*
Phase 2, Step 2: Create the Core Data Model

This step will walk you through the following activities:

2.2.1 Walk through the four-column model as a starting point for data modeling.
2.2.2 Model business entities at a high level.
2.2.3 Use the business data glossary data classification to develop a data model.
2.2.4 Identify master data elements to define dimensions.
2.2.5 Set up lookup tables based on reference data.
2.2.6 Define requirements for history tracking in the data warehouse.
2.2.7 Define current and real-time needs in the data warehouse.

This step involves the following participants:

• Business Stakeholders (from units of targeted processes)
• Data Warehouse Architects
• Integration Specialists
• CIO
• Project Manager

Outcomes of this step

• Foundation for the core tables that the data warehouse will be built around.
• Understanding of in-depth characterization of data up-front facilitates easier data modeling.
Use Info-Tech’s four-column model as a starting point for data modeling

Data modeling in the data warehouse is ultimately reliant on understanding how data originates, flows, and is used throughout the organization.

Detailing the data elements that support particular business processes will inform the relationships created between data tables in the data warehouse, as well as highlight points where other methods of integration should be used to augment connectivity.
Use the four-column model to start data modeling

2.2.1
Leverage the four-column model to organize the data, its flow, the systems that it touches, and the business use of the data to understand more about your existing data. With this fundamental knowledge, you will better architect and construct your model in the data warehouse.

1. Schedule business unit interviews, capture the four key components of the business unit's data, and map how they interact with each other:
   a) Start with a **Business Unit**. Schedule interviews or working sessions with that particular business unit to go over the four-column model.
   b) Populate the **Usage of the Data** column. Discuss the usage of the data at the business-unit level. (This can filled in with the content from the *Data Warehouse Foundation Project Plan Template*.
   c) Identify the **Systems** the data is housed in for the data usage identified above.
   d) Information and insight captured can then be used to populate the **Business Data Glossary** column, which will be used to generate the business data glossary.

2. Translate some of the data understanding into granular functional, data, and technical requirements. The requirements will inform later data modeling and technological planning.

**INPUT**
- Existing data flow diagram and data dictionary
- Process documentation from *Data Warehouse Foundation Project Plan Template*

**OUTPUT**
- The four-column model of data flow

**Materials**
- Whiteboard
- Markets

**Participants**
- Data warehouse architects
- BAs
- Business units
Use business entity modeling to highlight the relationships between business units and the data they use

The foundational data architecture will include a high-level business entity map (BEM). The BEM contains fundamental information about the business entities and how they are related. With the BEM, future modeling is accelerated as business knowledge is well kept and maintained as a corporate history.

Developing a BEM is a four-step refinement process. First, “things” in the corporation are identified and documented. The relationships among the entities are formulated. Attributes of the entities are collected and recorded. Finally, entities are organized into hierarchies.

Define “Things” Define Relationships Define Attributes Define Hierarchy

People

Product

Account

Place

Event

1..1 One-to-one relationship

1..M One-to-many relationship

M..M Many-to-many relationship

M or O Mandatory or optional

Account Number
Account Holder
Account Start Date
Payment Method
Cycle Date
Invoice Method
Account Limit
Account Language

Region
State
District
Dealer

The business entity map, for the most part, remains unchanged unless there are significant events such as a major shift in product/service offerings and/or a merger and acquisition. Even a source system replacement may not change BEM much. The business entity map serves as a foundational data understanding for you to build the data model on.
Model business entities at a high level

Instructions:

1. Schedule a working session with the data architect, data analysts, and business analysts.
2. Identify all the business entities (moving parts) of the organization.
3. Discuss how the business entities are associated with each other.
4. Define the attributes of the business entities (data points that describe the business entities).
5. Organize the entities into hierarchies.
6. For anything (business entities, relationships, attributes) that is uncertain at the moment, document these in a “parking lot” area. Additional inquiries and/or discussion may be needed.
7. Verify the business entity map with the business units to make sure the map represents the actual business model.

INPUT
- Existing documentation around business organization/architecture

OUTPUT
- High-level business entity maps

Materials
- Whiteboard
- Markers
- Data Warehouse Foundation Project Plan Template

Participants
- Business analysts
- Data architect
- Data analysts
Leverage the **Rosetta Stones** to land incoming data in the data warehouse

### Build multiple views of your data assets
For almost all data warehouse implementations, there are four core pieces of data reside in the data warehouse. They are considered to be the “Rosetta Stones” of a data warehouse. They provide the backbone for the data warehouse and act as a window to land upstream data in the data warehouse.

---

**Business Data Glossary**  
Provides definitions of key business terms. It is used to understand the data and its meaning to better model a data warehouse.

**Master Data Domain**  
Informs dimensional design. Master data is usually the moving parts that a company relies on. They are good dimension candidates.

**Data Classification**  
Informs the sensitivity of the data. The classification outlines which security level a data element should assume.

**Reference Data Element**  
Provides guidance to structure look-up tables. Reference data usually supplements the dimensions in the data warehouse.
Use your business process documentation and four-column model to complete the business data glossary

The business data glossary (BDG) serves as a repository for business definitions of data items. The attributes for each data item are also documented in the BDG.

In data modeling, the business data glossary:

- Supports understanding of the business entities in your organization (what they are and how they are related).
- Documents attributes associated with the specific business entities. Eventually the attributes will turn into data fields in the data warehouse tables.
- Identifies the source(s) of the data used by the business entities.
- Serves as metadata to provide definitions of terms, including the business-unit variations in definitions for similar terms (customer, account, etc.).

Download Info-Tech’s Business Data Glossary.
Account for data warehouse security by creating a data classification scheme

Use this tool to classify data items into security classification categories based on governance, risk, regulatory, and compliance requirements.

Classification Examples:

- A transaction for a corporate-level customer that is over $100,000 in value is considered to be a semi-classified transaction. Only senior-manager level employees can have access to it.

- Credit scores for the customers can be used by many warehouse users. However, detailed activities in the credit reports are restricted and they can be accessed by the credit department only.

Usage in the data warehouse:

- To tag data items in the warehouse to enable row-level security.

- To set up role-based security models in data warehouse and/or BI.

- To drive retention policy and inform multi-tiered storage.

- To search sensitive data in case of a legal proceeding.
Use the business data glossary and data classification scheme to develop a data model

2.2.3

The business data glossary and any existing data classification documentations are very useful in describing the nature and the characteristics of the data elements. Leverage the two to effectively model your data.

1. Review the business data glossary to familiarize yourself with the definition for each data element, especially these fields:

<table>
<thead>
<tr>
<th>Field in the Business Data Glossary</th>
<th>Data Modeling Implications</th>
</tr>
</thead>
<tbody>
<tr>
<td>Definition</td>
<td>Tells you “what” the data element is all about.</td>
</tr>
<tr>
<td>Possible Values</td>
<td>Informs the data type and the data distribution.</td>
</tr>
<tr>
<td>Data Sensitivity and Major Stakeholders</td>
<td>Informs content-level security.</td>
</tr>
<tr>
<td>Retention</td>
<td>Informs slowly changing dimension and historical tracking requirements.</td>
</tr>
<tr>
<td>Data Steward</td>
<td>Informs the contact person(s) when clarification is needed.</td>
</tr>
</tbody>
</table>

2. The classification of the data warehouse data can be informed by the data classification artifact with some modification. A data warehouse classification might be more stringent as source data has moved up the data value chain since the data is more integrated, organized, and refined. Discuss the classification needs in terms of the data value chain.

Base-line, source data classification requirements

<table>
<thead>
<tr>
<th>Data Value Chain</th>
<th>Stringent, data warehouse classification requirements</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data is integrated?</td>
<td>Data is organized?</td>
</tr>
</tbody>
</table>

INPUT
- Business data glossary
- Data sensitivities based on business environment

OUTPUT
- Classification schema for the data warehouse

Materials
- Whiteboard
- Markers
- Data Classification Inventory Tool

Participants
- Data warehouse architects
- BAs
- Information security representative
Defining **master data domains** will shape data warehouse dimensions

**Master data** “represents the business objects which are agreed on and shared across the enterprise” (IBM, 2008). Data assets that have both business **value** and **risk** will fall under master data classification.

### Master Data Examples:

- **Customer master.** This is master data that is sourced from multiple data sources (billing records, contact records, online login, etc.). The sources are consolidated and integrated to create a single record for the customer, often with history.

- **Product master.** Product master may be sourced from product codes used by internal departments and external partners. Product master may be used to share product information with partners, suppliers, and distributors. Product master may have a hierarchy associated with it that defines product categories and subcategories.

### Benefits of defining master data during data modeling:

- Give insights and intelligence of the dimensions and hierarchy.
- Serve as data sources for the dimension.
- Provide dimensional attributes for the data warehouse.
- Reuse consolidation and integration logic.

Identifying master data elements during data modeling will aid in creating consistent and reliable views in data marts and cubes that are sourced from data warehouse content.
Track how your current systems are informing your master data domains

Overview

Use this tool to define your master data domains, the data elements/attributes that are linked to the domains, and how your systems and databases are interacting with the domains.

The template contains three common master domains (party, product, and place) and sample data elements that inform the master record on tabs 2-4.

As you identify more master data domains, copy the blank sample table on tab 5 and fill it with content relevant to your organization.

Download Info-Tech’s Master Data Mapping Tool.
Identify master data domains to define dimensions

2.2.4

Review your master data template to create dimensions in your warehouse, and determine the sources and attributes of the dimensions.

1. Review the master data template to identify business entities that need to be mastered. Those business entities are candidates for data warehouse dimensions. Determine which master data entities should be created as dimensions.

2. Review the attributes of the identified dimensions. Determine if those attributes are strictly attributes or if some of them are metrics.

<table>
<thead>
<tr>
<th>Attribute</th>
<th>Dimension?</th>
<th>Metric?</th>
</tr>
</thead>
<tbody>
<tr>
<td>Customer Name</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>Customer Tenure</td>
<td>No</td>
<td>Yes</td>
</tr>
</tbody>
</table>

3. Brainstorm if the attributes suggested by the master data template are sufficient to provide information about the master data. If not, define additional master data attributes to be included.

4. Larger master data may need to be denormalized or organized into a hierarchy. Discuss the needs for denormalization and determine if the dimension is part of a hierarchy.

Info-Tech Insight

Distinguish between master data in master data management (MDM) and in a data warehouse. MDM data is highly operational in nature and is often focused in one or two subjects whereas the data warehouse’s data is often cross-functional and analytical, not operational, in nature.

Input

- Master Data Mapping Tool

Output

- A list of dimensions
- Dimension attributes
- Metrics

Materials

- Whiteboard
- Markers
- Sticky notes
- Data Warehouse Foundation Project Plan Template

Participants

- DW foundation team
- SMEs if necessary
Create consistency in data warehouse table values by incorporating reference data

**Reference data** is used to create a list of permissible values and attached textual descriptions. It serves as an important data quality control point by minimizing the input of invalid data or duplication of records through the use of non-standardized codes.

**Use reference data sets to create look-up tables for the following fields:**

- ZIP/postal code formats
- Currency conversion
- Holidays
- Customer status/classification
- Gender
- Title prefixes
- Role/position titles

In many ways, dimensions (master data) and reference tables are closely related. However:

- The distinction is that reference data is based in defining values and providing conversion/translation, but does not change frequently.
- Dimensions are more dynamic in nature and they have a lot more attributes.

When reference data is neglected...

Within their customer management database, an organization had “IBM” stored in **13 different formats** in the “company” field, such as:

- IBM
- I.B.M.
- IBM Corporation
- I.B.M. Corp.
- International Business Machines

You can often use industry standards like **ISO, SIC codes, time zones, and currency codes** as templates to build your reference data. You can also enable historical tracking in reference tables (e.g. changes in customer status).

**Info-Tech Insight**

**Tap into the value of external data.** Consider leveraging commercial data providers such as Dunn Bradstreet, Experian, Nielson, and ISBN as a consistent source for updating your reference data. Blending external data and internal data may not be an easy task due to the absence of common keys. Data matching tools may be needed to fuzzy match the datasets.
Set up lookup tables based on reference data

2.2.5

Standardized values are best managed by lookup tables. Leverage this exercise to determine what reference data to be managed.

1. Review the reference data template and classify the reference data into two categories: internal reference data and external reference data.

<table>
<thead>
<tr>
<th>Internal Reference Data</th>
<th>External Reference Data</th>
</tr>
</thead>
<tbody>
<tr>
<td>Status Codes</td>
<td>ZIP Code</td>
</tr>
<tr>
<td>Cost Centers</td>
<td>SIC Industry Codes</td>
</tr>
</tbody>
</table>

2. For internal reference data, invite the affected business units to harvest the potential values on the reference data.

3. For external reference data, explore if some external standards can be used to seed those data points.

4. Document the reference data in the data warehouse plan, and later on the data modelers and architects can use that to model and populate the lookup tables.

Info-Tech Insight

Enrich your lookup tables to add value to the data. You can always attach additional attributes to the lookup tables to provide more insight. For example, you can add income data, population statistics, or latitude/longitude to the ZIP code table to enable more analysis at the ZIP-code level.

INPUT

- Reference data in your organization

OUTPUT

- A list of lookup tables
- Potential values in those lookup tables

Materials

- Whiteboard
- Markers
- Sticky notes
- Data Warehouse Foundation Plan

Participants

- DW core team
- SMEs if necessary
Model historical data to empower point-in-time analysis

Successful data warehouses have **historical data to support time series and trending analyses**. Define requirements for historical data in the foundational data architecture to ensure all future iterations comply to the same standards.

### Possibilities

<table>
<thead>
<tr>
<th>Time Series</th>
<th>Time Comparison</th>
<th>Track Changes</th>
</tr>
</thead>
<tbody>
<tr>
<td><img src="image" alt="Sales Trend" /></td>
<td></td>
<td><img src="image" alt="Eastern Region" /></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>YTD</th>
<th>YTD (last yr.)</th>
<th>YTD (avg. 5 yrs.)</th>
</tr>
</thead>
<tbody>
<tr>
<td>85M</td>
<td>82M</td>
<td>74.5M</td>
</tr>
</tbody>
</table>

Tracking historical data helps you to construct a point-in-time view of the data. You can reconstruct the data snapshot as of a particular date in the past.

Point-in-time is very beneficial for drill-down analyses in which you want to understand all the changes and assess if the changes are causing fluctuations in your metrics.

### Historical Requirement Checklist

- How much historical data do you want?
- What granularity of historical data is required?
- Need historical data for dimensions?
- Need historical data for fact?
- Need historical data for relationships among entities?
- How do you want to track changes?
Deliver real-time (RT) data to support use cases

Apply real-time analytics to empower employees with timely information to help them to make decisions. Some RT analytics can be supported by data warehouses while others require specialized technology that integrates with the data warehouse.

How data warehouses deliver RT information:
- Data warehouses can be run on an **in-memory database platform** to provide faster loading and retrieval. In-memory warehouses can be used to deliver time-sensitive analytics.
- **Operational data stores** integrate with data warehouses so that RT information is combined with enterprise-level information and/or historical information.
- Many data warehouses can deliver **near real-time analytics** by using integration methods such as message queuing, data virtualization, materialized views, and frequent incremental updates.

Real-Time Use Cases

- Trading
- Emergency Services
- Manufacturing Floor
- 24/7 Operation
- Call Center Analytics

Do not enable real-time analytics for your entire data warehouse, as it will **slow it down instead of speed it up**. Identify essential business operations that require time-sensitive analytics and enable RT only for those operations.
Build an integration layer to process and feed source data to the warehouse

Data integration is often called **ETL (extract, transform, and load)** or **ELT (extract, load, and transform)**. The only difference is the location in which the transformation takes place. ETL transforms data in the staging database whereas ELT transforms data in the target database. The main goal for integration is to process the incoming data in a way that is suitable for analytic purposes. It also integrates multiple sources using common keys to create a holistic view.

Data integration is typically run in batch mode, **creating some data latency** in the data warehouse. Depending on your real-time requirements, you may reduce the data latency by **complementing ETL with a change data capture (CDC) solution**. CDC detects and accumulates changes, and ETL can then be used to load the changed data into the data warehouse.

**Extract, Transform, Load (ETL)**

- **Data Source**
- **Batch Changes**
- **ETL**
- **Target: Data Warehouse**

- Support daily or even weekly analytics
- Large incremental changes
- Infrequent updates
- Very sensitive to ETL failures
- Missing an ETL job is like “missing a plane”

**Change Data Capture (CDC)**

- **Data Source**
- **Capture Changes**
- **Change Queue**
- **ETL**
- **Target: Data Warehouse**

- Support near real-time or even real-time analytics
- Small incremental changes
- Very frequent updates
- Less susceptible to ETL failures
- Missing a CDC update is like “missing a subway”
Define requirements for history tracking in the data warehouse

**2.2.6**

Leverage this exercise to scope out historical requirements for your data warehouse.

1. **Brainstorm with your business units to ensure there is a case for tracking and storing history. Consider:**
   - Do the business units need to access historic data from a few years ago? Is the required historical data available in the source systems or is it purged?
   - Is there a need for constructing a point-in-time view of the data? How often are the dimensions changing? How many dimensions change over time?

2. **Determine the needs for slowly changing dimensions (SCDs). There are five major types of SCDs:** *type 1, 2, 3, 4 and 6*. The data warehouse team should discuss the fit for each of the SCD type. Higher SCD levels provide more robust SCD functionalities but require additional modeling effort.

3. **Also determine the needs for storing historical transactions in the data warehouse. The needs will inform requirements around storage space and how to optimize storage cost, performance, and ETL efficiencies. There are factors to consider in optimizing the mix:**
   - How often do the end users access historical data?
   - How much data is stored in the fastest disk? How much data can be stored in slower disk?
   - How frequently does the historical data get updated?
   - Is it possible to use increment load to optimize loading performance?
Define current and real-time needs for the data warehouse

2.2.7
On top of historical data, many data warehouses support some real-time analytics

1. Review your business model with the business units and brainstorm if the real-time use cases are applicable:

   **Real-Time Use Cases**
   - Trading
   - Emergency Services
   - Factory Floor
   - 24/7 Operation
   - Call Center Analytics

2. If one or more of the use cases are applicable, detailed real-time requirements should be established to plan for the right real-time architecture. Consider:
   - What data sets/data elements have to be real-time enabled?
   - What is the volume of data that you want to be real-time enabled?
   - What are timestamp fields to use in order to determine the real-time requirements?
   - How do you define real-time? No delay? Seconds of delay or even minutes?

3. Once the business requirements are defined in the previous steps, the data warehouse project team should reconvene and brainstorm some ideas on setting the right architecture to support real-time. Options are:
   - Change data capture (CDC)
   - Micro-batch ETL
   - Operational reporting generated from ODS

**INPUT**
- Inputs from the business SMEs

**OUTPUT**
- Real-time requirements and architecture

**Materials**
- Whiteboard
- Markers

**Participants**
- Select business users
- Data warehouse architects
Phase 2, Step 3: Create an Architectural Strategy

This step will walk you through the following activities:

2.3.1 Distinguish between ETL and EAI needs.
2.3.2 Create a fit-for-purpose data warehousing architecture model.

This step involves the following participants:

- Business Stakeholders
- IT Representatives

Outcomes of this step

- A high-level data warehouse architecture model informed by business information needs.
- Distinction between how a data warehouse facilitates integrations compared to other integration technologies.
Realize the need for a solid yet flexible data architecture foundation

A foundational data architecture has to be built to provide a foundation for future iterations. Successful data warehouses address enterprise-level concerns, and as a result, the extensible data warehouse leverages a data foundation that forms the basis of enterprise-level planning. With an enterprise focus, future iterations will be built to address more specialized needs while still adhering to an overarching strategy.

A foundational data architecture has to be scalable, comprehensive, and systematic in nature so that it can be used as a cornerstone for future iterations to be built on top.

<table>
<thead>
<tr>
<th>Scalable</th>
<th>Systematic</th>
<th>Comprehensive</th>
</tr>
</thead>
</table>
| • Future-friendly, ready for future additions  
• Integrate with the overall enterprise data architecture  
• High-level understanding of business entities | • Plan for quality data and continuous improvements  
• Boundaries are set to define sensitive information  
• Requirements for historical data are established | • Plan for quality data and continuous improvements  
• Boundaries are set to define sensitive information  
• Requirements for historical data are established |
With an immense variety of data architectural options, identify the most effective pattern for your business situation

Determine the fit for a data warehouse based on:

1. **Business model**
2. **Data landscape**
3. **Use cases**
4. **Resource availability**

By evaluating your organization against these four factors, you will get a clear picture of if you need a data warehouse or an alternative solution to support your information needs.

**Business Model:** Some organizations have **very simple business models** in which they offer simple products/services. The need for integrated information may not be present.

**Data Landscape:** Some organizations have a **very consolidated data landscape**. They are able to manage their data in one or two applications.

**Use Cases:** Some organizations are interested in **operational analytics ONLY**. Operational analytics are best performed with real-time or near real-time data.

**Resource Availability:** Some organizations have smaller IT departments and **simply do not have resources** to implement data warehousing solutions.

**Alternative Solutions**
- Leverage the reporting capability of the production systems.
- Leverage a BI tool to report directly from the production systems or the replicas of the production systems.
- Create reports using Excel and/or Access.
A data warehouse is a special type of database; leverage conventional RDBMS concepts when planning a warehouse.

A data warehouse, regardless of which architecture methodology (Kimball, Immon, data vault, hybrids, etc.) it is modeled with, is still a **database** at its core.

By default, a database stores data and allows read and write operations to be performed by multiple parties. However, the data warehouse does provide some additional capability.

---

**Operational Database**

Often called a transactional DB (OLTP). It mainly supports many write transactions and those transactions are in real-time and smaller in size. Uptime and transaction performance are the keys to transactional databases. They need to be available to support many concurrent transactions.

---

**Data Warehouse**

Often called an analytical DB (OLAP). It is modeled in a way to facilitate a large number of queries. The inputs of an analytics database are very batch-oriented. The data in an analytical DB is usually organized to support business consumption, often via business intelligence and reporting tools.
Understand **ETL** and the **EAI** in relation to the data warehouse

ETL and EAI both help you send data between systems, but the nature of the communication is different.

**EAI**

**Enterprise application integration** (EAI) is a methodology to integrate a set of enterprise applications. It is an integration framework that leverages middleware or some kind of middle-tier platform to enable integration of systems and applications across an enterprise. EAI is usually bi-directional, and real-time in nature. It helps to coordinate communications between applications that use differing data schemas or formats.

Application programming interfaces are the most basic form of EAI as they provide point-to-point connections between applications as needed. Enterprise service bus (ESB) is a family of middleware that facilitates larger scale message coordination and data transformation between many applications and systems.

**ETL**

In computing, **extract, transform, and load** (ETL) refers to processes in data warehousing in which upstream data from source systems is extracted, transformed, and loaded into the data warehouse. ETL is usually unidirectional and focuses on integrating datasets, cleansing data, and re-organizing data, as opposed to allowing applications to communicate in EAI platforms.
Distinguish between ETL needs and EAI needs

Given the similar in nature and functionalities, it is a best practice to spend some time to review your integration objectives to determine which technology to use. You can leverage this sample decision tree to get to right tool.

**Start with your integration objectives**

- **Objectives**

**What are you integrating?**

- **Applications** (E.g. the Salesforce application and Outlook)
- **Databases** (E.g. the Salesforce database with the marketing database)

**What integration approach to use?**

- **Point-to-point integration**
- **Middleware**
- **Move data in batch**
- **Move data in very small increments in real-time**

**Tools to leverage**

- **Point-to-point connections**
- **Enterprise Service Bus (ESB)**
- **ETL**
- **Micro-batch ETL or Change Data Capture (CDC)**

**OUTPUT**

- A clarification on the tools needed to meet your integration needs

**Materials**

- Whiteboard

**Participants**

- CIO
- Business stakeholders
- Data warehouse team (project manager architects, integration specialists)
Architect holistically – your enterprise architecture should inform your data warehouse architecture

Enterprise architecture can be defined as “organizing logic for business processes and IT infrastructure reflecting the integration and standardization requirements of the company’s operating model. The operating model is the desired state of business process integration and business process standardization for delivering goods and services to customers” (Center of Information Systems Research, 2007).

Data warehousing is a small component of enabling a successful enterprise architecture. Other integration strategies, such as enterprise application integration (APIs, enterprise service bus) and service-oriented architecture, work to accomplish the same goal with different approaches.
Initially we thought that we would have everything pumped into the EDW and it would serve as a hub, followed by data being delivered to the different spokes. However, with the burst in technology with cheaper storage and processing, users are looking for quicker turnaround times. Having a single central hub is becoming a bottleneck.

— Jaison Dominic
Lead Architect, Enterprise Data Warehouse, Moffitt Cancer Center

Possibilities can be found in **all levels** of the data warehouse: data source level, data integration level, data organization level, data analysis level, and data delivery level. The possibilities are all about providing **new analytics capabilities**, **improving existing capabilities**, or doing more for less.
Challenge

• Intel IT was initially able to consolidate several data warehouses to create a single source of truth in an enterprise data warehouse (EDW).

• While this brought standardization to its data and created consistency in its analyses, the expansion of its business and the development of big data has exposed its EDW’s inability to handle raw, unstructured, and semi-structured data.

Solution

In addition to keeping the existing EDW, the new data platforms included:

• Apache Hadoop, for raw and unstructured data sets.
• Extreme data warehouse (XDW) for handling large volumes of structured data.
• In-memory databases for real-time processing.
• Custom, independent data warehouses that were informed by specific business needs for analyzing structured data.

Results

• By creating an extensible and diversified model for delivering BI, Intel IT was able to avoid excessive use of the EDW, which produced a cost avoidance of USD $250,000 in the first year alone.

• The determination for which platforms to implement involved an in-depth set of questions for business, which was provided in an automated selection tool that requested information about data source, data type, data latency, availability, and other parameters.
Leverage the operational data store to quickly consolidate transactional data

An operational data store is a storage structure for storing operational data that is needed to support operational analytics and management. An ODS is required to provide a snapshot view in which the data is pulled from transactional applications on a regular basis, often in a real-time nature, to ensure data currency.

An ODS is ideal when the variety of information in the data warehouse is not needed and the queries are simple.

<table>
<thead>
<tr>
<th>Data Source</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Report</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Query</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Transaction ID</strong></td>
<td><strong>Transaction Datetime</strong></td>
<td><strong>Refresh Datetime</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1001</td>
<td>Sysdate 14:16</td>
<td>Sysdate 14:16</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1002</td>
<td>Sysdate 14:16</td>
<td>Sysdate 14:17</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1004</td>
<td>Sysdate 14:18</td>
<td>Sysdate 14:18</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1006</td>
<td>Sysdate 14:18</td>
<td>Sysdate 14:18</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1007</td>
<td>Sysdate 14:20</td>
<td>Sysdate 14:22</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Benefits:**
- Provide data in real-time.
- Possess some integration and cleansing capabilities.
- Avoid direct reporting from production systems.
- Repackage data to satisfy reporting/analytics needs.

**Enables:**
- Operational BI
- Real-time analytics
- Alerts
- Feeds to downstream operational applications

**Limitations:**
- Data integration is often limited at the ODS level.
- Data structure is not optimized for complex analytics loads.
- Data extraction may slow down upstream transactional systems.
- Lack of a full historical view due to lack of/little data integration.
Create the single-source of truth in an enterprise data warehouse

A data warehouse “is a subject-oriented, integrated, time-variant, non-volatile collection of data used to support the strategic decision-making process for the enterprise. It is the central point of data integration for business intelligence and is the source of data for the data marts, delivering a common view of enterprise data” (Corporate Information Factory).

Benefits:
- Top-down architectural planning.
- Subject areas from all lines of business are integrated.
- Time variant – changes to the data are tracked.
- Non-volatile – data is never over-written or deleted.

Limitations:
- A massive amount of corporate information.
- Slower delivery.
- Changes are harder to make.
- Data format is not very business friendly.

Enables:
- Time-series analyses with historical data
- Enterprise-level, common-view analyses
- Integrated, comprehensive customer profiles
- One-stop repository of all corporate information
Create tailored views of your data assets using **data marts and cubes**

Depending on the depth of analysis business users are requesting, the enterprise data warehouse may not offer enough granularity or customization to be the source of reporting.

Data marts can be created as a subset of the EDW to provide a specific view of the data that is more closely aligned with business operations (usually at the department level). They can also be created independently of the EDW.

**Info-Tech Insight**

Data warehouse performance is negatively impacted when multiple users are relying on it as the direct source of reporting. Regular reporting activities that query the same subsets of data every time (i.e. quarterly financial reporting) are ideal candidates for creating separate data marts and cubes to reduce the risk of EDW failure.

<table>
<thead>
<tr>
<th>Data Source</th>
<th>Data Source</th>
<th>Data Source</th>
<th>Data Source</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
</tr>
</tbody>
</table>

**Benefits:**

- Aligned with a specific business process or activity.
- Business-friendly data format.
- Easy to build and use.
- Easy to make changes.

**Limitations:**

- Bottom-up approach.
- Hard to integrate with other subject areas once dimensions are set.
- Complex point-to-point ETL integration when managing multiple marts joined to the EDW.
- Discrepancies among data marts can manifest.
Reduce the need for distinct architecture and data migration with a data virtualization layer

Data virtualization provides an abstraction layer for simplified and direct access to data from many data sources, without the need for physical movement. Data is accessed, combined, and then federated to deliver a virtual view or even a virtual database. Data virtualization can be used to reorganize your data to create data services.

Enables:
- Rapid prototyping
- Reformat data into more business-friendly format
- Subject-area analyses
- Real-time analytics

Benefits:
- Ideal for rapid prototyping.
- Flexible and scalable.
- Reduce time to analytics.
- Repackage data to suit business needs.

Limitations:
- Performance issues may arise.
- Complex point-to-point integration.
- Slow down source systems.
- One extra layer of business logic.
Leverage the *Data Architecture Planning Tool* to inform your architectural decision

Information needs are **seldom satisfied by using only one pattern** as different user groups require different information at different times. Patterns can be **mixed and matched** to create a customized data warehouse environment specific to your information landscape.

Our tool provides several questions around current information and reporting patterns to help your team identify the optimal combination of data warehouse architecture components.

### Data Architecture Questions

<table>
<thead>
<tr>
<th>The need for an enterprise view...</th>
<th>Explanation</th>
<th>Answer</th>
</tr>
</thead>
<tbody>
<tr>
<td>Do you need a central repository to integrate and store the majority of your corporate data sources?</td>
<td>A centralized repository that contains corporate data such as finance data, marketing data, operational data, HR data, customer data, etc.</td>
<td>Yes</td>
</tr>
<tr>
<td>Do you need to construct an enterprise view of your corporate data?</td>
<td>An enterprise view allows users to look at the corporate data across different subject areas and/or business domains.</td>
<td>Yes</td>
</tr>
<tr>
<td>Do you need to reorganize some of your corporate data into data services?</td>
<td>A data service is created to ease data access by extracting and presenting data from two data marts and from the marketing database.</td>
<td>Yes</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Data warehouse approach...</th>
<th>Explanation</th>
<th>Answer</th>
</tr>
</thead>
<tbody>
<tr>
<td>Are you looking to prototype with your data sources?</td>
<td>A prototype with sales data is created before a new sales data mart is created.</td>
<td>Yes</td>
</tr>
<tr>
<td>Are you looking for a standard data modelling method to fast track the creation of your data warehouse?</td>
<td>A standard data model provides streamlined data modelling principles and processes that provide a step-by-step approach to data modelling.</td>
<td>No</td>
</tr>
<tr>
<td>Do you want to “virtualize” some of your physical data into virtualized data?</td>
<td>Instead of creating a marketing data mart and persisting it into a database, a virtual data mart is created.</td>
<td>Yes</td>
</tr>
<tr>
<td>Are you currently reporting from live production systems?</td>
<td>Live reporting is done on production systems to report on real-time data.</td>
<td>No</td>
</tr>
</tbody>
</table>

*Tab 1. Data Architecture Questions*
Leverage the *Data Architecture Planning Tool* to inform your architectural decision (continued)

After answering questions about your current needs around data and analytics, the tool will provide a recommended architecture diagram with a description of the components.
Create a fit-for-purpose data warehousing model

2.3.2

60 minutes

Instructions

1. Schedule a working session to identify data warehouse patterns. Key participants are the CIO, business unit stakeholders, data warehouse architect, project manager, and business analysts.

2. Review the foundational information needs to get some ideas on the analytic needs of the business as well as the architecture diagram from the Data Architecture Planning Tool.

3. Discuss each DW pattern and document these data points:

<table>
<thead>
<tr>
<th>DW Pattern</th>
<th>Use Cases</th>
<th>Rationale</th>
<th>Pros</th>
<th>Cons</th>
<th>Constraints</th>
</tr>
</thead>
</table>

4. Vote on the DW patterns. Again, you do not need to select only one pattern. Combinations can be selected.

5. Document the pattern in the Data Warehouse Foundation Project Plan Template.

Consider these examples as a starting points:

- **DW Patterns**
  - EDW
  - Data Marts
  - ODS
  
  Patterns for organizations with mature data warehousing

- **DW Patterns**
  - Data Vault
  - Data Marts
  - ODS

  Patterns for organizations that want to accelerate the DW project

- **DW Patterns**
  - Data Marts
  - ODS

  Patterns for organizations that want a department-level focus

INPUT

- Data warehouse architecture patterns
- Results from Data Architecture Planning Tool

OUTPUT

- Consensus on data warehouse architecture model

Materials

- Data Warehouse Foundation Project Plan Template
- Whiteboard
- Markers

Participants

- CIO
- Business stakeholders
- Data warehouse team (project manager architects, integration specialists)
Complement your data warehouse with big data technology

Big data technology such as the Hadoop file system (HDFS) is helpful in storing a large amount of captured data with low cost commodity hardware. Data scientist tools or some BI tools can be used to access and analyze the captured big data on HDFS.

"Today, the relational structure is still relevant because of the disparities and differences in data, structured and unstructured, that can work together."

– Geoffrey Amu
Head of Cloud & Digital Transformation Service, Americas, Capegemini

Data warehouses can be enhanced with big data in the following ways:

<table>
<thead>
<tr>
<th>Data Lake</th>
<th>Data Discovery</th>
<th>Inexpensive Storage</th>
<th>Staging Layer</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data lake is an environment that stores data of any format, schema, and type. It is based on connected, inexpensive commodity hardware. It should be designed as a controlled environment that allows for much more flexibility in associating and merging a wider variety of data sources.</td>
<td>Data discovery is made possible by analyzing the data in the data lake. Data scientist/BI tools may be needed to access and analyze big data.</td>
<td>Big data storage such as HDFS can be used to store historical and/or &quot;cold&quot; data for an active data warehouse. HDFS is also a good candidate for an archive location.</td>
<td>The operation of ETL usually requires temporary storage space to perform integration and transformations. The ability of HDFS to scale up and scale out makes it a prospect.</td>
</tr>
</tbody>
</table>
Take advantage of other Info-Tech resources for a deep dive into building your data architecture

A data warehouse can be a crucial component of getting the right information to the necessary people in the appropriate time frame.

Data warehouse success is heavily dependent on:

- An understanding of how data is meant to flow through the organization.
- Detailed models that enable clarity of how data assets are related and flexibility to respond to changes.

Info-Tech’s approach is centered around creating a hybrid architecture that balances traditional (designs for supporting structured data types) and modern (designs for supporting emerging/disruptive trends and less structured data) elements.

Leverage this blueprint to fully prepare your data architecture for data-driven decision making:

Modernize Data Architecture for Measureable Business Results
Phase 2, Step 4: Explore Technology Options

This step will walk you through the following activities:
2.4.1 Develop a technology capability map to visualize your desired state.

This step involves the following participants:
- Business Stakeholders (from units of targeted processes)
- Data Warehouse Architects
- BI Specialist/Business Analysts
- Project Manager

Outcomes of this step
- Identification of potential vendors that will meet data warehouse architecture needs.
- Definition of immediate and future data warehouse capabilities.
Build a scalable technology stack with the future in mind

A foundational data architecture must be built to provide a foundation for future iterations. The foundation must be scalable to support expansion to address future needs. With the help of some of the emerging technologies—open source, cloud, in-memory computing, virtualization, and big data—options are available. The foundation technology stack should address the following criteria: scalable, Agile, and performance.

### Foundational Technology Stack

<table>
<thead>
<tr>
<th>Scalable</th>
<th>Agile</th>
<th>Performance</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Be able to scale</strong> from a small data warehouse implementation to a larger data warehouse implementation.</td>
<td>Leverage emerging technologies that support an Agile methodology to build the Agile data warehouse.</td>
<td>Be able to support advanced data operations and more complex analytic queries. Ensure quick response to end users.</td>
</tr>
</tbody>
</table>

Leverage open-source platforms, cloud infrastructure, and server virtualization to get started. Don’t over-invest at the beginning. You want to use minimal resources to test drive your approach. Once your approach is refined and evolved after 1-2 iterations, you can easily migrate to more stable technologies.
Take advantage of emerging technologies to build your foundation

Emerging technologies can be used in conjunction with the “iterations on foundation” approach to build a scalable foundation. This lowers the entry point and provides an easy on-ramp for data warehousing.

<table>
<thead>
<tr>
<th>Benefit</th>
<th>Technology</th>
<th>Contributions to the Foundational Technology Stack</th>
</tr>
</thead>
<tbody>
<tr>
<td>Scalable</td>
<td>Cloud DW</td>
<td>• Pay-per-use pricing model</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Managed infrastructure services</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Quick initial start-up</td>
</tr>
<tr>
<td></td>
<td>Hadoop</td>
<td>• Hadoop as the secondary and/or archival storage (cold storage)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Hadoop as the backup media</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Hadoop as the staging area of ETL processes</td>
</tr>
<tr>
<td>Agile</td>
<td>Data virtualization</td>
<td>• Rapid prototyping to create early value</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Package data sources as data services</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Virtualize part of/entire data warehouse</td>
</tr>
<tr>
<td></td>
<td>Materialized view</td>
<td>• Rapid prototyping</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Real-time presentation of data</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Virtualize part of/entire data warehouse</td>
</tr>
<tr>
<td>Performance</td>
<td>In-memory DB</td>
<td>• Improve query/analytics performance</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Auto-tuning capabilities</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Seamless switching between physical and memory storages</td>
</tr>
<tr>
<td></td>
<td>Columnar DB</td>
<td>• Improve query/analytics performance</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Optimized for analytics workload</td>
</tr>
<tr>
<td></td>
<td>DW appliances</td>
<td>• Managed platform with pre-installed software</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Tested and well integrated</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Technical support and services by the same vendor</td>
</tr>
</tbody>
</table>
Get familiar with the vendor landscape and find a suitable product to build the stack

Data warehouses are powered by databases; you must choose a database to build the foundation technology stack. Your initial data warehousing scope might be small and may not qualify for a large seeding budget to buy an expensive database. Alternatively, you can get started by choosing open-source and/or cloud solutions. By leveraging cloud and/or open source, you have the option down the road to scale up by moving from cloud to on-premise or moving from an open-source version to paid commercial versions.

### Cloud Database
Databases that are hosted in the cloud. Vendors provide different options when it comes to hosting:
- **SaaS** – the database is fully managed and you do not need to manage the software and the hardware.
- **PaaS** – a database server in the cloud is provided, and you install and manage the DBMS software.

### Open-Source Database
Open-source databases are available via GNU, MIT, and other community licenses. Some are totally free while others are free for non-commercial uses with an option to upgrade to commercial uses with a minimal fee. Note that not all of the open-source databases offer the features (triggers, stored procedures, etc.) of a commercial database.

Legal and privacy requirements should not scare you away. Many cloud/open-source vendors are flexible and work with clients’ specific legal/privacy needs. Bring your legal counsel and the privacy officer when you meet with the potential cloud or open-source vendors.
Etix leverages a cloud-based DW to simplify operations and optimize cost

**Challenge**

• Etix is the third largest ticketing service provider in North America. It processes more than 50 million tickets per year for more than 4,100 clients in 40 countries.

• Etix ran queries against its production databases for analytics purposes. Given the huge numbers of ticket transactions, system resources were overwhelmed while supporting analytics.

• Etix was searching for a solution that did not require significant capital investment and development resources.

**Solution**

• A three-part solution was purposed: Amazon Redshift as the cloud data warehouse; Attunity CloudBeam to accelerate data transfer to the cloud data warehouse; and Jaspersoft as the front-end BI tool for analytics.

• The Amazon Redshift provides a cloud-based platform to host the data warehouse so that Etix did not have to invest in the data warehouse infrastructure.

• CloudBeam provides good connectivity and integration capability to transform and load internal data to the cloud.

**Results**

• Efficient data loading between on-premise Oracle databases and Amazon Redshift, eliminated the need for a full-time DBA.

• Realized an annual savings of $80,000 as a result of reduced project work and maintenance.

• Ability to update the cloud data warehouse in real-time, which allows decision support/making in real-time.

• The cloud-based technology stack is a flexible architecture that allows Etix to add new data sources easily.
Leverage data warehouse automation tools to accelerate construction

Automation tools speed up development by designing, generating, and loading source data into normalized data warehouse models. Some products also handle mapping from data sources and generation of ETL code. If those basic chores are handled for you, you can dedicate more time and resources to other tasks that will accelerate business value.

Data Warehouse Automation

Automate much of the data modeling and ETL coding. Data schemas for the data warehouse are automatically generated based on the analysis and the mappings of the source data. The core or iteration project team may need to modify the generated ETL codes.
Enhance the traditional data warehouse with **massively parallel processing** and **in-database analytics solutions**

Though a traditional data warehouse integrates multiple data sources to create “a single source” through clearly defined and linked tables, the rigidity of the formatting puts some limitations on the speed at which data can be pushed through to business users.

**In-database analytics**

Newer data warehouse platforms are able to drastically reduce data delivery times by eliminating the need for data to be migrated to another area for analysis. This capability allows for more agile data exploration, manipulation, and model development.

**Something to consider:**
Though bringing the analysis to the data loosens the restraints on speed, it opens up the risk of compromising the integrity of data assets if the appropriate levels of governance are not in place regarding:
• Access rights to data.
• Where the transformations are allowed to occur.
• What is done with the outputs of the analytics activities.

**Massively parallel processing (MPP)**

In order to manage larger volumes of data in a traditional data warehouse, the data is often aggregated to provide more frequent, “real-time” summaries to a business intelligence layer. This process compromises the flexibility of analysis that business users can conduct as the granularity of data is lost.

MPP spreads the processing power of the data warehouse across multiple physical nodes to maintain atomicity of the data as well as speed of delivery.

Performance needs can also be increased in a linear fashion by adding more processing units, compared to traditional data warehouses that have an upper limit of processing power before a larger unit is required.
Get familiar with the ETL and CDC market space to customize the technology stack

**ETL Suites**
Suites that are used to extract data from sources, transform the data, and load the data into the data warehouse. Consider suites with data quality capabilities to improve quality while performing ETL. Also evaluate performance and parallelism of the suite to plan for an increase in data volume in the future.

**CDC Solutions**
CDC solutions identify, capture, and relay data changes in the source systems to the targets. CDC solutions should not degrade source system performance while managing data changes. Make sure the CDC solutions support connections to the source and target systems.

Platform standardization doesn’t lock you in with that vendor for tools; looking elsewhere is often more cost-effective. ETL suites and CDC solutions are modular so you don’t need to get a solution from the vendor that you are already using.
Using a BI platform, connect the warehouse with the business and power users

Business intelligence (BI) platforms serve as a presentation layer/gateway that allows end users to access information through a BI portal. End users can use the artifacts (reports, dashboards, ad hoc queries, scorecards, self-service data sets, etc.) created on the BI portal to make business decisions. In some setups, BI platforms can be used to mashup data that originated from a data warehouse with some other data sources to create impactful analyses.

Selecting the right BI platform is time consuming. Consider this fast-track approach:

1. Group the products into three categories: traditional BI platforms, data visualization platforms, and platforms in-between.

<table>
<thead>
<tr>
<th>Traditional BI Platforms</th>
<th>In-Between Platforms</th>
<th>Data Visualization Platforms</th>
</tr>
</thead>
<tbody>
<tr>
<td>SAP Business Objects</td>
<td>Birst</td>
<td>Qlikview</td>
</tr>
<tr>
<td>IBM Cognos</td>
<td>TIBCO Spotfire</td>
<td>Tableau</td>
</tr>
<tr>
<td>Oracle BI</td>
<td>Microsoft BI</td>
<td>Logi Analytics</td>
</tr>
<tr>
<td>SAS BI</td>
<td>MicroStrategy</td>
<td>YellowFin</td>
</tr>
</tbody>
</table>

2. IT performs a primary evaluation to find the champion for each platform group:

<table>
<thead>
<tr>
<th>Traditional Champion is:</th>
<th>In-Between Champion is:</th>
<th>Data Visualization Champion is:</th>
</tr>
</thead>
</table>

3. Run an end-user pilot with the three champions to select the most suitable BI platform. In some organizations, the use cases are so diverse that multiple BI platforms are selected as a result.

Excel is the number one BI tool. No matter how good your BI platform is, there will be some users who are not willing to switch to BI. When you select a BI platform, be sure that it supports integration with Excel so that those users can look for information on the BI platform and then they can dump the information into Excel for further analyses.
Alternatives to traditional databases if high performance is required

<table>
<thead>
<tr>
<th>Group 1 – Appliance</th>
</tr>
</thead>
<tbody>
<tr>
<td>A data warehouse appliance is pre-configured hardware coupled with a database management software that is engineered for analytical processing. A typical use case is allowing organizations to implement a high-performance data warehouse platform in a relatively short time.</td>
</tr>
<tr>
<td>Greenplum</td>
</tr>
<tr>
<td>Microsoft APS</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Group 2 – ADBMS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Analytic database management systems (ADBMS) specialize in providing high performance for analytic workloads. They are designed mostly for reading purposes while the data in the database is updated on a regular basis or near real-time basis to incorporate recent recent transactions.</td>
</tr>
<tr>
<td>Actian Matrix</td>
</tr>
<tr>
<td>EXASolution</td>
</tr>
<tr>
<td>HPE Vertica</td>
</tr>
<tr>
<td>IBM DB2 with BLU Acceleration</td>
</tr>
<tr>
<td>SAP HANA</td>
</tr>
</tbody>
</table>
Build a technology capability map to visualize your desired state

With an understanding of the available technology and the help of the service-level agreement (SLA), you can develop a technology capability map. The data warehouse stack contains four layers: **data ingestion, data foundation, data analysis, and data presentation.** Different technology will be required to build the capabilities in those layers.

This is a sample data warehouse capability map. You can **customize** the capabilities by applying these steps:

1. **Refer to information and decision needs of your business.** For example, if real-time analytics are required, then the real-time capability is needed in the data foundation layer. That capability will be acquired or built.

2. **Refer to the foundational data architecture.** For example, when data marts are required, you may determine if the data marts are built in the data foundation or in a BI platform.

3. **Refer to the emerging technologies presented in this section.** For example, if a big data platform is required, you may need to build a Hadoop cluster.

---

### The DW Stack

<table>
<thead>
<tr>
<th><strong>Data Ingestion</strong></th>
<th><strong>Data Foundation</strong></th>
<th><strong>Data Analysis</strong></th>
<th><strong>Data Presentation</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td>Connect to Excel</td>
<td>Data warehouse</td>
<td>OLAP cube</td>
<td>Standardized reports</td>
</tr>
<tr>
<td>Connect to ERP</td>
<td>Data marts</td>
<td>Data mash-up</td>
<td>Dashboards</td>
</tr>
<tr>
<td>Connect to CRM</td>
<td>Data vault</td>
<td>Third-party data</td>
<td>Scorecards</td>
</tr>
<tr>
<td></td>
<td>Data warehouse automation</td>
<td>Data virtualization</td>
<td>Location mapping</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th><strong>Basic Capabilities</strong></th>
<th><strong>Advanced Capabilities</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td>Alert</td>
<td>Alerts</td>
</tr>
<tr>
<td>Self-serve BI</td>
<td>Predictive analytics</td>
</tr>
<tr>
<td>Mobile BI</td>
<td>Social analytics</td>
</tr>
<tr>
<td>Data visualization</td>
<td>Big data analytics</td>
</tr>
</tbody>
</table>

---
Develop a technology capability map to visualize your desired state

1. Schedule a technology planning session with the project manager, business analysts, technical lead, data integration specialist, and data warehouse architect to develop the capability map collectively.

2. Review and analyze:
   - Foundational information needs
     - Operational, tactical, and ad hoc activities
   - Foundational data architecture
     - Data warehouse architecture model patterns and BEM (business entity map)
   - Foundation technology stack
     - The art of the possible slides and reference technology capability map

3. Think of the technology requirements and work backwards to figure out the technology capabilities required to support the technology requirements.

<table>
<thead>
<tr>
<th>Data presentation</th>
<th>Capabilities</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Standardized reports</td>
</tr>
<tr>
<td></td>
<td>Dashboard</td>
</tr>
<tr>
<td></td>
<td>Scorecards</td>
</tr>
<tr>
<td>Data analysis</td>
<td>OLAP cube</td>
</tr>
<tr>
<td>Data foundation</td>
<td>Data marts</td>
</tr>
<tr>
<td>Data ingestion</td>
<td>Connect to Excel, ERP, CRM</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Data presentation</th>
<th>Capabilities</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mobile BI</td>
</tr>
<tr>
<td></td>
<td>Location analytics</td>
</tr>
<tr>
<td>Data analysis</td>
<td>Data visualization</td>
</tr>
<tr>
<td>Data foundation</td>
<td>EDW</td>
</tr>
<tr>
<td>Data ingestion</td>
<td>Data cleansing</td>
</tr>
<tr>
<td></td>
<td>Fuzzy matching</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Data presentation</th>
<th>Capabilities</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Alerts</td>
</tr>
<tr>
<td>Data analysis</td>
<td>Predictive analytics</td>
</tr>
<tr>
<td>Data foundation</td>
<td>Data mashup</td>
</tr>
<tr>
<td>Data ingestion</td>
<td>Data appliance</td>
</tr>
<tr>
<td></td>
<td>DW automation</td>
</tr>
<tr>
<td></td>
<td>Connect to social media and big data</td>
</tr>
</tbody>
</table>
# Visualize the possibilities with a technology capability roadmap

Based on the reference technology capability map, you can evaluate your existing technology capabilities to get a sense of the current maturity. The findings can be presented in a heat map to establish an enterprise view of capabilities and to find gaps and opportunities.

## Overall observations:

1. Most organizations are doing some kind of reporting in the current state. You can improve on that capability.

2. Capabilities in the data foundation layer need a lot of development. This blueprint helps you to establish them.

3. You may define advanced capabilities in the reference map. You cannot deploy them at once. Prioritize them and deploy the most beneficial ones.

## Legend
- **Green**: Existing capabilities
- **Yellow**: Partial capabilities
- **Red**: Capabilities not presented

## Basic Capabilities

<table>
<thead>
<tr>
<th>Data presentation</th>
<th>Basic Capabilities</th>
<th>Advanced Capabilities</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>• Standardized reports</td>
<td>• Alerts</td>
</tr>
<tr>
<td></td>
<td>• Dashboards</td>
<td>• Self-serve BI</td>
</tr>
<tr>
<td></td>
<td>• Scorecards</td>
<td>• Mobile BI</td>
</tr>
<tr>
<td></td>
<td>• Location mapping</td>
<td>• Data visualization</td>
</tr>
</tbody>
</table>

## Advanced Capabilities

<table>
<thead>
<tr>
<th>Data analysis</th>
<th>Basic Capabilities</th>
<th>Advanced Capabilities</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>• OLAP cube</td>
<td>• Predictive analytics</td>
</tr>
<tr>
<td></td>
<td>• Data mashup</td>
<td>• Social analytics</td>
</tr>
<tr>
<td></td>
<td>• Third-party data</td>
<td>• Big data analytics</td>
</tr>
<tr>
<td></td>
<td>• Data virtualization</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Data foundation</th>
<th>Basic Capabilities</th>
<th>Advanced Capabilities</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>• Data warehouse</td>
<td>• Real-time analytics</td>
</tr>
<tr>
<td></td>
<td>• Data marts</td>
<td>• Hadoop data warehouse</td>
</tr>
<tr>
<td></td>
<td>• Data vault</td>
<td>• In-memory DBMS</td>
</tr>
<tr>
<td></td>
<td>• Data warehouse automation</td>
<td>• Data warehouse automation</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Cloud data warehouse</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Data ingestion</th>
<th>Basic Capabilities</th>
<th>Advanced Capabilities</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>• Connect to Excel</td>
<td>• Connect to unstructured data</td>
</tr>
<tr>
<td></td>
<td>• Connect to ERP</td>
<td>• Connect to social media</td>
</tr>
<tr>
<td></td>
<td>• Connect to CRM</td>
<td>• Connect to NoSQL</td>
</tr>
<tr>
<td></td>
<td>• Data cleansing</td>
<td>• HDFS as ETL staging area</td>
</tr>
</tbody>
</table>
If you want additional support, have our analysts guide you through this phase as part of an Info-Tech workshop:

**Book a workshop with our Info-Tech analysts:**

- To accelerate this project, engage your IT team in an Info-Tech workshop with an Info-Tech analyst team.
- Info-Tech analysts will join you and your team onsite at your location or welcome you to Info-Tech's historic Toronto office to participate in an innovative onsite workshop.
- Contact your account manager ([www.infotech.com/account](http://www.infotech.com/account)), or email Workshops@InfoTech.com for more information.

**The following are sample activities that will be conducted by Info-Tech analysts with your team:**

**Articulate the mission of the data warehouse program and identify the most important business processes**

Our analyst team will ensure that the data warehouse initiative is closely aligned with the overall business strategy by involving business and IT stakeholders in the creation of a program charter, as well as fully characterizing the most critical reporting, analytics, and data-related activities that need to be enabled by the data warehouse foundation.

**Develop the core data model**

Our analyst team will take your project team through understanding and creating the four “Rosetta Stones” for data modeling, as well as defining business entity relationships and needs for history tracking in the data warehouse.
If you want additional support, have our analysts guide you through this phase as part of an Info-Tech workshop.

**Book a workshop with our Info-Tech analysts:**

### 2.3.1-2.3.3

**Pinpoint your data warehouse architecture pattern**

Our analyst team will educate IT and business stakeholders on the different components of data warehouse architecture, as well as explore architecture combinations that will support your organization’s use cases.

### 2.4.1

**Review the data warehouse technology landscape**

Using your high-level architecture model and business drivers, our analyst team will give a brief overview of the vendor landscape, followed by assisting your project team in articulating technology capabilities that will support your desired data warehouse environment.
PHASE 3

Plan for Data Warehouse Governance

Build an Extensible Data Warehouse Foundation
Phase 3 outline

Call 1-888-670-8889 or email GuidedImplementations@InfoTech.com for more information.

Complete these steps on your own, or call us to complete a guided implementation. A guided implementation is a series of 2-3 advisory calls that help you execute each phase of a project. They are included in most advisory memberships.

Guided Implementation 3: Plan for Data Warehouse Governance
Proposed Time to Completion (in weeks): 1-2 weeks

Step 3.1: Plan for the Centre of Excellence

Start with an analyst kick off call:
• Discuss the advantages of creating a data warehouse centre of excellence (DW CoE).
• Review the necessary roles and responsibilities required for a DW CoE.

Then complete these activities…
• Identify membership, responsibilities, and goals for the DW CoE.
• Consolidate the outputs from this blueprints activities in a project roadmap.

Step 3.2: Define Operating Procedures

Review findings with analyst:
• Discuss considerations that should be made regarding data warehouse availability.
• Discuss the importance standardizing data warehouse procedures (extraction, transformation, loading, validation, consumption).

Then complete these activities…
• Work with the business users to define service-level agreements.
• Create the data warehouse standard operating procedures.

With these tools & templates:
• Data Warehouse SLA Template
• Data Warehouse Standard Operating Procedures Template

Phase 3 Results & Insights:
• Build a center of excellence and treat the members as data warehouse gurus. Consult them on future extensions to your data warehouse or changes to the existing data warehouse.
• All data warehouses ingest, process, and supply data regardless of industry and organization size. Build templates from those data operations to inject standards and to accelerate future warehouse extensions.
**Phase 3, Step 1: Plan for the Center of Excellence**

**This step will walk you through the following activities:**

3.1.1 Create a plan for the data warehouse center of excellence (DW CoE).
3.1.2 Create a comprehensive project roadmap.

**This step involves the following participants:**

- Project Sponsor
- Data Warehouse Foundation Team

**Outcomes of this step**

- Identify membership and goals for your data warehouse center of excellence.
- Consolidate activity outputs in a foundation project roadmap.
Establish a data warehouse **center of excellence** to create consistency in data warehouse operations

The DW program needs to be agile to adopt the changing needs of different phases of the DW project. Essentially the data warehouse program is comprised two project types: **foundation building** and **improvement projects via iterations**. There should be two types of project team: one establishes the initial data warehouse and helps co-ordinate later projects, and the other executes iterations to extend the program.

The core project team is the backbone of the data warehouse program and should serve as the initial **center of excellence**. The team will **make the key decisions** around the foundation of the data warehouse, and help set long-term goals for performance.

The iteration project team is peripheral to the core project team. They execute iterations to enable departmental needs. They **work collaboratively** with the core team to ensure iterations are aligned with the program.
Use the center of excellence as a governance authority, but also as a point of knowledge sharing

As the data warehouse program expands, governance needs to scale accordingly to keep new initiatives in line with the foundation and preserve the data warehouse vision.

The data warehouse center of excellence should be a point of contact for subsequent project teams to ensure that:

- Policies and procedures are communicated to and understood by project teams.
- Duplication of architecture/technology does not occur.
- Data models are understood and any changes that need to be made are validated so as to not disrupt existing data warehouse activities.

If you’re heavily executing in the project, it’s difficult to keep the higher level view of the project itself [and overall program]. You need to structure your team and your roles and responsibilities very well in advance.

– Gus Marquez
Senior IT Project Manager, CIBC

Use the data warehouse foundation team as initial members of the CoE.

The best practices get brought back to the governance committee, who will consider whether or not to implement those practices across the entire user base.

– Sree Pulapaka
Director of Enterprise Business Innovation and Analytics, MWAA

The CoE should also function as a place where those interested in data warehousing activities can share their experiences and best practices to grow the organization’s collective wisdom and continually evolve the program.
Identify a **BI/data warehouse director** to be the head of your center of excellence

As your data warehouse program grows, there needs to be an identified champion and manager who can continue to:

- Encourage the sharing of best practices across the organization and drive a culture of creating data-driven insight.
- Co-ordinate efforts to evaluate the success of the data warehouse program, via user-satisfaction interviews and other metrics.
- Serve as a key point of contact between the iteration project teams and the data warehouse steering committee.

These responsibilities can be scaled down initially, or spread across multiple members of the foundation project team.

Leverage Info-Tech’s **Director of Data Warehousing/Business Intelligence Job Description Template** to source for candidates externally, or identify a current member of staff with the right capabilities.
Establish a data warehouse center of excellence and consult the center with all things related to the data warehouse

3.1.1

Your data warehouse program should be directed by a data warehouse center of excellence to ensure quality and consistence. This exercise helps you to establish such an organization.

1. Take an inventory of your data management professionals. The DW CoE is very likely to draw personnel from that group of employees. Consider roles with the following capabilities:

<table>
<thead>
<tr>
<th>Capability</th>
<th>Requirement</th>
</tr>
</thead>
<tbody>
<tr>
<td>Business Analysis</td>
<td>Leadership</td>
</tr>
<tr>
<td>Project Management</td>
<td>Data Architecture and Model</td>
</tr>
<tr>
<td></td>
<td>Business Intelligence</td>
</tr>
</tbody>
</table>

2. Define DW CoE roles and responsibilities and assign professionals to the roles. You can refer to the data management inventory above.

3. Develop mandate and responsibilities for the DW CoE. Responsibilities should be focused around the goals of having a DW CoE:

<table>
<thead>
<tr>
<th>Goals</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ensure consistency between the data warehouse phases.</td>
</tr>
<tr>
<td>Ensure a data warehouse is architected with the “big picture” in mind.</td>
</tr>
<tr>
<td>Promote integration of siloed data.</td>
</tr>
<tr>
<td>Preserve institutional memory around the data warehouse.</td>
</tr>
<tr>
<td>Provide guidance and advice to projects involving the data warehouse.</td>
</tr>
</tbody>
</table>

INPUT

- Organizational chart for the data management organization

OUTPUT

- DW CoE staffing model
- DW CoE functions and responsibilities

Materials

- Whiteboard
- Data Warehouse Project Plan

Participants

- CIO
- SVP of information management
- Data warehouse project team
Create a data warehouse development roadmap

3.1.2

Using the outputs of the previous phases’ activities, create a **roadmap** in the form of a Gantt chart to organize the initiatives and tasks for the data warehouse project and program.

### OUTPUT
- Summary roadmap

### Materials
- Data Warehouse Foundation Project Plan Template

### Participants
- Data warehouse foundation team

<table>
<thead>
<tr>
<th>Q1</th>
<th>Q2</th>
<th>Q3</th>
<th>Q4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Communicate vision and mission statement for data warehouse</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Formalize steering committee membership</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Formalize foundation project team roles/responsibilities</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Develop data governance policies</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Build business data glossary and data classification scheme</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Research and select external data feeds for reference data tables</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Perform technology requirements gathering based on business processes</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hold inaugural centre of excellence meeting</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Implement annual data operations policy review</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Define success metrics for the data warehouse program

1. Revisit the driving statements for the data warehouse created in phase 2 (vision, mission, and goals).
   - Whereas project metrics are focused on the evaluating the foundation and iteration projects, program metrics should be more aligned with how the data warehouse is evolving from an enterprise standpoint.

2. Using the overall data warehouse goals, develop metrics that articulate the expansion of your program, whether it is by:
   - Size of user base
   - Variety of reports produced
   - Architecture that supports specific business units/functions

3. Document these metrics in the Data Warehouse Program Charter Template.

Program Metrics

Goals:
1. Efficient data repository
2. Effective data repository
3. Comprehensive data repository

Metrics:
- Time spent fulfilling data/reporting requests
- Data quality score (accuracy, completeness)
- Number of integrated data sets
- Number of self-service requests occurring
- Membership of data warehouse steering committee and center of excellence
Leverage other Info-Tech resources to create a robust center of excellence that optimizes your data warehouse

By framing the data warehouse as a program instead of a project, your data warehouse needs to be treated as an enterprise application that brings together stakeholders from all lines of the organization.

The center of excellence is an integral piece in ensuring the success and sustainability of the data warehouse environment by mitigating the following risks:

- Processes pertaining to managing the application are inconsistent and do not drive excellence.
- There is a lack of interdepartmental collaboration between different teams pertaining to the application.
- There are no formalized roles and responsibilities for governance and support around enterprise applications.

Download the *Maximize the Benefits from Enterprise Applications with a Center of Excellence* blueprint.
**Phase 3, Step 2: Define Operating Procedures**

**This step will walk you through the following activities:**

3.2.1. Define service-level agreements.
3.2.2. Customize and improve the data warehouse standard operating procedures.

**This step involves the following participants:**

- Project Sponsor
- Data Warehouse Foundation Team

**Outcomes of this step**

- Business-IT alignment on the availability of the data warehouse.
- Document the processes related to data delivery, from end to end.
Work collaboratively with the business to define SLAs

The latency and availability of the data is an important factor in data warehouse design. IT teams will forget to ask business users what frequency of refreshes is adequate for supporting their activities.

– Sree Pulapaka
Director, Enterprise Business Innovation and Analytics, MWAA

Help the business to understand and picture service levels by discussing all these scenarios:

**Scenario 1:** What if the data warehouse is down?

**Scenario 2:** What if the network is down but the DW is still online?

**Scenario 3:** What if the BI suite is down?

**Scenario 4:** What if the ETL is delayed for two hours?

**Scenario 5:** What if the month-end closing process is using so much data warehousing power that it holds back other analytics?

**Scenario 6:** What if incorrect data is found in the data warehouse?

- Hold the IT accountable for agreed-upon service levels.
- Protect the IT department from unrealistic expectations.
- Increase IT department credibility and perceived value.
- Improve communication between IT and the business.
Work collaboratively with the business to define an SLA

Instructions

1. Schedule an SLA working session with the business representatives individually. Ideally the business representatives should be business executives. Business executives have the power to make final decisions on the SLA.

2. Invite core project team members including the project manager, business analysts, project technical lead, and system integrator.

3. Present scenarios to the business representatives to help them to picture what may happen in case of an outage.

4. Analyze from two angles:
   - Brainstorm the measures needed to revert the outages.
   - Brainstorm the measures needed to prevent the outages.

5. Document the SLA information in the Agile Data Warehouse Service Level Agreement Template.

INPUT
- Business needs

OUTPUT
- Data warehouse service-level agreement

Materials
- Whiteboard
- Markers
- Agile Data Warehouse Service Level Agreement Template

Participants
- The data warehouse project team
- Business representatives (ideally business executives)

Scenario Illustrations
- Invite business executives to the SLA working sessions. They are so high up in rank that they are able to make final decisions on SLA items such as uptime, performance expectations, and recovery time.

Measures that revert outage

Measures that prevent outage
Leverage a simple data warehouse SOP to govern data operations in the warehouse

In a data warehouse environment you would manipulate the data before the data is delivered to the end users or business intelligence. The most common data operations are extraction, transformation, loading, validation, and consumption. These general steps are very universal and they are applicable to different data sets that you want to onboard. That is the reason you can leverage a standard operating procedure (SOP) to inject governance into these steps.

Benefits of leveraging a data warehouse standard operating procedure (DW SOP):
• Serves as a guiding document. It outlines best practices around data operations to help you to structure your project effort.
• Works as a scoping template. You can use the considerations to understand and define the scope of the data operations.
• Provides a checklist. Use as a checklist to ensure essential requirements are covered in new data warehouse iterations.
Standard operating procedures are essential for data warehouse program growth

The DW SOP is a living document that should be continually improved. As your data warehouse program expands and new iterations are added, any improved methodologies and processes for onboarding new data sets should be incorporated into the SOP.

The DW SOP is divided into five sections:

- **Data Extraction**: The data operation that extracts data from the upstream source systems.
- **Data Transformation**: The data operation that transfers data into a new format or calculates new values.
- **Data Loading**: The data operation that loads data into staging and the EDW.
- **Data Consumption**: The data operation that re-organizes data into business views.
- **Data Validation**: The data operation that enhances data to increase its data value.

**Info-Tech Insight**

Take the DW SOPs to the next level by creating usable templates or data patterns in your ETL platform that use the best practices suggested by the SOP. This approach ensures the best practices are “built into” the ETL platform.
Understand the extraction, transformation, and loading operations

Extraction, transformation, and loading operations involve the source systems and the data warehouse (EDW and data vault).

<table>
<thead>
<tr>
<th>Operation</th>
<th>Objectives</th>
<th>Example</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data Extraction</td>
<td>• Extract data from multiple source systems.</td>
<td>Extract a subset of Salesforce data on an incremental basis efficiently using the Salesforce API.</td>
</tr>
<tr>
<td></td>
<td>• Plan for extraction methodology and frequency.</td>
<td>Calculate derived metrics such as YTD revenue, MTD cost, and revenue per sales person.</td>
</tr>
<tr>
<td></td>
<td>• Present best practices in data extraction.</td>
<td>Load a large volume of data into a live data warehouse without affecting the access and query performance.</td>
</tr>
<tr>
<td>Data Transformation</td>
<td>• Transform data so that operational data is transformed in a way that supports analytics consumption.</td>
<td></td>
</tr>
<tr>
<td></td>
<td>• Optimize calculations and complex transformation.</td>
<td></td>
</tr>
<tr>
<td></td>
<td>• Present best practices in data transformation.</td>
<td></td>
</tr>
<tr>
<td>Data Loading</td>
<td>• Load data accurately and effectively to the predefined data warehouse model.</td>
<td></td>
</tr>
<tr>
<td></td>
<td>• Optimize loading speed and suggest fit-for-purpose loading approach using tips and tricks.</td>
<td></td>
</tr>
<tr>
<td></td>
<td>• Present best practices in data loading.</td>
<td></td>
</tr>
</tbody>
</table>
Understand the validation and consumption operations

Validation and consumption operations optimize the ease of data consumption from a business end-user perspective.

<table>
<thead>
<tr>
<th>Operation</th>
<th>Objectives</th>
<th>Example</th>
</tr>
</thead>
</table>
| Data Consumption        | • Present data for business consumption via business intelligence or direct access.  
                          | • Consider re-formatting data to improve analytics capabilities.          | Construct a business facing semantic layer and load less-organized data to that layer. |
|                         | • Present best practices in data consumption.                              | Validate customer names and addresses and segment the customers according to the ZIP code. |
| Data Validation         | • Cleanse data to ensure accurate and complete data is extracted, processed, and stored.  
                          | • Enhance data with additional data attributes to add value and context to the data.  
                          | • Present best practices in data validation.                              |
Customize and improve the data warehouse SOP

3.2.2

Info-Tech’s *Data Warehouse Standard Operating Procedures Template* is a good standing point to build your maturity in data operations. Use this exercise to customize the SOP and/or improve it.

1. Reflect on your organization and your data landscape. Think about:
   - Are you similar to many organizations out there?
   - Are you unique in some areas?
   - Are you more or less mature in the context of data management?

   Document the ideas you came up with. They will inform the customization.

2. Chances are, you are already performing the same data operations. What have you learned from your experience? Leverage the Agile retrospective methodology to ideate:

   - **Start**
     - Documenting data lineage
   - **Continue**
     - Using out-of-the-box templates in the ETL suite
   - **Stop**
     - Stop using ETL; use ELT for customer data

3. Inform ideas by the previous activities. Brainstorm what best practices and considerations should be incorporated or updated in the SOP in these areas:
   - Extraction, Transformation, and Loading
   - Validation and Consumption

**INPUT**

- *Data Warehouse Standard Operating Procedures Template*

**OUTPUT**

- Customized *Data Warehouse Standard Operating Procedures Template*

**Materials**

- Whiteboard
- Markers

**Participants**

- The data warehouse project team
- DW CoE (if any)
If you want additional support, have our analysts guide you through this phase as part of an Info-Tech workshop

Book a workshop with our Info-Tech analysts:

- To accelerate this project, engage your IT team in an Info-Tech workshop with an Info-Tech analyst team.
- Info-Tech analysts will join you and your team onsite at your location or welcome you to Info-Tech’s historic Toronto office to participate in an innovative onsite workshop.
- Contact your account manager (www.infotech.com/account), or email Workshops@InfoTech.com for more information.

The following are sample activities that will be conducted by Info-Tech analysts with your team:

**Develop a plan for a data warehouse center of excellence and project roadmap**

Our analyst will guide your team through laying the groundwork for a data warehouse CoE, from best practices for identifying candidates to helping you define what the goals and expectations for the CoE should be, as well as consolidating the blueprint activity outputs in an activity roadmap as part of the project plan.

**Define operating procedures**

Our analysts will work with your project team to develop the right questions for eliciting the service-level requirements for the data warehouse, and document data operation procedures (validation, extraction, transformation, loading, and consumption).
# Summary of accomplishment

## Knowledge Gained
- Understanding of governance as a significant component of building a successful data warehouse.
- “Rosetta Stone” views of data that accelerate data modeling.
- Appropriate structure of data warehouse teams.
- Vendor technologies and capabilities.

## Processes Optimized
- Identification of business activities that can be supported by a data warehouse.
- Connecting business needs to data warehouse architecture.

## Deliverables Completed
- Data Warehouse Foundation Project Plan Template
- Data Warehouse Program Charter Template
- Data Warehouse Steering Committee Charter
- Data Warehouse Work Breakdown Structure Template
- High-Level Data Warehouse Architecture Diagram
- Data Warehouse Service-Level Agreements
- Data Warehouse Standard Operating Procedures Template
Insight breakdown

Insight 1

A data warehouse is a project, but successful data warehousing is a program. An effective data warehouse requires planning beyond the technology implementation. In addition to supporting immediate needs, appraising future corporate goals and objectives will inform how the data warehouse can remain a relevant and sustainable source of business value.

Insight 2

Governance, not technology needs to be the core support system for enabling a data warehouse program. How the data warehouse functions and supports the goals of your environment is more dependent on which stakeholders are involved, as well as the policies and procedures that dictate how data can and should be handled.

Insight 3

Understand business processes at the operational, tactical, and ad hoc levels to ensure a fit-for-purpose data warehouse is built. For the data warehouse to be worth the investment, it needs to be able to support a wide variety of user needs that demand different information at different times. By beginning the data warehouse planning with descriptions of business activities, data models can be constructed in a way that provides the appropriate views of corporate data.
Appendices

Data Management Framework
Contributors
Bibliography
Data management needs to stand on a solid framework; we used three: DMBOK2, Mike 2.0, and COBIT 5

As part of our research process, we leveraged the frameworks of COBIT 5, Mike 2.0, and DAMA DMBOK2. Contextualizing information management within these frameworks clarifies its importance and role, and ensures that initiatives are focused on key priority areas.

The DMBOK2 data management framework by the Data Asset Management Association (DAMA) provided a starting point for our classification of the components in our data management framework.

Mike 2.0 is a data management framework that helped guide the development of our framework through its core solutions and composite solutions.

The COBIT 5 framework and its business enablers were used as a starting point for assessing the performance capabilities of the different components of information management.

Info-Tech’s Data Management Framework outlines the key data enablers and business information that should likewise be enhanced by and support data governance.
Info-Tech’s Data Management Framework

Info-Tech’s Data Management Framework is designed to show how an organization’s business model sits as the foundation of its data management practice. Drawing from the requirements of the underpinning model, a practice is designed and maintained through the creation and application of the enablers and dimensions of data management.

Please note the components of the model are not meant to reflect a flow diagram, but to instead reflect a taxonomy of capabilities and components needed for effective data management.

Adapted from DAMA DM80K and Advanced Knowledge Innovations Global Solutions
Decoding Info-Tech’s Data Management Framework – Layer 1: Business Information

**Business Information**

Data subject areas provide high-level views of the data assets that are used in business processes and enable an organization to perform its business functions.

Classified by specific subjects, these groups reflect data elements that, when used effectively, are able to support analytical and operational use cases of data.

This layer is representative of the delivery of the data assets and the organization’s consumption of the data.

For a data management practice to be effective, it ultimately must show how its capabilities and operations better support the business in accessing and leveraging its key data assets.
Decoding Info-Tech’s Data Management Framework – Layer 2: Information Dimensions

Information Dimensions
Components at the information dimensions layer manage the different types of data and information present within an environment.

At this layer, data is managed based on its type and how the business is looking to use and access the data.

Custom capabilities are developed at this level to support:
- Structured data
- Semi-structured data
- Unstructured data

The types, formats, and structure of the data are managed at this level, using the data management enablers to support their successful execution and performance.
Decoding Info-Tech’s Data Management Framework – Layer 3: Data Management Enablers

**Data Management Enablers**
Info-Tech categorizes data management enablers as the processes that guide the management of the organization's data assets and support delivery.

- **Govern and Direct**
  - Ensures data management practices and processes follow the standards and policies outlined for them.
  - Manages the executive oversight of the broader practice.

- **Align and Plan**
  - Aligns data management plans to the business' data requirements.
  - Creates the plans to guide the design and execution of data management components.

- **Build, Acquire, Operate, Deliver, and Support**
  - Executes the operations that manage data as it flows through the business environment.
  - Manages the business' risks in relation to its data assets and the level of security and access required.

- **Monitor and Improve**
  - Analyzes the performance of data management components and the quality of business data.
  - Creates and executes plans to improve the performance of the practice and the quality and use of data assets.
Research contributors

Internal Contributors

• Chris O’Connor, Manager- Business Analysis and Data Management
• Rob Anderson, Database Administrator and Business Intelligence Manager
• Keith Tudor, Data Architect

External Contributors

• Chris Debo, Senior Manager, Schneider Downs & Co., Inc.
• Geoffrey Amu, Head of Cloud & Digital Transformation Service, Americas, Capgemini
• Gus Marquez, Senior IT Project Manager, CIBC
• Jaison Dominic, Lead Architect- Enterprise Data Warehouse, Moffitt Cancer Center
• Liselle Ramcharan, Project Manager, TD Insurance
• Randy Piscione, Enterprise Data Architect, BMO Financial Group
• Sree Pulapaka, Director of Enterprise Business Innovation and Analytics, Metropolitan Washington Airports Authority
• 1 anonymous contributor
Research contributors and experts

Chris Debo, Senior Manager, Technology Advisory
Schneider Downs

Chris Debo, CISA, is a Senior Manager in Schneider Downs’ Technology Advisory practice and an Adjunct Professor of Management Information Systems at The Ohio State University’s Fisher College of Business. His 15 years of information technology experience spans multiple disciplines including risk management, regulatory compliance, cybersecurity, software development, data warehousing, business intelligence, ERP implementation, and technology optimization.

Jaison Dominic, Lead Architect, Enterprise Data Warehouse
Moffitt Cancer Center

Jaison has strong data warehouse and information management experience providing leadership and customer-focused solutions architecture in the development of high performance data warehouse and business intelligence applications. He is a results-driven data solutions architect with 16 years' experience in building enterprise-wide business intelligence systems in the healthcare industry, leading onshore and offshore application development teams. His areas of expertise include enterprise data strategy and execution, value-added BI solutions, technical program management, team and relationship building, data architecture leadership, master data management, and healthcare performance management using Oracle EHA and HDWF.
Research contributors and experts

Liselle Ramcharan, Former Project Manager
TD Insurance

Liselle, PMP, has over ten years of experience in information technology and project management, at the business, technical, and consultative levels. She holds a BSc (Hons) in Electrical Engineering and an MSc in Digital Systems, among other qualifications. She has extensive knowledge in the fields of information and project management with her most recent experience being in the finance industry. She has an emerging passion for organizational change management, which she sees as a perfect complement to truly successful information and project management implementations. Her specialties are metadata management, project management, and master data management.

Randy Piscione, Enterprise Data Architect,
BMO Financial Group

Randy started his professional life as a musician and moved into C coding after a couple of years touring Canada. Since beginning his data architect career in 1994, Randy has worked in retail, banking, distribution, railroading, agriculture, and other industries, for organizations such as Canadian Tire, CP Rail, TJX, Ontario Hydro, and many others. He recently moved from a 22-year stint as an independent consultant to become an enterprise data architect for BMO.
Bibliography


