DATA WAREHOUSES – part6

1. **Introduction to the DWH discipline**
2. **Brief history** (Inmon &Linstedt,2015) **and Data Architecture** (Kimball&Ross,2016)
3. **Dimensional Modeling Fundamentals** (Kimball&Ross,2016)
4. **Technical Architecture Considerations** (Kimball&Ross,2016)
5. **Extract Transform Load and Data Quality** (Kimball&Ross,2016)
6. **DWH Lifecycle** (Kimball&Ross,2013)
7. **Trends in the evolution of DWH:**

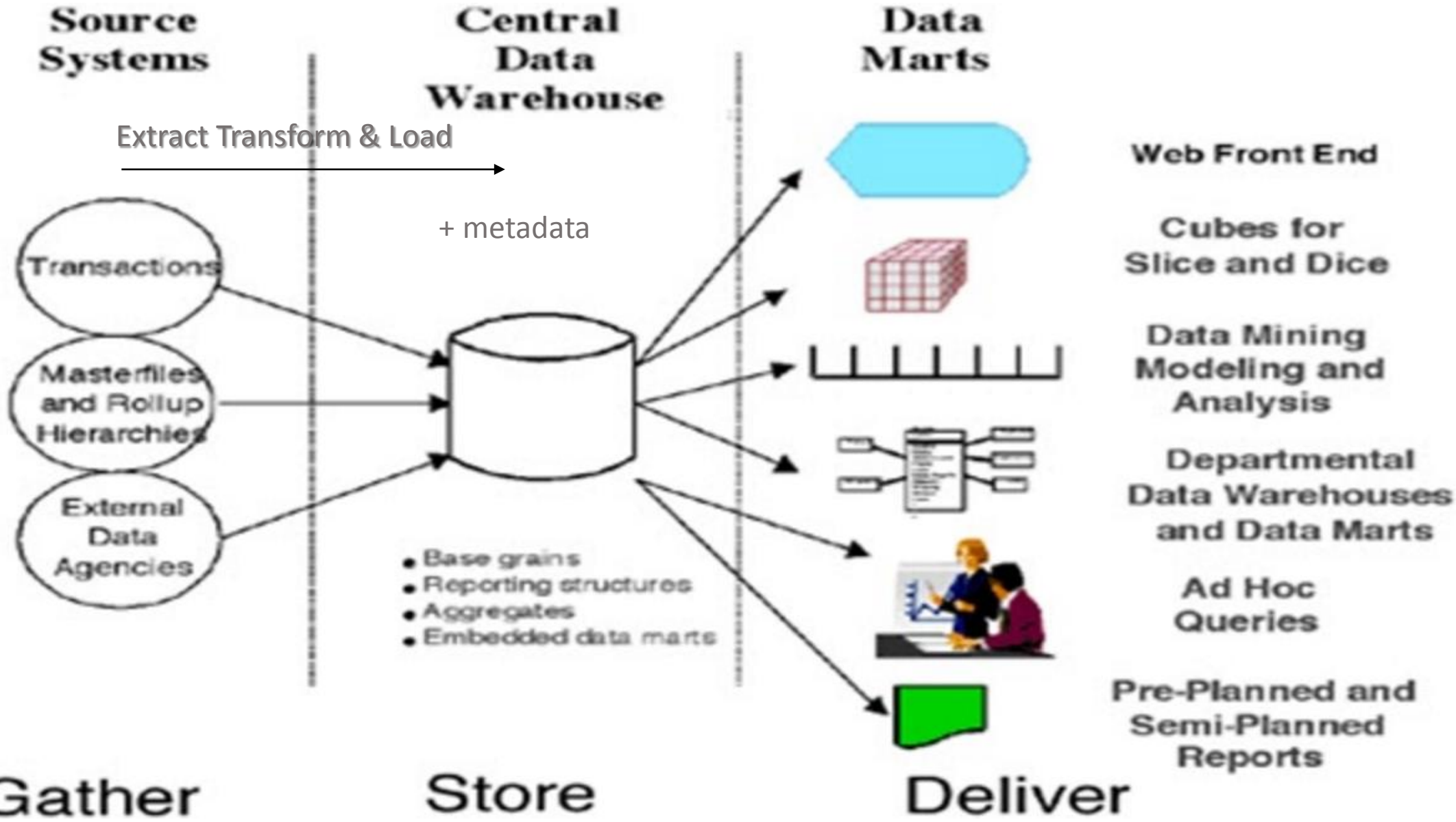
   Extended RDBMS Architecture (Kimball&Ross,2013)**.**

   Pushing into the Future (Reeves,2009)**.**

   DWH 1.0 vs. 2.0 (Krishnan,2013)

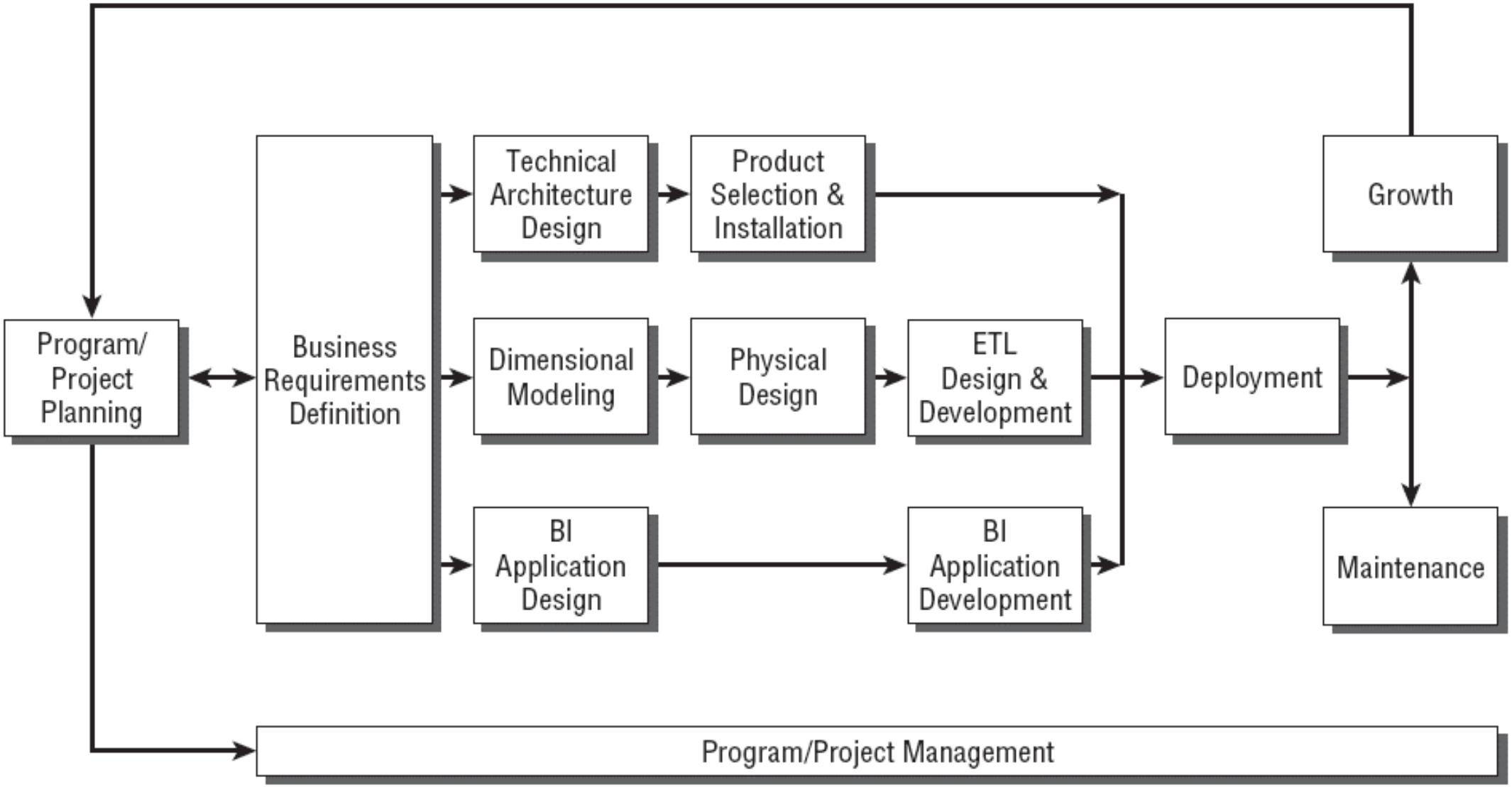
DWH – overview

- *Data mart* has been replaced with *business process dimensional model*, *business process subject area*, or just *subject area*, depending on the context. (Kimball&Ross,2016)



DWH Life Cycle Diagram

Kimball &Ross,2013



DWH Life Cycle – Program/Project Planning

Kimball &Ross,2013

Assessing Readiness

Before moving ahead with a DW/BI effort, it is prudent to take a moment to assess the organization's readiness to proceed. Based on our cumulative experience from hundreds of client engagements, three factors differentiate projects that were predominantly smooth sailing versus those that entailed a constant struggle.

The most critical readiness factor is to have a strong executive business sponsor. Business sponsors should have a clear vision for the DW/BI system's potential impact on the organization. Optimally, business sponsors have a track record of success with other internal initiatives. They should be politically astute leaders who can convince their peers to support the effort

DWH Life Cycle – Program/Project Planning

Kimball &Ross,2013

The second readiness factor is having a strong, compelling business motivation for tackling the DW/BI initiative. This factor often goes hand in hand with sponsorship. The DW/BI project needs to solve critical business problems to garner the resources required for a successful launch and healthy lifespan

The third factor when assessing readiness is feasibility. There are several aspects of feasibility, including technical and resource feasibility, but data feasibility is the most crucial. Are you collecting real data in real operational source systems to support the business requirements? Data feasibility is a major concern because there is no short-term fix if you're not already collecting reasonably clean source data at the right granularity.

DWH Life Cycle – Program/Project Planning

Kimball &Ross,2013

Scoping and Justification

When you're comfortable with the organization's readiness, it's time to put boundaries around an initial project. Scoping requires the joint input of the IT organization and business management. The scope of a DW/BI project should be both meaningful to the business organization and manageable for the IT organization. You should initially tackle projects that focus on data from a single business process; save the more challenging, cross-process projects for a later phase. Remember to avoid the *Law of Too* when scoping—too brief of a timeline for a project with too many source systems and too many users in too many locations with too diverse analytic requirements.

Justification requires an estimation of the benefits and costs associated with the DW/BI initiative. Hopefully, the anticipated benefits grossly outweigh the costs. IT usually is responsible for deriving the expenses. DW/BI systems tend to expand rapidly, so be sure the estimates allow room for short-term growth. Unlike operational system development where resource requirements tail off after production, ongoing DW/BI support needs will not decline appreciably over time.

The business community should have prime responsibility for determining the anticipated financial benefits. DW/BI environments typically are justified based on increased revenue or profit opportunities rather than merely focusing on expense reduction. Delivering "a single version of the truth" or "flexible access to information" isn't sufficient financial justification. You need to peel back the layers to determine the quantifiable impact of improved decision making made possible by these sound bites. If you are struggling with justification, this is likely a symptom that the initiative is focused on the wrong business sponsor or problem.

DWH Life Cycle – Program/Project Planning

Kimball &Ross,2013

Staffing

From the business side of the house, we'll need representatives to fill the following roles:

- **Business sponsor.** The sponsor is the DW/BI system's ultimate client, as well as its strongest advocate. Sponsorship sometimes takes the form of an executive steering committee, especially for cross-enterprise initiatives.
- **Business driver.** In a large organization, the sponsor may be too far removed or inaccessible to the project team. In this case, the sponsor sometimes delegates less strategic DW/BI responsibilities to a middle manager in the organization. This driver should possess the same characteristics as the sponsor.
- **Business lead.** The business project lead is a well-respected person who is highly involved in the project, communicating with the project manager on a daily basis. Sometimes the business driver fills this role.
- **Business users.** Optimally, the business users are the enthusiastic fans of the DW/BI environment. You need to involve them early and often, beginning with the project scope and business requirements. From there, you must find creative ways to maintain their interest and involvement throughout the project. Remember, business user involvement is critical to DW/BI acceptance. Without business users, the DW/BI system is a technical exercise in futility.

DWH Life Cycle – Program/Project Planning

Kimball &Ross,2013

Staffing

Several positions are staffed from either the business or IT organizations. These straddlers can be technical resources who understand the business or business resources who understand technology:

- **Business analyst.** This person is responsible for determining the business needs and translating them into architectural, data, and BI application requirements.
- **Data steward.** This subject matter expert is often the current go-to resource for ad hoc analysis. They understand what the data means, how it is used, and where data inconsistencies are lurking. Given the need for organizational consensus around core dimensional data, this can be a politically challenging role, as we described in Chapter 4: Inventory.
- **BI application designer/developer.** BI application resources are responsible for designing and developing the starter set of analytic templates, as well as providing ongoing BI application support.

DWH Life Cycle – Program/Project Planning

Kimball &Ross,2013

Staffing

The following roles are typically staffed from the IT organization:

- **Project manager.** The project manager is a critical position. This person should be comfortable with and respected by business executives, as well as technical resources. The project manager's communication and project management skills must be stellar.
- **Technical architect.** The architect is responsible for the overall technical architecture. This person develops the plan that ties together the required technical functionality and helps evaluate products on the basis of the overall architecture.
- **Data architect/modeler.** This resource likely comes from a transactional data background with heavy emphasis on normalization. This person should embrace dimensional modeling concepts and be empathetic to the requirements of the business rather than focused strictly on saving space or reducing the ETL workload.
- **Database administrator.** Like the data modeler, the database administrator must be willing to set aside some traditional database administration truisms, such as having only one index on a relational table.
- **Metadata coordinator.** This person helps establish the metadata repository strategy and ensures that the appropriate metadata is collected, managed, and disseminated.
- **ETL architect/designer.** This role is responsible for designing the ETL environment and processes.
- **ETL developer.** Based on direction from the ETL architect/designer, the developer builds and automates the processes, likely using an ETL tool.

DWH Life Cycle – Program/Project Planning

Kimball &Ross,2013

Developing and Maintaining the Plan

Any good project manager knows key team members should develop estimates of the effort required for their tasks; the project manager can't dictate the amount of time allowed and expect conformance. The project plan should identify acceptance checkpoints with business representatives after every major roadmap milestone and deliverable to ensure the project remains on track.

DW/BI projects demand broad communication. Although project managers typically excel at intra-team communications, they should also establish communication strategy describing the frequency, forum, and key messaging for other constituencies, including the business sponsors, business community, and other IT colleagues.

Finally, DW/BI projects are vulnerable to scope creep largely due to a strong need to satisfy business users' requirements. You have several options when confronted with changes: Increase the scope (by adding time, resources, or budget), play the zero-sum game (by retaining the original scope by giving up something in exchange), or say "no" (without actually saying "no" by handling the change as an enhancement request). The most important thing about scope decisions is that they shouldn't be made in an IT vacuum. The right answer depends on the situation. Now is the time to leverage the partnership with the business to arrive at an answer that everyone can live with.

DWH Life Cycle – Business Requirements Def.

Kimball &Ross,2013 (..more details at pp.410-415)

Requirements Preplanning:

-Choose the Forum

-Identify and Prepare the Requirements Team

-Select, Schedule, and Prepare Business Representatives

Collecting Business Requirements:

-Launch

-Interview Flow

-Wrap-Up

Conducting Data-Centric Interviews

Documenting Requirements

Prioritizing Requirements

(BP = business process)

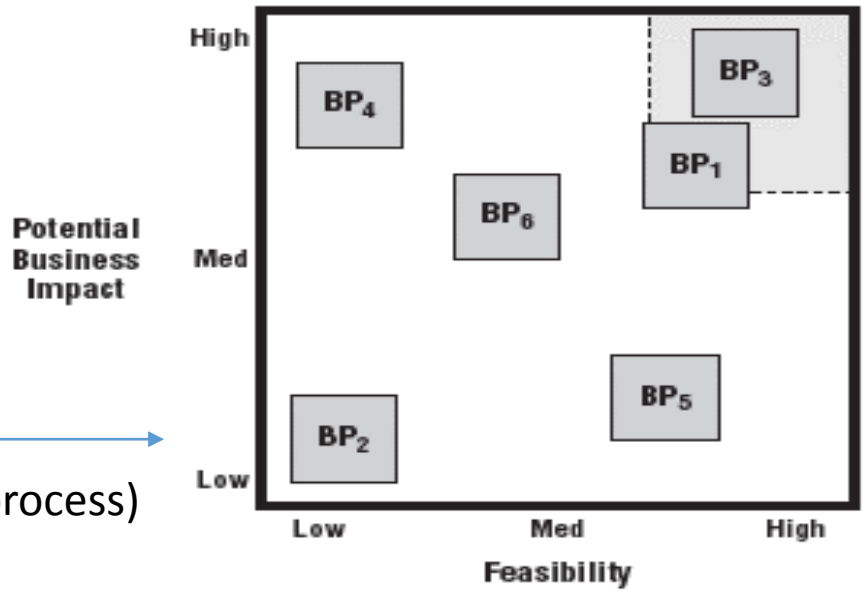


Figure 17-2: Prioritization grid based on business impact and feasibility.

DWH Life Cycle – Tech. Architecture Design

Kimball &Ross,2013 (..more details at pp.417-418)

Establish an Architecture Task Force

Collect Architecture-Related Requirements

Document Architecture Requirements

Create the Architecture Model

Determine Architecture Implementation Phases

Design and Specify the Subsystems

Create the Architecture Plan

Review and Finalize the Technical Architecture

DWH Life Cycle – Product Selection & Install.

Kimball &Ross,2013 (..more details at pp.418-420)

Understand the Corporate Purchasing Process

Develop a Product Evaluation Matrix

Conduct Market Research

Evaluate a Short List of Options

If Necessary, Conduct a Prototype

Select Product, Install on Trial, and Negotiate

DWH Life Cycle – Dim. Modeling & more..

Kimball &Ross,2013 (..more details at pp.420-424)

Physical Design:

Develop Naming and Database Standards

Develop Physical Database Model

Develop Initial Index Plan

Design Aggregations, Including OLAP Database

Finalize Physical Storage Details

ETL Design and Development

BI Application Specification

BI Application Development

DWH Life Cycle – Deploy, Maintain & Grow

Kimball &Ross,2013 (..more details at pp.424)

In the case of DW/BI deployment, the data is the main entrée. “Cooking” the data in the ETL kitchen is the most unpredictable task.

Unfortunately, even if the data isn’t fully cooked, you often still proceed with the DW/BI deployment because you told the warehouse guests they’d be served on a specific date and time.

Because you’re unwilling to slow down the pace of deployment, you march into their offices with undercooked data. No wonder users sometimes refrain from coming back for a second helping.

DWH Life Cycle – Deploy, Maintain & Grow

Kimball &Ross,2013 (..more details at pp.425)

You need to continue to manage the existing environment by investing resources in the following areas:

■ **Support.** User support is immediately crucial following the deployment to ensure the business community gets hooked. You can't sit back in your cubicle and assume that no news from the business community is good news. If you're not hearing from them, chances are no one is using the DW/BI system. Relocate (at least temporarily) to the business community so the users have easy access to support resources. If problems with the data or BI applications are uncovered, be honest with the business to build credibility while taking immediate action to correct the problems. If the DW/BI deliverable is not of high quality, the unanticipated support demands for data reconciliation and application rework can be overwhelming.

■ **Education.** You must provide a continuing education program for the DW/BI system. The curriculum should include formal refresher and advanced courses, as well as repeat introductory courses. More informal education can be offered to the developers and power users to encourage the interchange of ideas.

■ **Technical support.** The DW/BI system needs to be treated as a production environment with service level agreements. Of course, technical support should proactively monitor performance and system capacity trends. You don't want to rely on the business community to tell you that performance has degraded.

■ **Program support.** The DW/BI program lives on beyond the implementation of a single phase. You must closely monitor and then market your success. Communication with the varied DW/BI constituencies must continue. You must also ensure that existing implementations continue to address the needs of the business. Ongoing checkpoint reviews are a key tool to assess and identify opportunities for improvement.

DWH Life Cycle – Common errors to avoid

while building a DW/BI system

Kimball &Ross,2013 (..more details at pp.426)

■ **Pitfall 10:** Become overly enamored with technology and data rather than focusing on the business's requirements and goals.
■ **Pitfall 9:** Fail to embrace or recruit an influential, accessible, and reasonable senior management visionary as the business sponsor of the DW/BI effort.
■ **Pitfall 8:** Tackle a galactic multiyear project rather than pursuing more manageable, although still compelling, iterative development efforts.
■ **Pitfall 7:** Allocate energy to construct a normalized data structure, yet run out of budget before building a viable presentation area based on dimensional models.
■ **Pitfall 6:** Pay more attention to back room operational performance and ease of-development than to front room query performance and ease of use.
■ **Pitfall 5:** Make the supposedly query-able data in the presentation area overly complex. Database designers who prefer a more complex presentation should spend a year supporting business users; they'd develop a much better appreciation for the need to seek simpler solutions.
■ **Pitfall 4:** Populate dimensional models on a standalone basis without regard to a data architecture that ties them together using shared, conformed dimensions.
■ **Pitfall 3:** Load only summarized data into the presentation area's dimensional structures.
■ **Pitfall 2:** Presume the business, its requirements and analytics &the underlying data and the supporting technology are static.
■ **Pitfall 1:** Neglect to acknowledge that DW/BI success is tied directly to business acceptance. If the users haven't accepted the DW/BI system as a foundation for improved decision making, your efforts have been exercises in futility.

Bibliography

-Airinei, D., Depozite de date, Polirom, Iaşi, 2002
portal.feaa.uaic.ro/Master/sia/Pages/default.aspx
-Airinei, D., Dospinescu, O., Huiban, A., Aplicatii practice cu sisteme OLAP si Depozite de date, Editura Sedcom Libris, Iaşi, 2008
-Homocianu, D., Sistemele de asistare a deciziilor in contextual societatii cunoasterii, Editura UAIC, Iasi, 2009 (ssrn.com/abstract=2384380)
-Inmon,W.H.,Linstedt, D., Data Architecture: A primer for the data scientist, MK, MA, 2015 (tinyurl.com/h7dcq66)
-Kimball, R., Ross, M., The Data Warehouse Toolkit Third Edition. The Definitive Guide to Dimensional Modeling, John Wiley & Sons, New York, 2013 (tinyurl.com/jogo5uy)
-Kimball, R., Ross, M., The Kimball Group Reader. Relentlessly Practical Tools for Data Warehousing and Business Intelligence – Remastered Collection, Second Edition, Wiley, New York, 2016 (tinyurl.com/johox4v)
-Krishnan, K, Data Warehousing in The Age of Big Data, Morgan Kaufmann (MK), MA, 2013 (tinyurl.com/zrjo75j)
-Reeves, L.L., A Manager's Guide to Data Warehousing, Wiley, New York, 2009 (tinyurl.com/htsslk8)
-Sarka, D., et. al., Implementing a Data Warehouse with Microsoft SQL Server 2012. Training Kit, O'Reilly Media, Sebastopol, 2012 (tinyurl.com/jk6lclk)
-Sheldon, B., et. al., Professional Visual Basic 2012 and .NET 4.5 Programming, John Wiley & Sons, Indianapolis, 2013 (tinyurl.com/hrb6j9u)