Muni University

BACHELOR OF INFORMATION SYSTEMS DEGREE

Year 3, Semester 2

Course Name: **Data Warehousing and Business Intelligence**

Course Code: **ISM 3203**

NKAMWESIGA Nicholas

Lecture Content

- **Course Objectives**
- **Reading List**
- **Introduction to the Kimball Lifecycle Approach**
  - Life cycle history

  - Lifecycle milestone

  - Using lifecycle roadmap

  - Lifecycle navigation aids

  - Lifecycle vocabulary

Course Objectives

Students will learn to:

- Explain how a data warehouse combined with good business intelligence can increase a company's bottom line.

- Describe the components of a data warehouse.

- Describe different forms of business intelligence that can be gleaned from a data warehouse and how that intelligence can be applied toward business decision-making.

- Develop dimensional models from which key data for critical decision-making can be extracted.

- Sketch out the process for extracting data from disparate databases and data sources, and then transforming the data for effective integration into a data warehouse.

- Load extracted and transformed data into the data warehouse.

Reading List

- Ralph Kimball, Margy Ross, Warren Thornthwaite, Joy Mundy and Bob Becker. *The Data Warehouse Lifecycle Toolkit*. 2nd ed., Wiley, 2008.

- Richard T. Watson. *Data Management: Databases and Organizations*. 5th ed., Wiley, 2005.

- Ralph Kimball and Margy Ross. *The Data Warehouse Toolkit: The Complete Guide to Dimensional Modeling*. 2nd ed., Wiley, 2002.

- Thomas C. Hammergren. *Data Warehousing For Dummies*. 2nd ed., For Dummies, 2009.

- Swain Scheps. *Business Intelligence For Dummies*. For Dummies, 2008.

- Jiawei Han and Micheline Kambler. *Data Mining: Concepts and Techniques*. 2nd ed., Morgan Kaufmann, 2006.

- Ian H. Witten and Eibe Frank. *Data Mining: Practical Machine Learning Tools and Techniques with Java Implementation.* 2nd ed., Morgan Kaufmann, 2005.

- Galit Shmueli, Nitin R. Patel and Peter C. Bruce. *Data Mining for Business Intelligence: Concepts, Techniques, and Applications in Microsoft Office Excel with XLMiner*. WileyInterscience, 2006.

- Robert Stackowiak, Joseph Rayman and Rick Greenwald. *Oracle Data Warehousing and Business Intelligence Solutions*. Wiley, 2007.

- Alan R. Simon and Steven L. Shaffer. *Data Warehousing And Business Intelligence For eCommerce*. 1st ed., Morgan Kaufmann, 2001.

Introduction to the Kimball Lifecycle Approach

Life cycle history

- The ***Kimball Lifecycle*** provides the overall framework that ties together the various activities of a DW/BI implementation.

- ***1980s:*** the Kimball methodology first took root at Metaphor Computer Systems.
    - Metaphor was a pioneering decision support vendor; its hardware/software product offering was based on LAN technology with a relational database server and graphical user interface client built on a 32-bit OS.
    - Analysts of large corporations used Metaphor to build queries and download results into spreadsheets and graphs.

- ***1984:*** Ralph Kimball et al worked together to implement decision support solutions during the early days at Metaphor.
    - There were no industry best practices or formal methodologies.
    - The sequential steps of decision support existed. Their training manual at that time described the steps as: ***extract***, ***query***, ***analysis*** and ***presentation***.

Cont…

Life cycle history…

- Ralph Kimball et al together with Metaphor colleagues honed new techniques and approaches to deal with the idiosyncrasies of decision support.
- They modified and enhanced the traditional development methodologies to address the challenges of providing data access and analytics to business users, while considering growth and extensibility over time.

- ***1998:*** All the lesions learned from the previous successes and failures of their methodologies were described in their first book ***The Data Warehouse Lifecycle Toolkit.***

- The book's name was officially changed to ***Business Dimensional Lifecycle.***
  - The belief was that successful data warehousing depends on the following fundamental concepts:
    - **Focus on the business**
    - **Dimensionality structure of the overall data that's delivered to the business via ad hoc queries or reports.**
    - **Iteratively develop the overall data warehouse environment in manageable lifecycle increments and not through use of big bang approaches.**
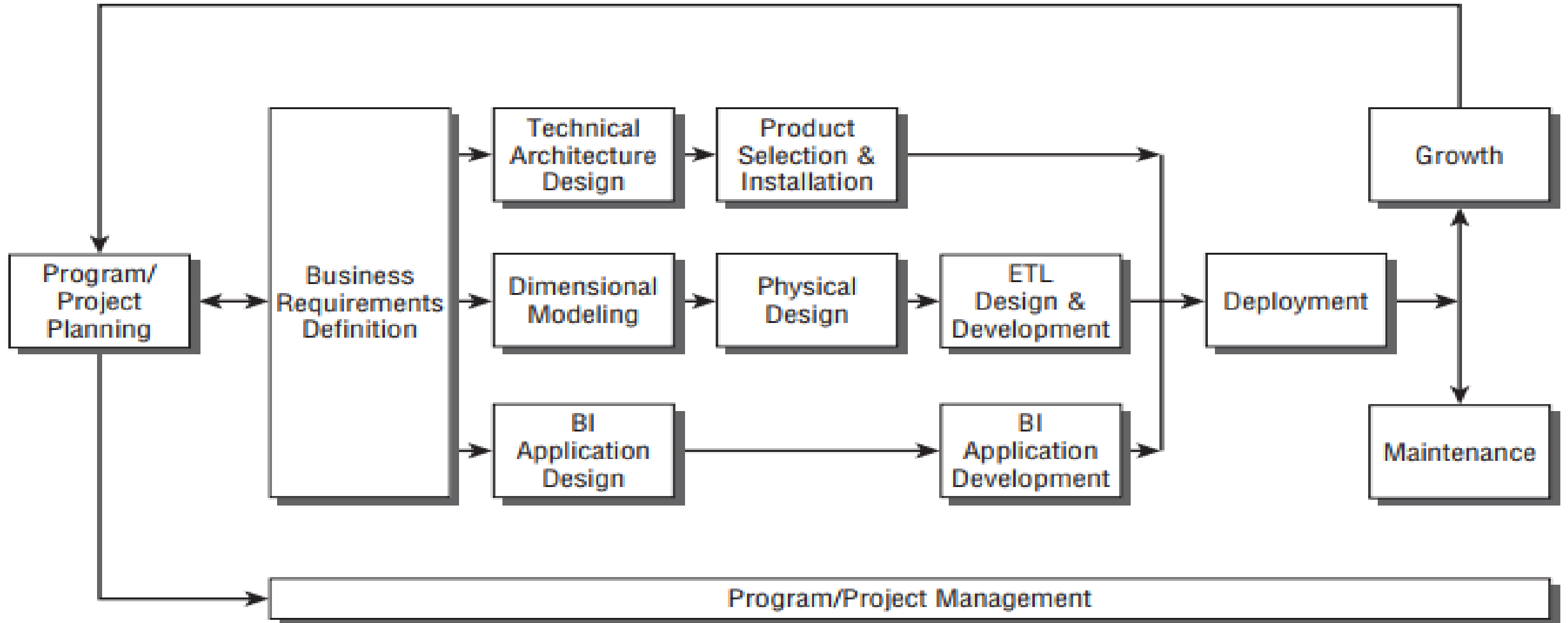
Cont...

Life cycle history...

- **2008:** Many users of Business Dimensional lifecycle, in the industry, had always referred to it as Kimball approach and this prompted the official name to change to Kimball Lifecycle.

- The basic constructs of Kimball Lifecycle have remained the same over time, their approach to designing, developing, and deploying DW/BI solutions has been tried and is true. The approach has been tested with numerous projects across virtually every industry, business function and platform. It has proven to work and work again. This formed the basis for the Kimball group's motto, i.e. **practical techniques, proven results**.

Cont...

Lifecycle milestone

Fig. 1: Kimball Lifecycle representation

Cont…

Lifecycle milestone…

- The overall *Kimball Lifecycle* approach is shown in Fig. 1. The diagram illustrates task sequence, dependency, and concurrency required for effective DW/BI design, development, and deployment. It serves as a roadmap to help teams do the right thing at the right time.

- Successful implementation of a DW/BI system depends on the appropriate integration of numerous tasks and components. The approach ensures that the project pieces are brought together in the right order and at the right time.

NB:

- The **Kimball Lifecycle** approach and agile methodologies share some common characteristics: **Focus on business value**, **collaborate with the business**, and **develop incrementally**.

- It is recommended that DW/BI system design and development needs to be built on a solid data architecture and governance foundation, driven by the bus architecture.

- It is recommended that many situations warrant the bundling of multiple agile *deliverables* into a more full-function release before being broadly deployed to the general business community.

Cont…

Lifecycle milestone…

- ***Program/Project Planning***
    - This is the first step of Lifecycle.
    - A *project* refers to a single iteration of Kimball Lifecycle from launch through deployment. Projects are known to have finite start and end.
    - *A Program* refers to the broader, ongoing coordination of resources, infrastructure, timelines, and communication across multiple projects. A program is, therefore, an overall umbrella of encompassing at least one projects. It should continuously renew itself and shouldn't have an abrupt end.
    - DW/BI initiative begins with a series of program and project planning activities.
    - Assessing Readiness
        - The most critical readiness factor is to have a strong executive business sponsor.
        - The 2nd readiness factor is having a strong, compelling business motivation for tackling the DW/BI initiative.
        - The 3rd factor when assessing readiness is feasibility: *technical & data* feasibility.

Cont…

Lifecycle milestone…

- *Program/Project Planning…*
  - *Scoping and Justification*
    - After being satisfied with the organizational 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.
    - Initially tackle projects that focus on data from a single business process; save the more challenging, cross-process projects for a later phase.
    - When scoping avoid; 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.
  - *Staffing*
    - DW/BI projects require the integration of a cross-functional team with resources from both the business and IT communities.

Cont...

Lifecycle milestone...

- ### *Program/Project Management*
    - It ensures that *Kimball Lifecycle* activities remain on track and on sync.
    - Program/ Project management activities focus on motoring the project status, issue tracking, and change control to preserve scope boundaries.
        - DW/BI projects are vulnerable to scope creep largely due to a strong need to satisfy business users' requirements. There are 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.
    - It involves the establishment of a comprehensive communication strategy that addresses both businesses and IT functions of the project.
        - Continuous communication is critical in managing expectations which is critical to achievement of project/program goal(s).

Cont…

Lifecycle milestone…

- ***Business requirements definition***
  - Collaborate with business users to understand their requirements and ensure their buy-in is absolutely essential to successful DW/BI. This section focuses on back-to-basic techniques for gathering business requirements.
  - ***Requirements Preplanning***
    - Before sitting down with business representatives to collect their requirements, consider the following to have productive sessions:
      - Choose the Forum
      - Select, Schedule, and Prepare Business Representatives
  - ***Collecting Business Requirements***
    - At this stage, one sits down face-to-face with business representatives to gather the business's requirements. The process usually flows from an introduction through structured questioning to a final wrap-up.
 - **Documenting Requirements**
    - Document what you've heard

Cont…

Lifecycle milestone…

- ## *Business requirements definition*

  - ### **Prioritizing Requirements**

    - prioritize using the prioritization grid shown in Fig. 2. Projects that warrant immediate attention are located in the upper-right corner because they're high-impact projects, as well as highly feasible

Fig. 2: Prioritization grid based on business impact and feasibility
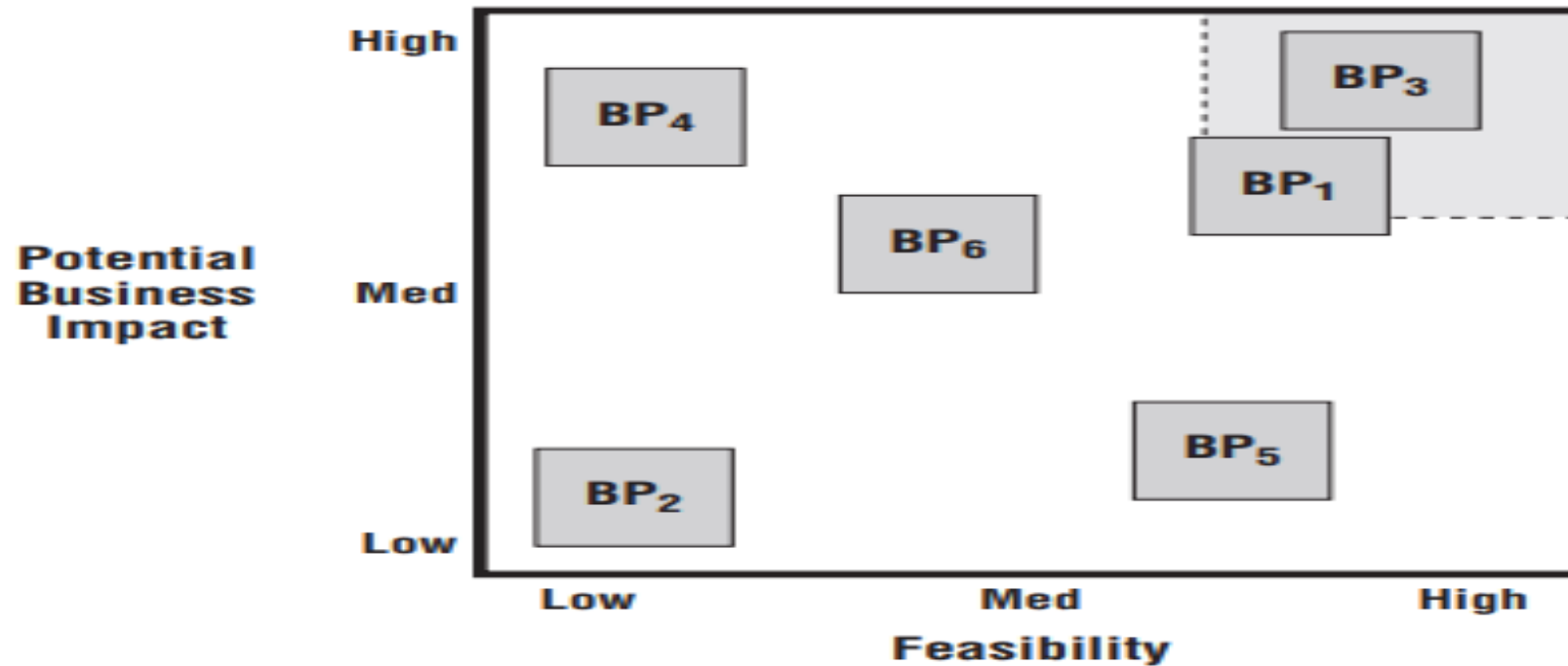
Cont...

Lifecycle milestone...

- ***Life Cycle Technology Track***

  - Technology track;

  - Data track; and

  - Business intelligence applications track

- ***Technical Architecture***

  - This is a blueprint for the DW/BI environment's technical services and infrastructure.

  - The technical architecture design establishes the overall architectural framework and vision.

  - **3 Factors**: *business requirements*; *currents technical environment*; and *planned strategic technical directions*, must be considered simultaneously to establish the appropriate DW/BI technical design.

Cont…

Lifecycle milestone…

- *Life Cycle Technology Track…*

- ***Technical Architecture…***

  - The technical architecture consists of a series of models that unveil greater detail regarding each major component.

    - The architecture enables you to catch problems on paper and minimize mid-project surprises.

    - It supports the coordination of parallel efforts while speeding development through the reuse of modular components.

    - The architecture identifies immediately required components versus those that will be incorporated at a later date.

    - Most important, the architecture serves as a communication tool. The DW/BI technical architecture supports communication regarding a consistent set of technical requirements within the team, upward to management, and outward to vendors.

Cont...

Lifecycle milestone...

- *Life Cycle Technology Track...*

- ***Technical Architecture...***

  - 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

Cont...

Lifecycle milestone...

- *Life Cycle Technology Track...*

- ***Product Selection and Installation***

- The technical architecture in the previous step acts a shopping list of needed capabilities.

- Architectural components such as hardware platform, database management system, extract-transformation-load (ETL) tool, or data access query and reporting tool are evaluated and selected.

- After selection, products are installed and tested to ensure appropriate end-to-end integration within your DW/BI environment.

Cont…

Lifecycle milestone…

- *Life Cycle Technology Track…*

- ***Product Selection and Installation…***

- The following six tasks associated with DW/BI product selection are quite similar to any technology selection.

  - 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

Cont…

Lifecycle milestone…

- *Life Cycle Data Track…*

- ***Dimensional Modeling***

  - During the gathering of business requirements, the organization's data needs are determined and documented in preliminary *enterprise data warehouse bus matrix* representing the organization's key business processes and their associated dimensionality.

    - The matrix serves as a data architecture blueprint to ensure that the DW/BI data can be integrated and extended across the organization over time.

  - Designing dimensional models to support the business' reporting and analytical needs requires a different approach than that used for transaction processing design.

    - By following a more detailed data analysis of a single business process matrix row, modelers identify the fact table granularity, associated dimensions, and attributes, and numeric facts.

Cont…

Lifecycle milestone…

- *Life Cycle Data Track…*

- ***Physical design***

  - The dimensional models developed and documented via a preliminary source-to target mapping need to be translated into a physical database.

    - With dimensional modeling, the logical and physical designs bear a close resemblance; you don't want the database administrator to convert the dimensional schema into a normalized structure during the physical design process.

  - Physical database implementation details vary widely by platform and project. In addition, hardware, software, and tools are evolving rapidly, so the following physical design activities and considerations merely scratch the surface.

    - **Develop Naming and Database Standards**

      - Table and column names are key elements of the users' experience, both for navigating the data model and viewing BI applications, so they should be meaningful to the business.

      - You must also establish standards surrounding key declarations and the permissibility of nulls.

Cont...

Lifecycle milestone...

- *Life Cycle Data Track...*

- ***Physical design...***

    - **Develop Physical Database Model**

        - This model should be initially built in the development server where it will be used by the ETL development team.

        - There are several additional sets of tables that need to be designed and deployed as part of the DW/BI system, including staging tables to support the ETL system, auditing tables for ETL processing and data quality, and structures to support secure access to a subset of the data warehouse

    - **Develop Initial Index Plan**

    - **Design Aggregations, Including OLAP Database**

    - **Finalize Physical Storage Details**

Cont...

Lifecycle milestone...

- *Life Cycle Data Track...*

- **ETL Design and Development**

  - The Lifecycle's data track wraps up with the design and development of the ETL system.

  - Design and development of the extract, transformation, and load (ETL) system remains one of the most vexing challenges confronted by a DW/BI project team; even when all other tasks have been well planned and executed, a great percentage of the risk and effort in the DW/BI project comes from this step

Cont...

Lifecycle milestone...

- *Lifecycle BI Applications Track...*

- **BI Application Design/ Specification**

  - While some DW/BI team members are working on the technical architecture and dimensional models, others should be working with the business to identify the candidate BI applications, along with appropriate navigation interfaces to address the users' needs and capabilities.

    - BI applications help to deliver business value from the DW/BI solution, rather than just delivering the data.

  - Before you start designing the initial applications, it's helpful to establish standards, such as common pull-down menus and consistent output look and feel.

    - Using these standards, you specify each application template and capture sufficient information about the layout, input variables, calculations, and breaks, so both the application developer and business representatives share a common understanding.

  - During the BI application specification activity, you should also consider the applications' organization. You need to identify structured navigational paths to access the applications, reflecting the way users think about their business.

    - Leveraging customizable information portals or dashboards are the dominant strategies for disseminating access.

Cont…

Lifecycle milestone…

- *Lifecycle BI Applications Track…*

- **BI Application Development**

    - Using the BI application specification, application development tasks include configuring the business metadata and tool infrastructure, and then constructing and validating the specified analytic and operational BI applications, along with the navigational portal.

    - The BI application template specifications should be revisited to account for the inevitable changes to the model since the specifications were completed.

    - The BI application quality assurance activities cannot be completed until the data is stabilized. You must ensure there is adequate time in the schedule beyond the final ETL cutoff to allow for an orderly wrap-up of the BI application development tasks.

Cont…

Lifecycle milestone…

- ***Deployment***
    - The technology, data, and BI application tracks converge at deployment.
    - The convergence does not happen naturally but requires substantial preplanning. Perhaps more important, successful deployment demands the courage and willpower to honestly assess the project's preparedness to deploy.
        - Deployment is similar to serving a large holiday meal to friends and relatives. It can be difficult to predict exactly how long it will take to cook the meal's main entrée. Of course, if the entrée is not done, the cook is forced to slow down the side dishes to compensate for the lag before calling everyone to the table.
    - For 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.

Cont…

Lifecycle milestone…

- *Deployment…*
    - Although testing has undoubtedly occurred during the DW/BI development tasks, you need to perform end-to-end system testing, including:
        - data quality assurance;
        - operations processing;
        - Performance; and
        - usability testing.
    - In addition to critically assessing the readiness of the DW/BI deliverables, you also need to package *it with education and support for deployment*. Because the user community must adopt the DW/BI system for it to be deemed successful, education is critical.
    - The DW/ BI support strategy depends on a combination of management's expectations and the realities of the deliverables. Support is often organized into a **tiered structure: T**he first tier is *website and self-service support*; the second tier is provided by *the power users residing* in the business area; *centralized support from the DW/BI team* provides the final line of defense.

Cont…

Lifecycle milestone…

- ## *Maintenance and Growth*

  - After deployment, 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

    - *Education:* You must provide a continuing education program for the DW/ BI system

    - *Technical support:* The DW/BI system needs to be treated as a production environment with service level agreements.

    - *Program support:* The DW/BI program lives on beyond the implementation of a single phase.

Cont…

Using lifecycle roadmap

- The general flow of DW/BI implementation is illustrated in the Kimball Lifecycle as shown in Fig. 1. it identifies task sequencing and highlights the activities that should happen concurrently throughout the technology, data, and BI application.

- The diagram does not reflect an absolute timeline; although the boxes are equally wide, there's a vast difference in the time and effort required for each major activity.

- With many approaches, one may need to customize the Kimball Lifecycle to address the unique needs of his/her organization.

- The Kimball is the most effective when used to implement projects of manageable, yet meaningful scope. It is nearly impossible to tackle everything at once, so don't let your business users, fellow team members, or management for that approach.

Cont…

Lifecycle navigation aids

- The book is riddled with references to the Kimball Lifecycle.

- Each chapter title page includes a miniature graphic of the Life cycle diagram, highlighting where you are within the overall framework.

- It should be noted that there's no direct (1:1) relationship between the chapters and Lifecycle phases. In some cases, a single chapter may cover at more than one Lifecycle phases and one Lifecycle phase may be covered in more than one chapters.

- At the end of each process-oriented chapter, the following guidance and recommendations are included:

  - Managing the effort and reducing risk

  - Assuring quality

  - Key project team roles involved in the process

  - Key deliverables

  - Estimating guidelines

  - Templates and other resources available on the common companion book website at www.kimballgroup.com

  - Detail grouping of project tasks.

Cont…

Lifecycle vocabulary

- ### *Data warehouse Vs. Business Intelligence*

  - Data ware house is the overall process of providing information to support business decision. The process include entire end-to-end solution, from the source extracts to the queries and applications that business users interact with i.e. presentation and access capabilities.

  - Business intelligence refers to the reporting and analysis of data stored in the data warehouse.

  - The two terms are abbreviated as DW/BI to emphasize the strong dependence on each other.

Cont...

Lifecycle vocabulary...

- ***ETL System***
    - This serves as a back room kitchen of the DW/BI.
    - It collects raw data from operational source systems and transform it into meaningful information for the business. Incoming data is checked for reasonable for quality; data quality conditions are continuously mo9nitored.

- **Business Process Dimensional model**
    - These fundamental front room deliverable.
    - Is a design discipline optimized to deliver on the twin goals of business users' ease of use and BI query performance.
    - Dimensional models contain the same content and relationships as models normalized into third normal form; they are just structured differently.

- **Business Intelligence Application**
    - These perform work querying, analyzing and presentation of information from the dimensional model.
    - The title substitutes the one previously know as *end user applications.*