C - Building the Data Warehouse

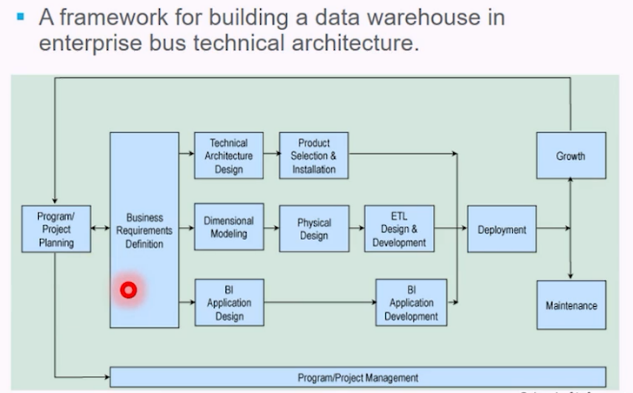
Development methodologies – User-centric: Data warehouses are about the organization’s data

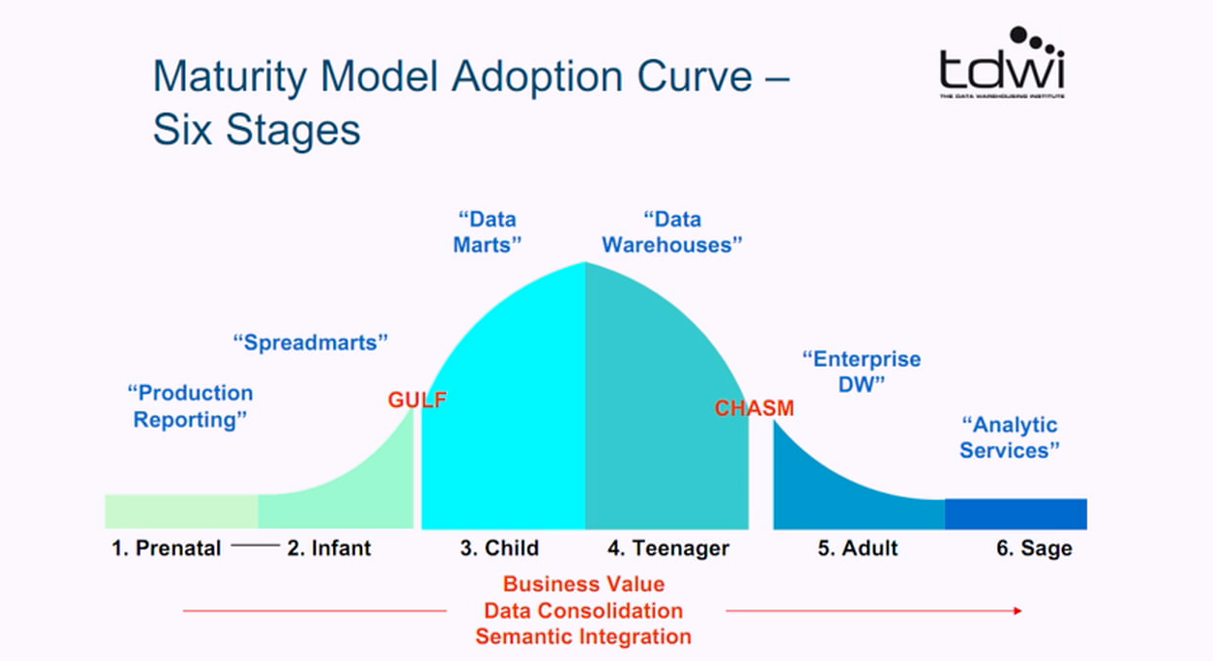
Data-centric: Data warehouses are about the people who use the data.

Two basic models:

1. Waterfall model: User-centric. Starts with the user’s needs and finds thee data within the organization to satisfy them. Linear in nature, top-down approach. Starts with feasibility and requirements and ends with testing and deployment.
2. Spiral Model: Data-centric. Starts with the data in the organization and ends with the user’s needs. Iterative in nature and has a bottoms-up approach that starts with architecture. Re-iterates based on the user’s needs.

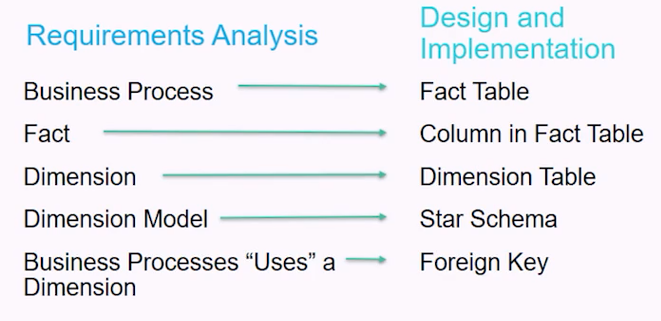
|  |  |
| --- | --- |
| Kimball | Inmon |
| User-centric | Data-centric |
| Enterprise bus TA DDS is user-facing | Hub and Spoke TA Nds is internal |
| Waterfall model | Spiral Model |
| Top-down  Program – Collection of coordinated projects. Several data marts will be created with conformed dimensions. E.g. Sales BI Program Program -> Project -> Data Mart | Bottom-up Project – Single Iteration of the entire cycle. Encompasses a business process resulting in a data mart. Each project team will work on one project within the same program. |





Event: A frequent activity within the business, also known as a business process which helps in identifying business transactions.   
Status: The condition of an object at a point in time, which helps to identify workflows.   
Level: Quantitative measurement of an object at a point in time, which helps to identify periodic snapshots.   
Roles: The who, what and when of the event, status, or level, which helps in understanding dimensions.  
Requirement gatherings: Interviews and observations of business users. Data audits, interview write-ups, identify business processes, enterprise bus matrix, prioritization grid and issues list.

1. How does dimensional modeling reflect business processes? What are dimensions in dimensional modeling? Video 3.6.6 Videos 4.2.1 – 4.3.1
   1. Dimensions provide context for facts. Identify dimensions using key words such as “by”, “for” etc.
   2. Dimensions have attributes that describe and categorize their values. E.g. student: major, year, dorms, gender.
   3. The attributes help constrain and summarize facts. 1 process, 1 fact table.
   4. Dimensional modelling: Fact table with the dimensions it requires.
      1. Logical design technique for structuring data with the following objectives:
         1. Intuitive: easy for business users to understand
         2. Fast: Speedy query performance
      2. Dimensional models are implemented in the relational DMBS as star schemas. In MOLAP databases they exist as cubes/data-mart.
   5. Components of dimensional model:
      1. Fact table: a database table of quantifiable performance measurements. FK’s are assigned to each of the dimensions. E.g. sales amount, days to ship.
      2. Dimension table: Table of contexts for the facts. E.g. date/time, location, customer.
      3. Attribute: Characteristics of the dimension. E.g. product: name, category, department.
      4. Star Schema: Connection among facts and dimensions of a business process. E.g sales, inventory management etc.



1. Explain the difference between a functional and non-functional requirement? Video 3.5.2
   1. Functional requirements – define what thee system does or should do. These requirements address needs of business users. Business users must be able to analyze sales of product over time. The functional requirements are used to describe the functionalities of a system. E.g.- An operating system requires that users enter a password and username when logging in so that the system can authenticate their identity.
   2. Non-functional requirements – Guide and constrains the system architecture, outline procedures, rules or regulations. Maximum query response time should be no longer than 20 seconds. The non-functional requirements describe the system quality characteristics or the quality attributes. E.g a system has added features that enhance usability.
2. What are the three business process types? Provide an example of each. Video 3.6.2

Business processes end up being fact tables in the ROLAP star schema. In the design phase they are called dimensional models. Business process types:

* 1. Events or transactions: Basic level. Easiest to identify. One row per line in a transaction. Corresponds to a point in space and time. Once inserted it is never updated. Rows are inserted when an event or transaction occurs. Eg. Sales, returns and telemarketing
  2. Workflows or Accumulating snapshots: Based on a status. Used to capture a business process workflow, fact row is initially inserted then updated as milestones occur. Fact table has multiple dates that correspond to each milestone and records a change in status. Special facts: milestone counters and lag facts for length of time between milestones. Measures if the event occurred and the lag-time between the two workflows. E.g Order fulfillment, job applicant tracking.
  3. Periodic snapshots: Predetermined intervals, snapshots of quantitative measurements are taken and stacked consecutively in the fact table. Taken either daily, weekly or monthly or hourly. Complements detailed transaction facts but does not replace them. Eg. Student’s GPA, financial report, bank account value.

1. What are the three types of facts? Provide an example of each. Video 3.6.4
   1. Facts are quantifiable numerical values associated with business process. How much, how many, how long and how often?
   2. Points scored in a game is a fact, player height is not a fact.
   3. Type of facts:
      1. Additive: fact can be summed across all dimensions. The most useful kind of fact. Quantities sold, hours billed.
      2. Semi-additive: Cannot be summed across all dimensions, such as time periods. Sometime these are averaged across the time dimension. E.g. time logged on to computer.
      3. Non-additive: Cannot be summed across any dimension. E.g. Basketball player height, retail price.
2. Come up with your own functional requirement for any business/department – state that as a

single question. Then, identify the business process, business process type, dimensions and

fact(s) based on that statement. Video 3.6.1

How many ceramic shipments arrived on time last week?

Arrival Time – Fact

Product type – Ceramic products

Business process – Count of items

Duration of time – Arrival time of the shipment