Please mute Your cell!
• Data Warehouse Architecture

• Approaches to build a Data Warehouse
  – Top Down Approach
  – Bottom Up Approach

• Best Practices

• Case Example
Okay, let’s shake this thing and see what we come up with.
Data warehousing is the **coordinated**, **architected**, and **periodic** copying of data from various **sources**, both inside and outside the enterprise, into a **centralized repository** optimized for **reporting** and **analytical processing**.
Data Warehouse Architecture

Source Layer

ERP
Other Data Sources
Social Media Content
Unstructured Content

Data Integration Layer

Data Profiling
Data Transformation
Data Cleansing
Data Validation
Data Auditing
Data Quality

Data Warehouse Layer

Staging/Operation Data Store

Enterprise Data Warehouse

Data Marts

Presentation Layer

BI Applications
Mobile
On Demand Services

Browsers
Data Warehouse Architecture

**Source Layer**
- Core Systems
- Unstructured / External Data
- Social Media

**Data Integration Layer**
- Extraction
- Cleansing
- Transformation
- Profiling
- Auditing
Data Warehouse Architecture

**Data Warehouse Layer**
- Staging Database
- Operational Data Store
- Data Warehouse
- Data Mart

**Presentation Layer**
- Reporting
- Dashboards
- Mobile Devices
Data Warehouse Approaches

Top Down Approach

- Suggested by Bill Inmon
- Build a centralized repository (DW) to house corporate wide business data
- The data in the DW is stored in a normalized form in order to avoid redundancy
- Once the DW is implemented, start building subject area specific data marts which contain summarized information based on the end users analytical requirements. This is to enable faster access to the data for the end users analytics
- Longer to implement - May fail due to the lack of patience and commitment

Bottom Up Approach

- Suggested by Ralph Kimball
- The bottom up approach is an incremental approach to build a data warehouse
- We build the data marts separately at different points of time as and when the specific subject area requirements are clear
- The data marts are integrated or combined together to form a data warehouse
- Separate data marts are combined through the use of conformed dimensions and conformed facts. A conformed dimension and a conformed fact is one that can be shared across data marts.
- Faster to implement - Implementation in stages
Data Warehouse Approaches

**Top Down Approach**

The requirements of users across different levels are gathered and one schema for the entire data warehouse is built. Afterwards, separate data marts are tailored according to the particular characteristics of each business area or process.

The top-down approach emphasizes the users’ requirements thus making it a more user driven approach.

**Bottom Up Approach**

A separate schema is built for each data mart, taking into account the availability of data for decision-making. Later, these schemas are merged, forming a global schema for the entire data warehouse.

The bottom-up approach aims at identifying all the star schemas that can be implemented using the available source systems making it a more source/data driven approach.
How to Choose?

<table>
<thead>
<tr>
<th></th>
<th>Favors Top Down</th>
<th>Favors Bottom up</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nature of the organization's decision support Requirements</td>
<td>Strategic</td>
<td>Tactical</td>
</tr>
<tr>
<td>Data Integration Requirements</td>
<td>Enterprise Wide integration</td>
<td>Individual Business areas</td>
</tr>
<tr>
<td>Nature of Source Data</td>
<td>High rate of change from source systems</td>
<td>Source systems are relatively stable</td>
</tr>
<tr>
<td>Time to Delivery</td>
<td>Organization's requirements allow for longer start-up time</td>
<td>Need for the first data warehouse application is urgent</td>
</tr>
<tr>
<td>Cost to Deploy</td>
<td>Higher start-up costs, with lower subsequent project development costs</td>
<td>Lower start-up costs, with each subsequent project costing about the same</td>
</tr>
<tr>
<td>Yesterday</td>
<td>Today</td>
<td></td>
</tr>
<tr>
<td>--------------------------------------------------------------------------</td>
<td>-----------------------------------------------------------------------</td>
<td></td>
</tr>
<tr>
<td>▪ Point-in-time business intelligence</td>
<td>▪ Real-time business intelligence</td>
<td></td>
</tr>
<tr>
<td>▪ Batch data warehousing</td>
<td>▪ Dynamic data warehouse environment</td>
<td></td>
</tr>
<tr>
<td>▪ Separation of data warehouse and transaction systems</td>
<td>▪ Consolidation of data warehouse and transaction systems</td>
<td></td>
</tr>
<tr>
<td>▪ Self-contained historical data warehouse</td>
<td>▪ Information integrated with other sources</td>
<td></td>
</tr>
<tr>
<td>▪ Latency in development and deployment of business applications</td>
<td>▪ Speed of delivery and development is critical</td>
<td></td>
</tr>
</tbody>
</table>
The Hybrid approach aims to harness the speed of the Bottom up approach and user orientation/integration of the Top down approach – Combination of User and Data Driven
• Best Practice #1
  – Use a Data model that is optimized for Information retrieval
    • Dimensional Model
    • Hybrid Approach

• Best Practice #2
  – Carefully design the data integration processes for your DW
    • Ensure the data is processed efficiently and accurately
    • Consider an ETL or Data Cleansing tool
    • Use them well!
Data Warehouse Architecture Best Practices

• Best Practice #3
  – Design a metadata architecture that allows sharing of metadata between components of your DW
    • Consider metadata standards such as OMG’s Common Warehouse Metamodel (CWM)

• Best Practice #4
  – Take an approach that consolidates data into ‘a single version of the truth’
    • Hybrid Model
    • OR ?

* OMG – Object Management Group
At a Glance:
• India's largest private bank
• Over 10 million customers
• 364 branches and 46 extension counters
• Network of 1,050 ATMs
• Multiple call center’s and Internet banking

• ICICI Bank considered various data warehousing and customer relationship management (CRM) solutions. Based on worldwide industry experience, Teradata emerged as the foremost solution
• The implementation of an enterprise data warehouse from Teradata allowed ICICI to analyse its customer database, which includes information from separate operational systems including retail banking, bonds, fixed deposits, credit cards, and more
Thank you for your attention!

Any Questions?