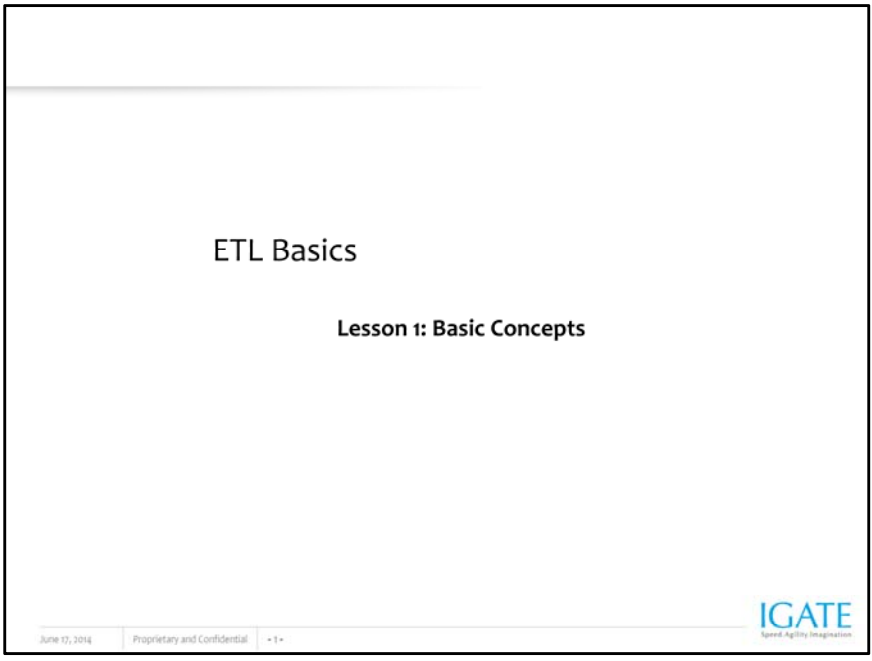

ETL Basics
Lesson 1: Basic Concepts
June 17, 2014
Proprietary and Confidential
- 1 -
IGATE
Speed. Agility. Imagination

Lesson Objectives

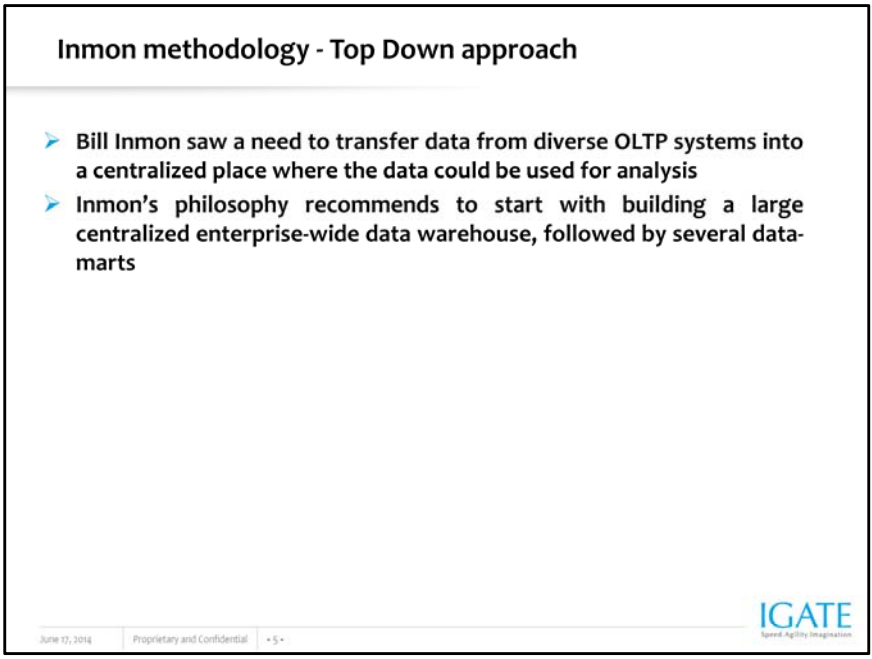
- **On completion of this lesson on ETL basics, you will be able to:**
  - Understand Data warehousing strategies and architecture
  - Know the meaning and need of ETL

Datawarehouse

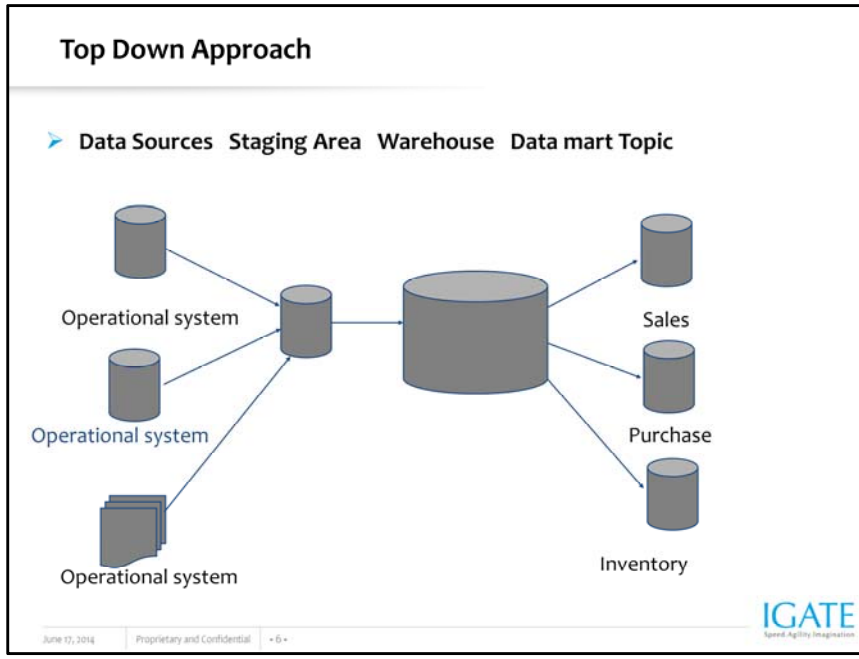
- A single, complete and consistent store of data obtained from a variety of different sources made available to end users in a format that they can understand and use in a business context.

Datawarehousing Strategies

- Enterprise-wide warehouse, top down, the Inmon methodology
- Data mart, bottom up, the Kimball methodology
- When properly executed, both result in an enterprise-wide data warehouse



The data marts are treated as sub sets of the data warehouse. Each data mart is built for an individual department and is optimized for analysis needs of the particular department for which it is created. The data flow in the top down OLAP environment begins with data extraction from the operational data sources. This data is loaded into the staging area and validated and consolidated. This data from the Staging area is then loaded in to the datawarehouse.

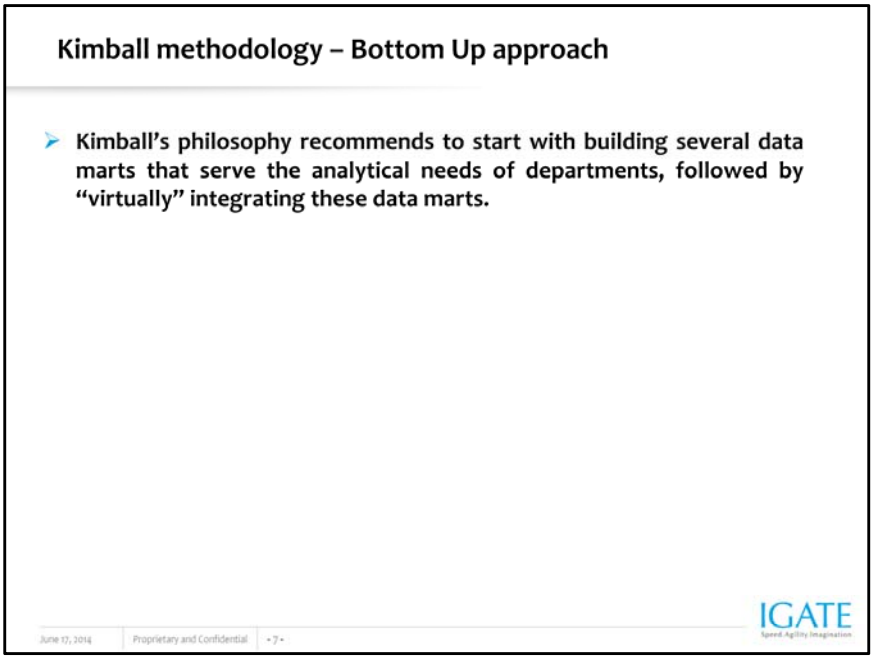


The data flow in the top down OLAP environment begins with data extraction from the operational data sources. This data is loaded into the staging area and validated and consolidated for ensuring a level of accuracy and then transferred to the Operational Data Store. (ODS). The ODS stage is sometimes skipped if it is a replication of the operational databases. Data is also loaded into the Data warehouse in a parallel process to avoid extracting it from the ODS.

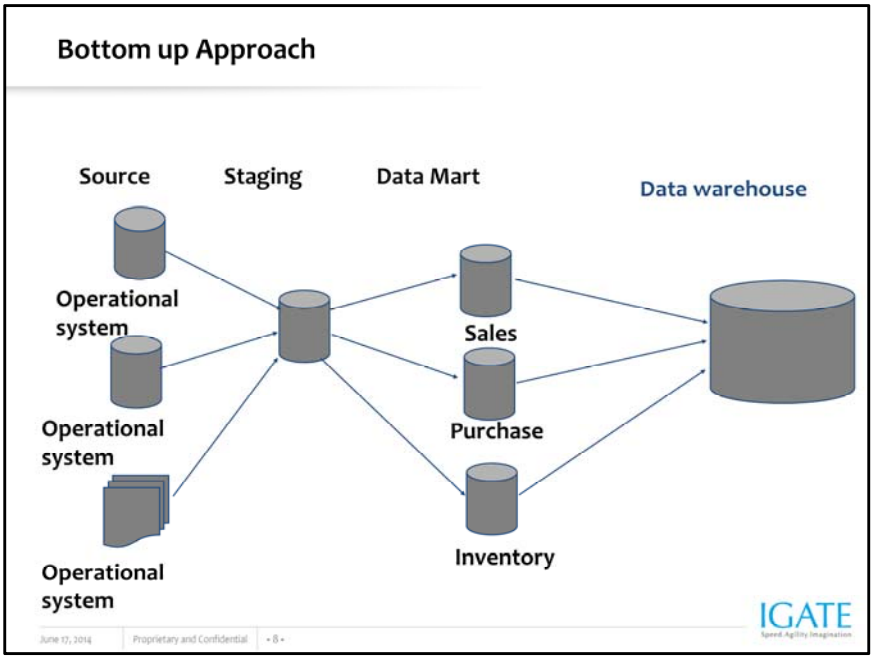
Detailed data is regularly extracted from the ODS and temporarily hosted in the staging area for aggregation, summarization and then extracted and loaded into the Data warehouse. The need to have an ODS is determined by the needs of the business. If there is a need for detailed data in the Data warehouse then, the existence of an ODS is considered justified. Else organizations may do away with the ODS altogether.

Once the Data warehouse aggregation and summarization processes are complete, the data mart refresh cycles will extract the data from the Data warehouse into the staging area and perform a new set of transformations on them. This will help organize the data in particular structures required by data marts. Then the data marts can be loaded with the data and the OLAP environment becomes available to the users.

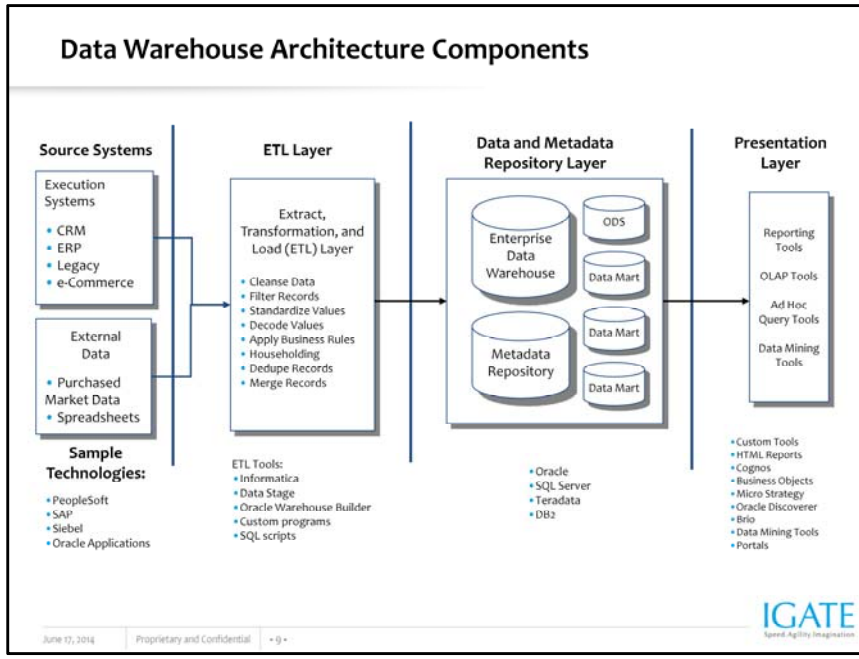
The data in a data warehouse is time variant in nature as it contains historical data. Inmon proposes a top-down model approach to create a centralized Enterprise Data Warehouse using traditional database modeling techniques (ER Model), where the data is stored in 3NF. The data warehouse acts as data source for the new data marts


Kimball methodology – Bottom Up approach
Kimball's philosophy recommends to start with building several data marts that serve the analytical needs of departments, followed by "virtually" integrating these data marts.
June 17, 2014
Proprietary and Confidential
- 7 -
IGATE
Speed. Agility. Imagination

.



The bottom-up approach reverses the positions of the Data warehouse and the Data marts. Data marts are directly loaded with the data from the operational systems through the staging area. The ODS may or may not exist depending on the business requirements.
The data flow in the bottom up approach starts with extraction of data from operational databases into the staging area where it is processed and consolidated and then loaded into the ODS. The data in the ODS is appended to or replaced by the fresh data being loaded. After the ODS is refreshed the current data is once again extracted into the staging area and processed to fit into the Data mart structure. The data from the Data Mart, then is extracted to the staging area aggregated, summarized and so on and loaded into the Data Warehouse and made available to the end user for analysis.


Data Warehouse Architecture Components
Source Systems
Execution Systems
• CRM
• ERP
• Legacy
• e-Commerce
External Data
• Purchased Market Data
• Spreadsheets
Sample Technologies:
• PeopleSoft
• SAP
• Siebel
• Oracle Applications
ETL Layer
Extract, Transformation, and Load (ETL) Layer
• Cleanse Data
• Filter Records
• Standardize Values
• Decode Values
• Apply Business Rules
• Householding
• Dedupe Records
• Merge Records
ETL Tools:
• Informatica
• Data Stage
• Oracle Warehouse Builder
• Custom programs
• SQL scripts
Data and Metadata Repository Layer
Enterprise Data Warehouse
ODS
Data Mart
Data Mart
Metadata Repository
Data Mart
• Oracle
• SQL Server
• Teradata
• DB2
Presentation Layer
Reporting Tools
OLAP Tools
Ad Hoc Query Tools
Data Mining Tools
• Custom Tools
• HTML Reports
• Cognos
• Business Objects
• Micro Strategy
• Oracle Discoverer
• Brio
• Data Mining Tools
• Portals
June 17, 2014
Proprietary and Confidential
• 9 •
IGATE
Speed. Agility. Imagination

What is ETL?

- ETL stands for Extract Transform & Load
- The process of updating the data warehouse
- ETL is the automated and auditable data acquisition process from source system that involves one or more sub processes of data extraction, data transportation, data transformation, data consolidation, data integration, data loading and data cleaning.

Need for ETL

- The process of ETL is required so that data from different heterogeneous sources can be combined and brought into one common source.
- The Advantage of having the process of ETL is that, as data from different sources can be brought together, highly complex and user friendly reports can be generated for decision making

Need for ETL

- Data stored in different formats in different types of databases
- Some data sources might be archives while others may be active operational systems
- Data extraction and cleansing - time-consuming and difficult Aggregation of data

Summary

- **In this module, you learned about the following:**
  - Datawarehousing strategies
  - Datawarehousing architecture
  - Need for ETL
  - Meaning of ETL

Add the notes here.