IST722
Data Warehousing

Components of the Data Warehouse

Michael A. Fudge, Jr.
Recall: Inmon’s CIF

The CIF is a reference architecture
Understanding the Diagram

The CIF is a reference architecture
CIF Components
The CIF is a reference architecture
External World & Applications

- **External World** – the people and systems that generate operational data.
- **Applications** – the systems which provide the source for the operational data.
- **Examples**: ERP’s, Business Applications, Internet data, external data streams.
- These are the **inputs** and **data sources** for the CIF.
- **OLTP Systems** – Operational data, transaction-oriented.
Integration & Transformation Layer

The CIF is a reference architecture
Integration & Transformation Layer

- **I&T Layer** - takes un-integrated data from multiple sources and integrates and consolidates it.
- Computer programs are written to transform data from the *external world* into *corporate data*.
- The data come from a variety of sources and in both *structured* and *un-structured* formats.
- Today’s Database Management Systems provide tooling to assist with this process.
- This is the most difficult and time-consuming component of the CIF.
- Two approaches: ETL and ELT
ETL – Extract Transform Load

- The data transformation occurs over staged data.
- The source data is not stored in the warehouse.
ELT – Extract Load Transform

- The data transformation occurs over warehoused data.
- The staged data is stored in the warehouse.
The CIF is a reference architecture
Operational Data Store

- **Integrated, detailed, and current** data from the External World and Applications.
- **Consolidated** from disparate sources.
- Does not grow over time.
- Performs similarly to a transactional database.
- Structured differently than a data warehouse, and therefore should be *stored as a separate database*.
- Receives data from I&T layer sends data to the **data warehouse**.
- The **data warehouse** can populate it, too.
- Think of it as a consolidated operational database.
Enterprise Data Warehouse

The CIF is a reference architecture
Enterprise Data Warehouse

- **Subject-oriented, integrated, summarized, and current** data from the External World and Applications.
- Optimized for *query performance*.
- Structured differently than operational data, typically in a *dimensional model*.
- Receives data from I&T layer and the ODS.
- Use as a source for *data marts* and *decision support systems*.
- Grows in size over time due to historical data.
- The heart of the CIF.
## ODS vs. EDW

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Operational Data Store</th>
<th>Data Warehouse</th>
</tr>
</thead>
<tbody>
<tr>
<td>Primary Purpose</td>
<td>Run the business on a current basis</td>
<td>Support managerial decision making</td>
</tr>
<tr>
<td>Design Goal</td>
<td>Performance throughput, availability</td>
<td>Easy reporting and analytics</td>
</tr>
<tr>
<td>Primary Users</td>
<td>Clerks, salespersons, administrators</td>
<td>Managers, business analysis, customers</td>
</tr>
<tr>
<td>Subject-Oriented</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Integrated</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Detailed Data</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Summary Data</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>Time of Data</td>
<td>Current data</td>
<td>Historical snapshots</td>
</tr>
<tr>
<td>Updates</td>
<td>Frequent small updates</td>
<td>Periodic batch updates</td>
</tr>
<tr>
<td>Queries</td>
<td>Simple queries on a few rows</td>
<td>Complex queries on several rows</td>
</tr>
</tbody>
</table>
Why No ODS in the EDW?

I need fast updates!

I need query performance!

You can’t have both!
The CIF is a reference architecture
Data Marts

- A collection of data tailored to the informational needs of a department or business process.
- Easy to control, low cost, and customizable due to their limited scope.
- Receive their inputs from the Enterprise Data Warehouse.
- Are source data for Online Analytical Processing (OLAP) engines.
OLAP

**ROLAP**
- Uses a **Relational Database Management System**
- Data design is the **Star Schema**
- Built on well-known relational concepts
- In the EDW.

**MOLAP**
- Uses a **Multi-Dimensional Database Management System**
- Data design is the **Cube**
- Highly flexible.
- Data Marts

Typical implementations have the ROLAP star schema feed the MOLAP cube.
ROLAP – Star Schema

- Stored in a **relational DBMS**
- Fact table is M-M relationship among dimensions.
MOLAP - Cube

- Stored in a Multi-Dimensional DBMS
- Facts are pre-aggregated across all dimensions for improved performance.
DSS Applications

The CIF is a reference architecture
Decision-Support Systems

- Business Intelligence.
- Front-ends to ROLAP and OLAP Engines.
- Help us explore and visualize information at a high level
In Summary...

- The CIF is a **reference architecture** for building out an information ecosystem.
- Applications from the external world are inputs into the CIF.
- The **Integration & Transformation Layer** transforms transactional data into corporate data.
- The **Operational Data Store** contains consolidated, non-historical data.
- The **Enterprise Data Warehouse** contains consolidated historical data.
- **Data marts** are tailored to the informational needs of a department or business process.
IST722
Data Warehousing

Components of the Data Warehouse

Michael A. Fudge, Jr.