



# Table of Contents

- *Recommended Tools and Standard*
- *The Data Warehouse Environment*
- *Data Quality Assurance*
- *Data Profiling using Axio*
- *Quality Manager*
- *Data Quality Methodology*
- *Data Quality Benefits*
- *Data Movement*
- *DataStage Architecture*
- *Metadata Management*

# Recommended Tools

Tools	Vendors
<b>Data Modeling</b>	<ul style="list-style-type: none"><li>• CA Erwin</li><li>• CA Model Validator</li></ul>
<b>Data Model Management</b>	<ul style="list-style-type: none"><li>• CA Model Manager</li></ul>
<b>Data Model Integration</b>	<ul style="list-style-type: none"><li>• MetaIntegration Bridge</li></ul>
<b>Database Design</b>	<ul style="list-style-type: none"><li>• CA ERWin</li><li>• Framework Manager</li><li>• Rose</li></ul>
<b>Database</b> <b>Online Transactional Processing</b> <b>Operational Data Store</b> <b>Data Warehouse</b> <b>Data Mart</b>	<ul style="list-style-type: none"><li>• DB2 UDB, Oracle, SQL Server</li><li>• Teradata, DB2UD, Oracle, SQL Server</li><li>• Teradata, DB2UDB, Oracle</li><li>• Teradata, DB2UD, Oracle, SQL Server</li></ul>

# Recommended Tools

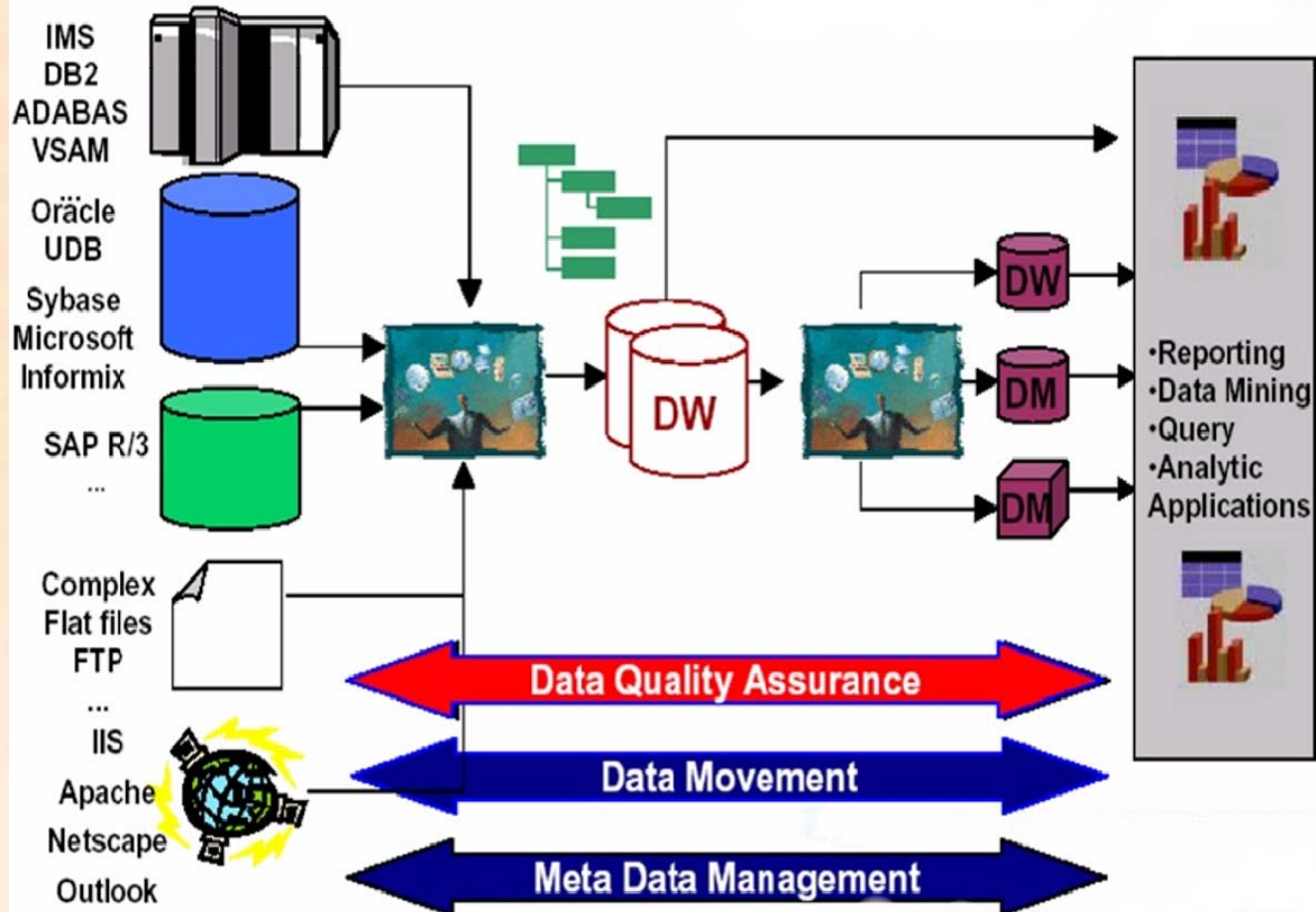
<b>Tools</b>	<b>Vendors</b>
<b>Data Quality Analysis and Metrics</b>	<ul style="list-style-type: none"><li>• QualityStage</li></ul>
<b>Extract-transform--load (ETL)</b>	<ul style="list-style-type: none"><li>• DataStage</li><li>• Informatica</li></ul>
<b>Data Profiling</b>	<ul style="list-style-type: none"><li>• AXIO</li><li>• ProfileStage</li></ul>
<b>Data Scrubbing</b>	<ul style="list-style-type: none"><li>• AuditStage</li></ul>
<b>Data Mapping</b>	<ul style="list-style-type: none"><li>• AXIO</li></ul>
<b>Reporting and Querying (Business Intelligence - OLAP)</b>	<ul style="list-style-type: none"><li>• Cognos ReportNet</li><li>• SAS</li><li>• Business Objects</li></ul>
<b>Metadata Management</b>	<ul style="list-style-type: none"><li>• MetaStage</li></ul>

# Data Warehouse Tools

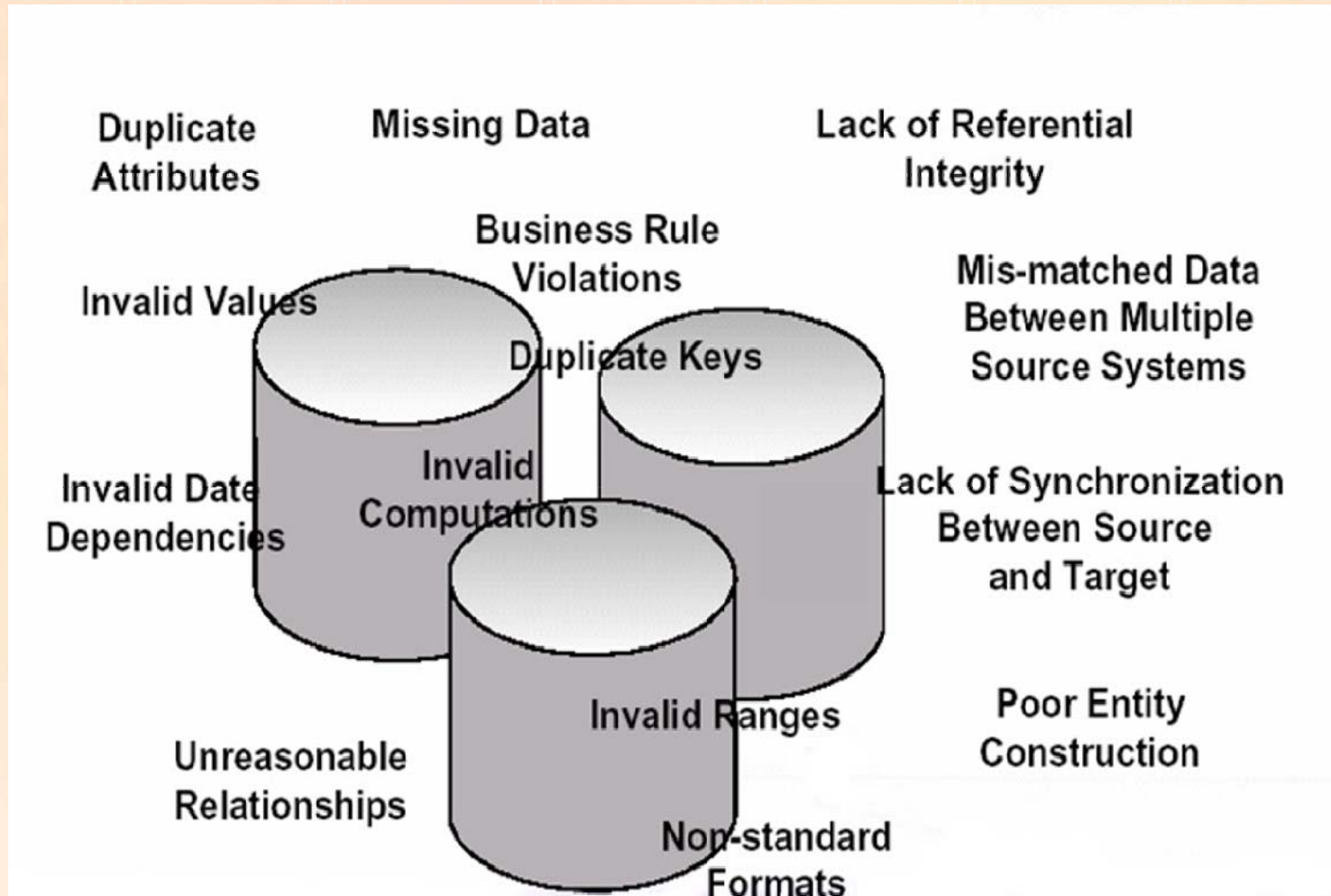
- *Data Quality Assurance*
  - *Data Profiling using **Evoke Axio***
  - *Data Quality Assessment using **Quality Manager***
  - *Data Cleansing using **Integrity***
- *Data Movement*
  - *Extract, Transform, Move, Load (ETML) using **DataStage***
- *Metadata Management*
  - *Capture ETML Design Metadata, ETML Process Metadata using **MetaStage***
- *Data Modeling*
  - *Creating Logical Data Model and Physical Data Model using **ERWin***



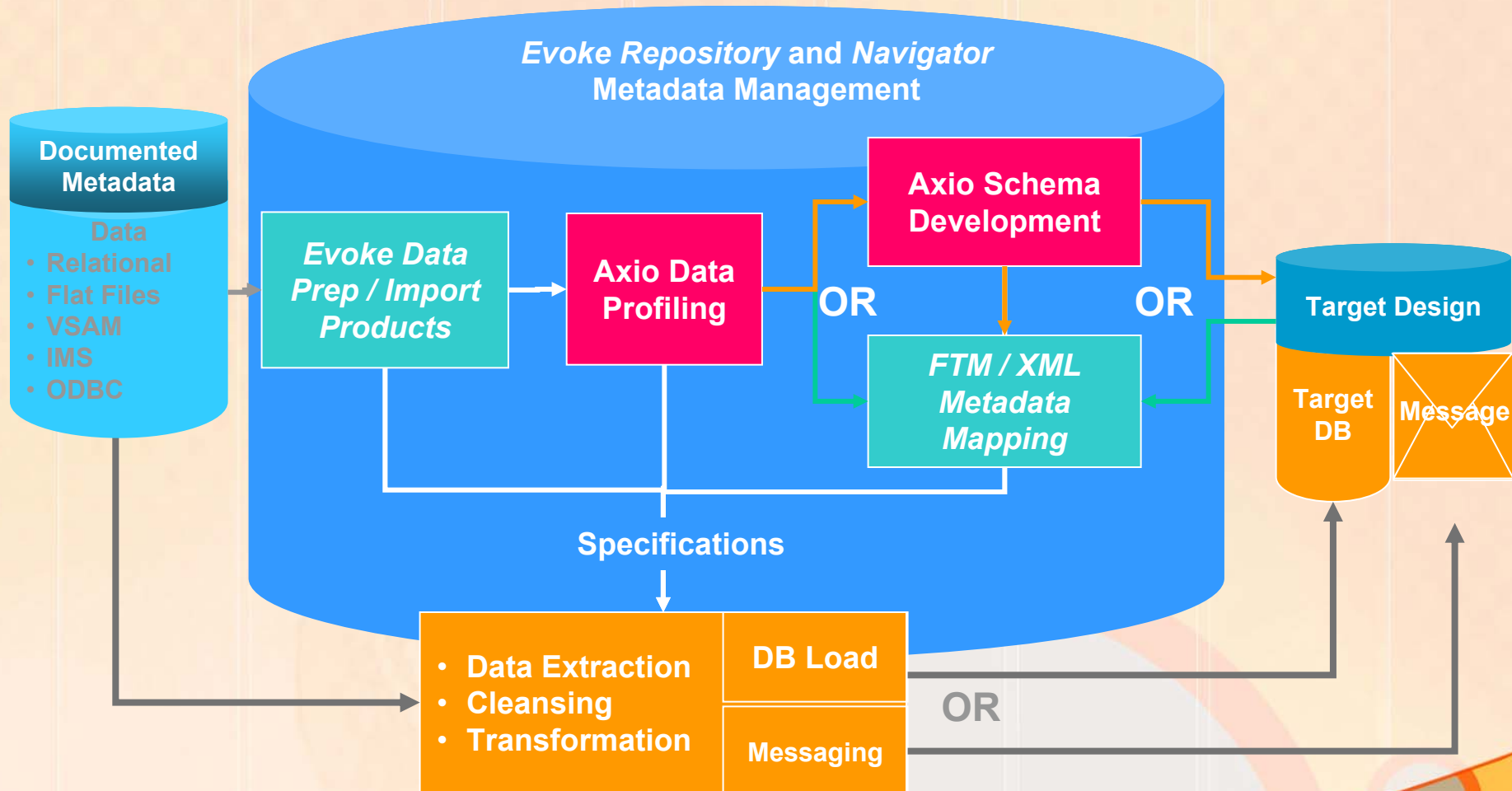
# The Data Warehouse Environment



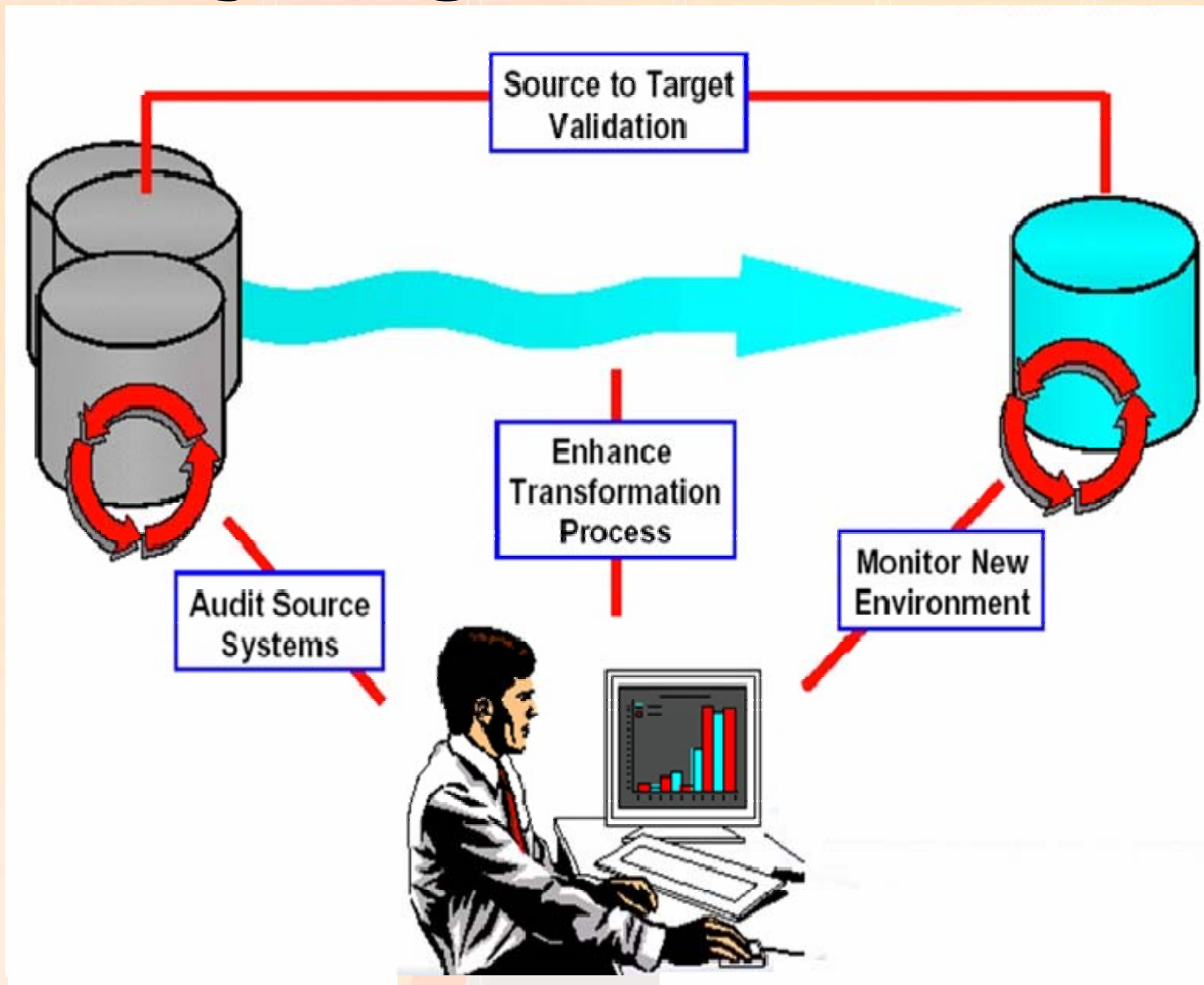
# Data Quality Assurance



# Data Profiling using AXIO

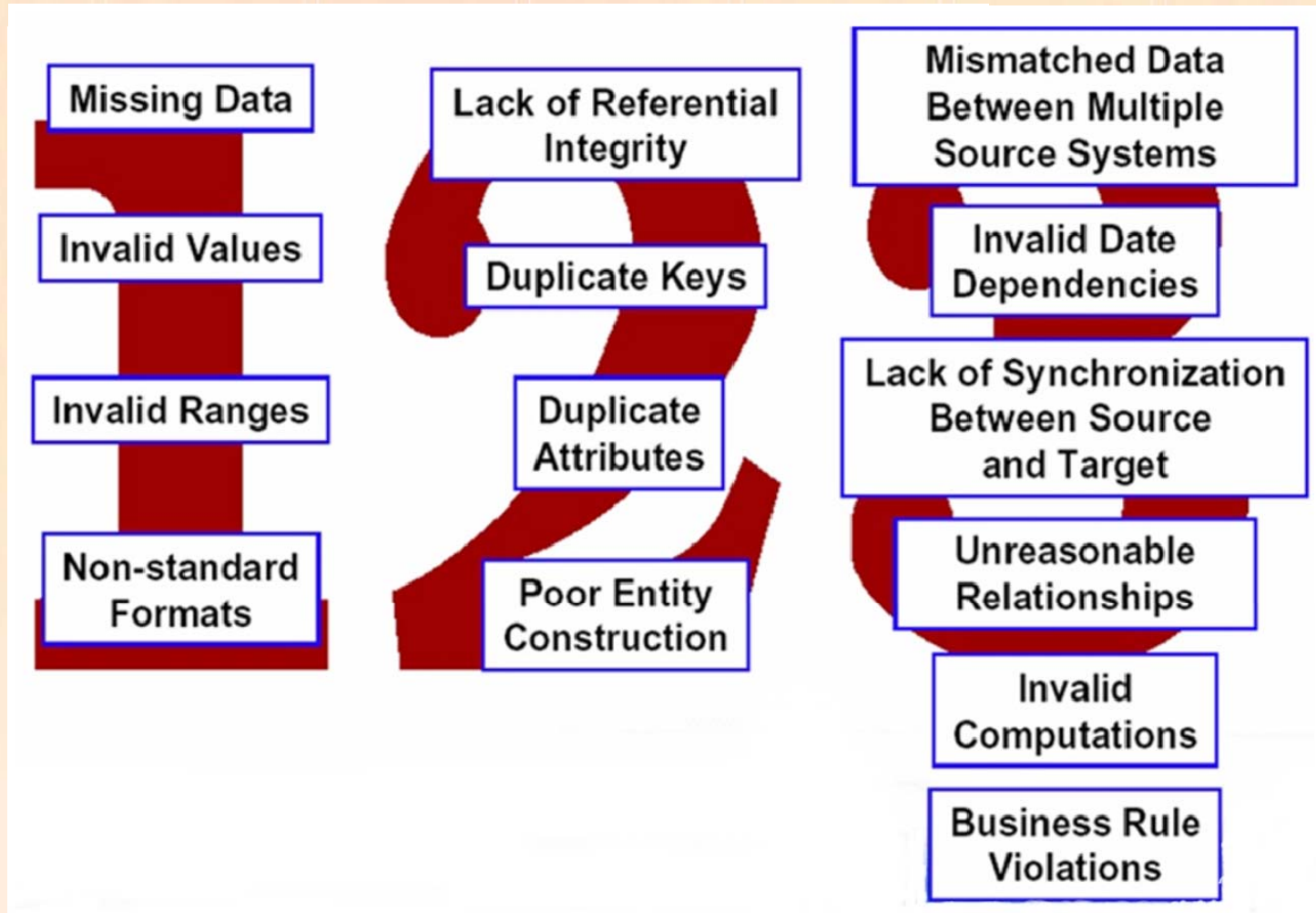


# QualityStage





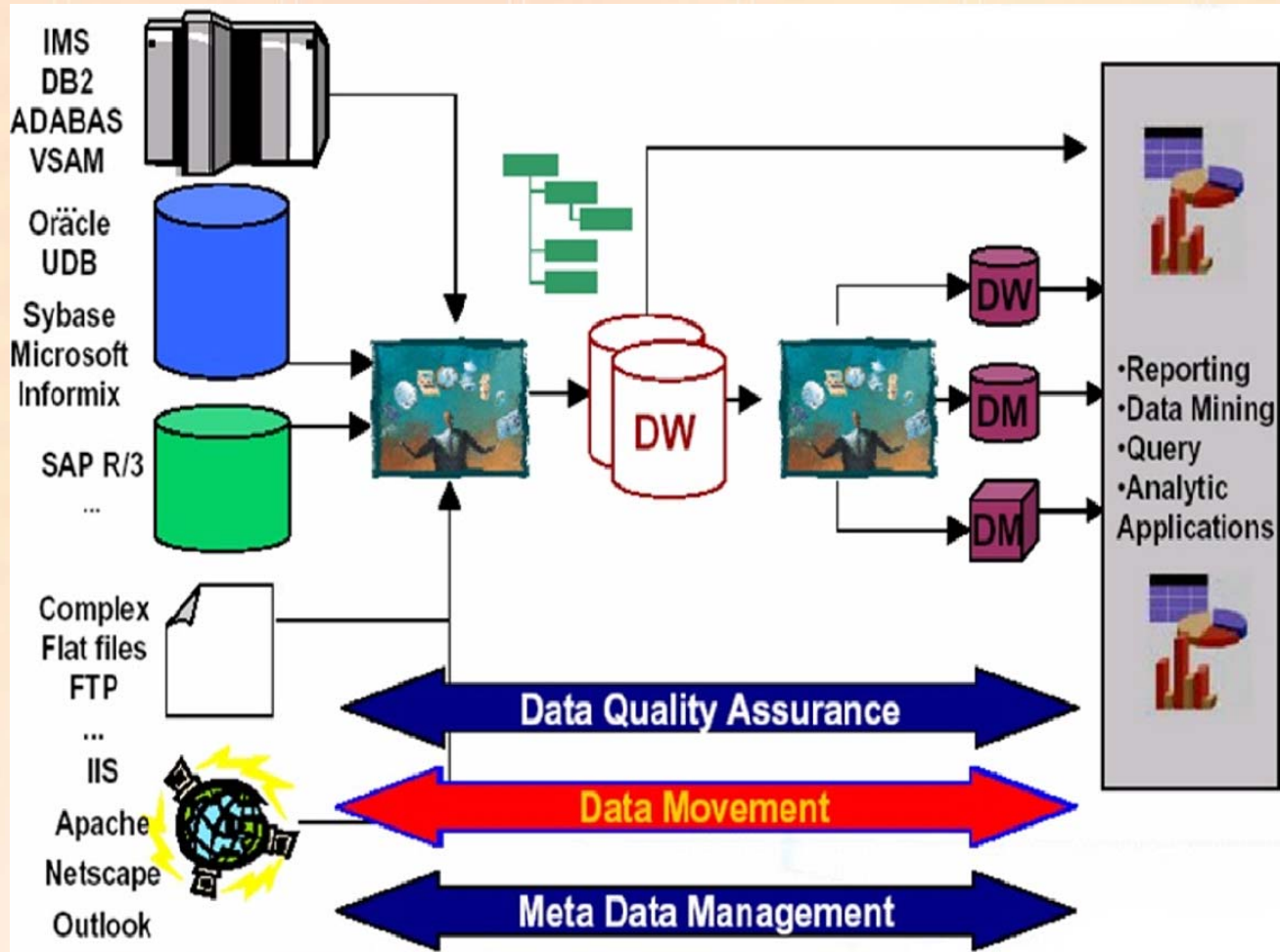
# Data Quality Methodology



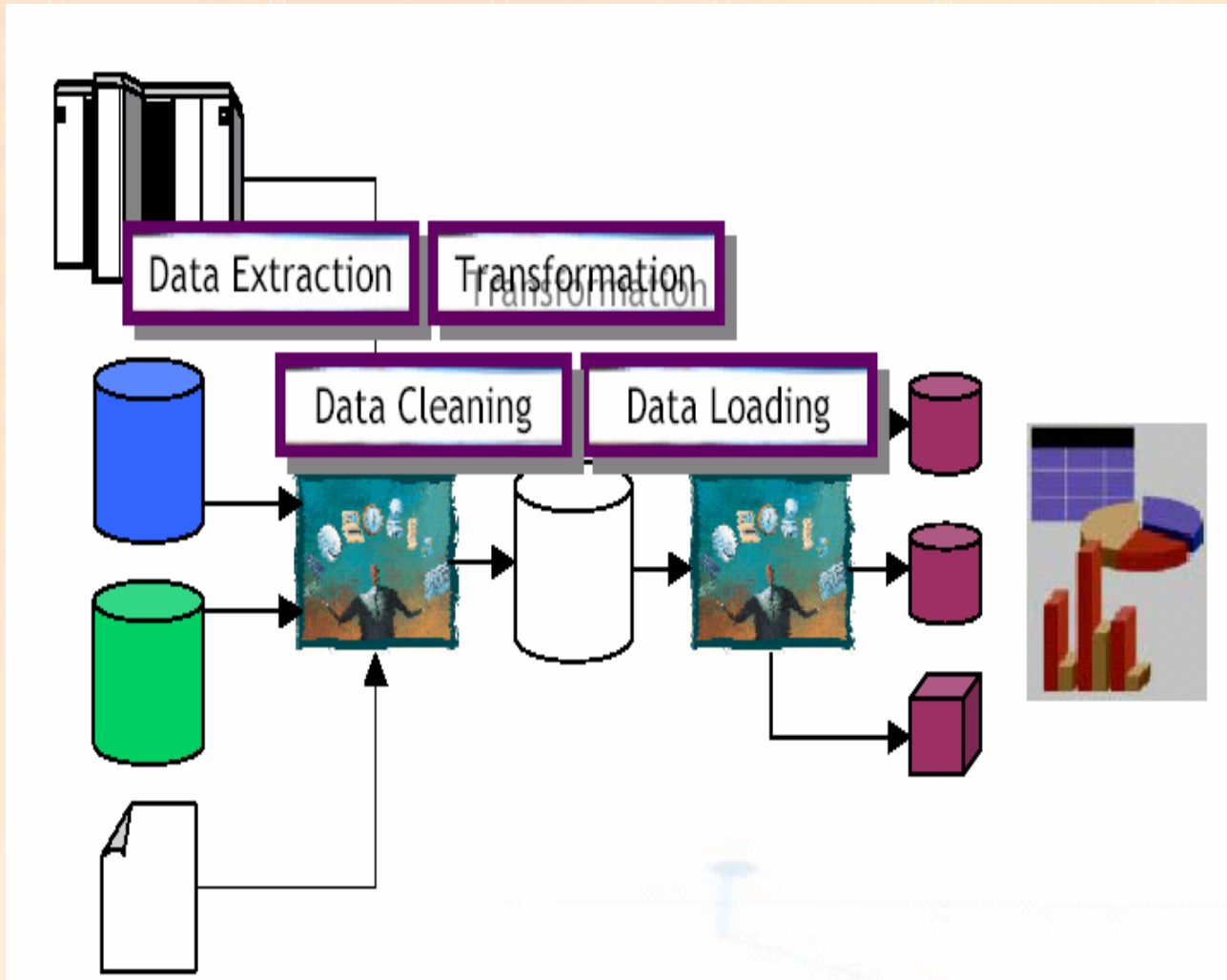
# Data Quality Benefits

- A measurable, repeatable process of ongoing improvement
- Identified, quantified data quality issues
- Successful data warehouse efforts
- Shortened development cycles
- Reasonable warehouse user expectations: real DSS and data mining results
- Demonstration of the return on the organization's investment in data quality
- Ability to respond to customers and regulators about data quality levels

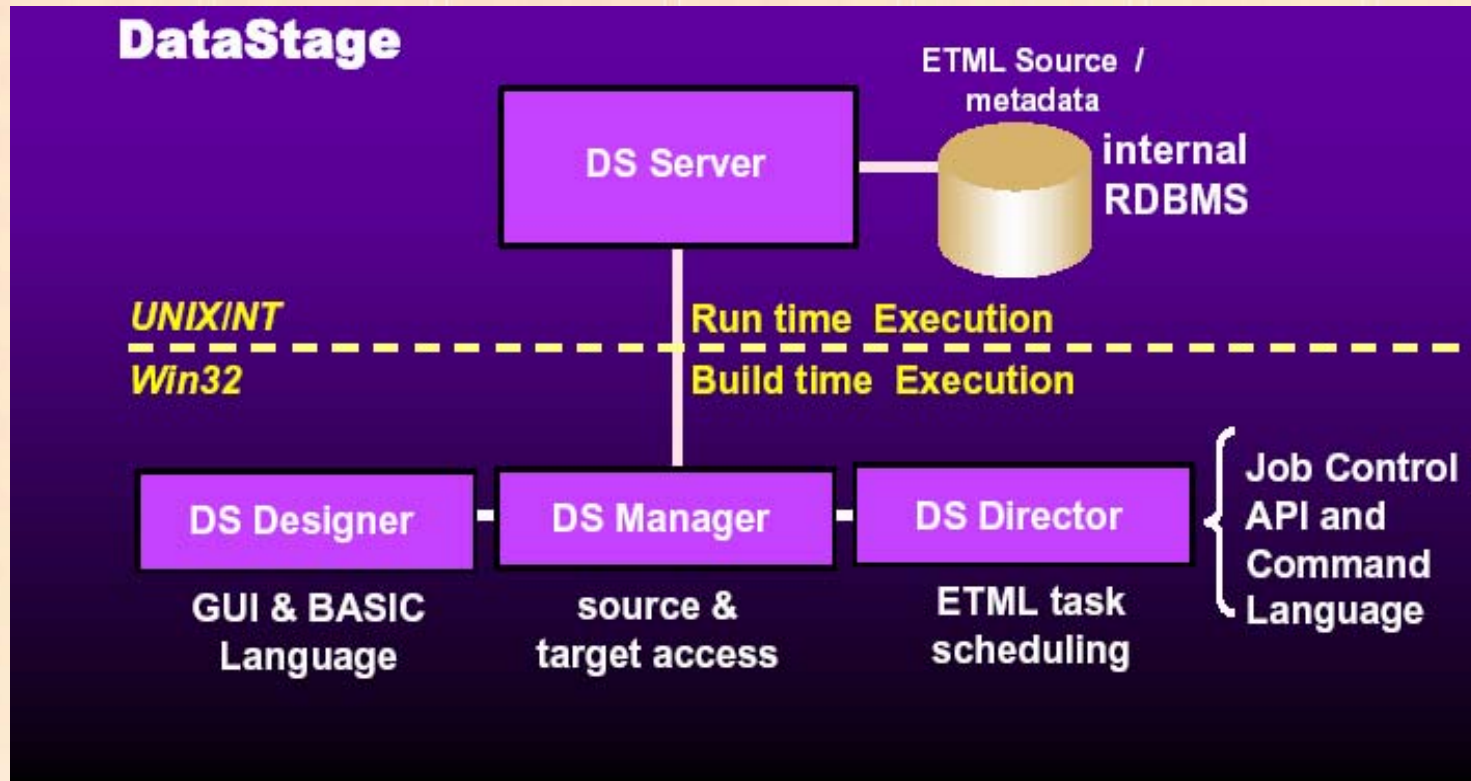
# The Data Warehouse Environment



# Data Movement

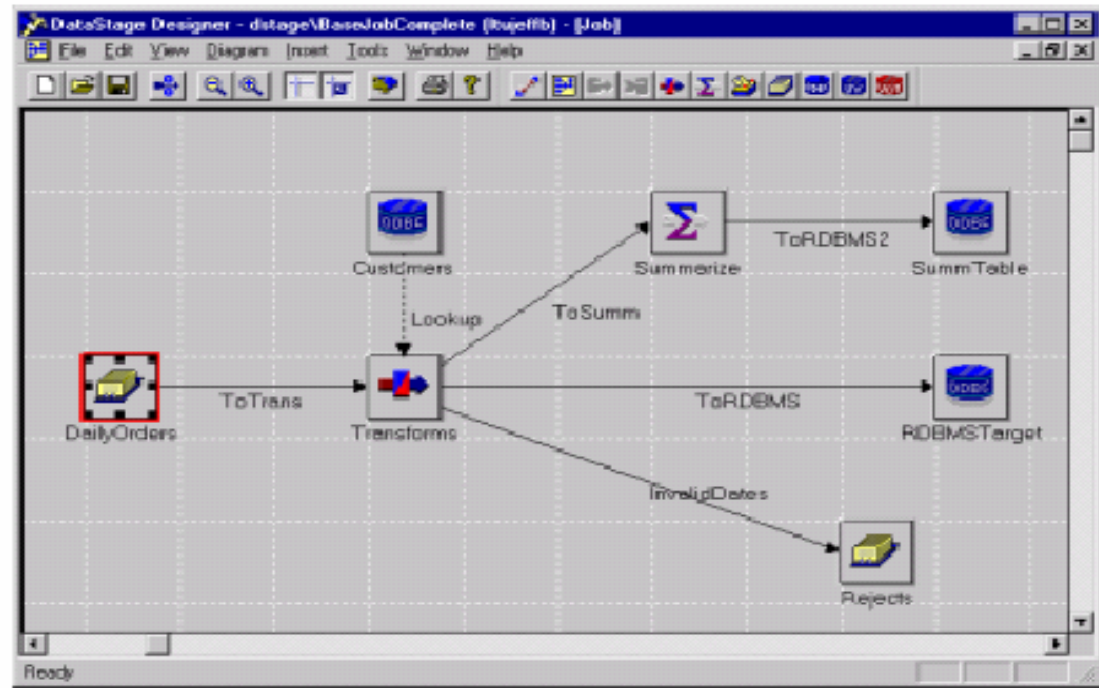


# DataStage Architecture



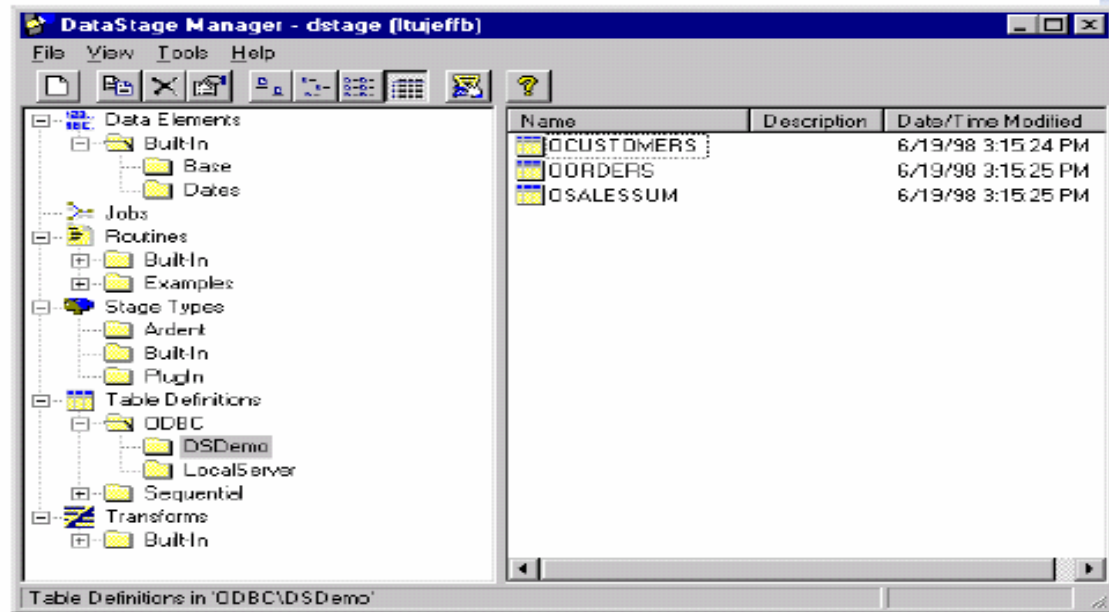
# DataStage Designer

- Design ETL process
- Each icon is a stage
- Each stage has editor
- Link stages to define data flow



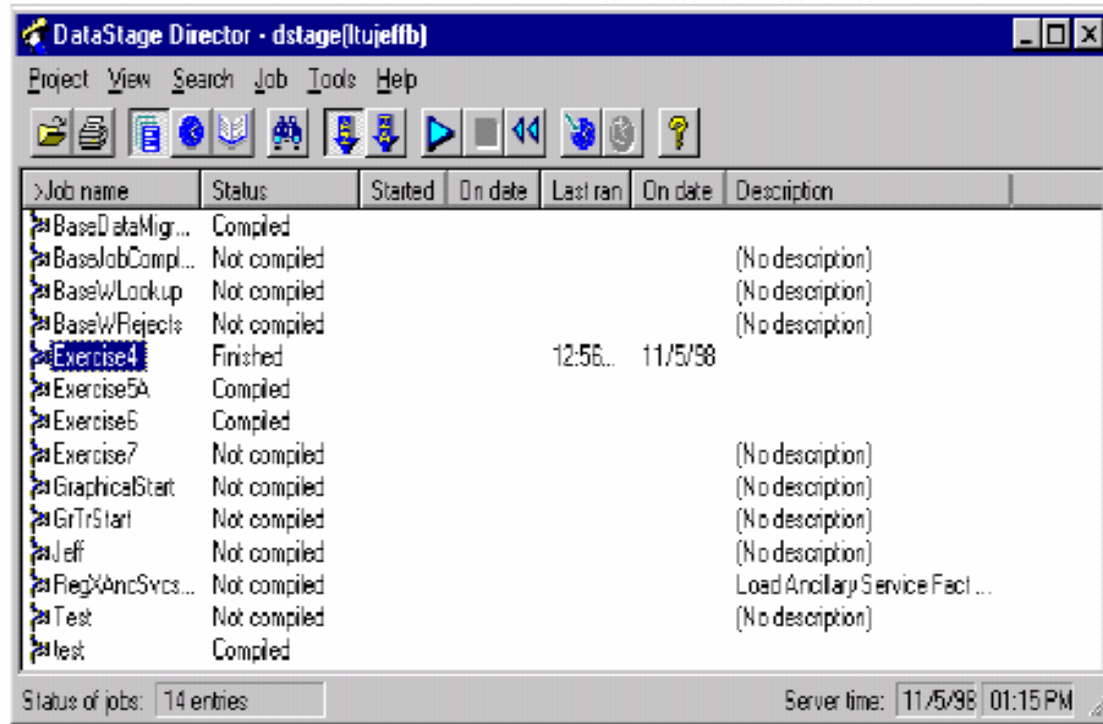
# DataStage Manager

- Manage definitions
  - Meta data
  - Transformations
  - Stages
- Import / export
  - ODBC or API
  - MetaBrokers
  - ActiveX transformations
  - XML DTD, Schema definitions



# DataStage Director

- Manage jobs
- Schedule jobs
- Review status of jobs
- Filter view for large environments
- Can use external scheduling engine



The screenshot shows the DataStage Director application window titled "DataStage Director - dstage(ltjjeffb)". The interface includes a menu bar (Project, View, Search, Job, Tools, Help) and a toolbar with various icons. Below the toolbar is a table displaying the status of various jobs. The table has columns for Job name, Status, Started, On date, Last ran, On date, and Description. The job "Exercise4" is highlighted in blue and shows a status of "Finished" with a "Last ran" time of "12:56.." and an "On date" of "11/5/98". Other jobs include "BaseDataMigr...", "BaseJobCompl...", "BaseW/lookup", "BaseW/Rejects", "Exercise5A", "Exercise6", "Exercise7", "GraphicalStart", "GrTrStart", "Jeff", "RegAncSvcs...", "Test", and "test".

Job name	Status	Started	On date	Last ran	On date	Description
BaseDataMigr...	Compiled					
BaseJobCompl...	Not compiled					(No description)
BaseW/lookup	Not compiled					(No description)
BaseW/Rejects	Not compiled					(No description)
Exercise4	Finished			12:56..	11/5/98	
Exercise5A	Compiled					
Exercise6	Compiled					
Exercise7	Not compiled					(No description)
GraphicalStart	Not compiled					(No description)
GrTrStart	Not compiled					(No description)
Jeff	Not compiled					(No description)
RegAncSvcs...	Not compiled					Load Ancillary Service Fact...
Test	Not compiled					(No description)
test	Compiled					

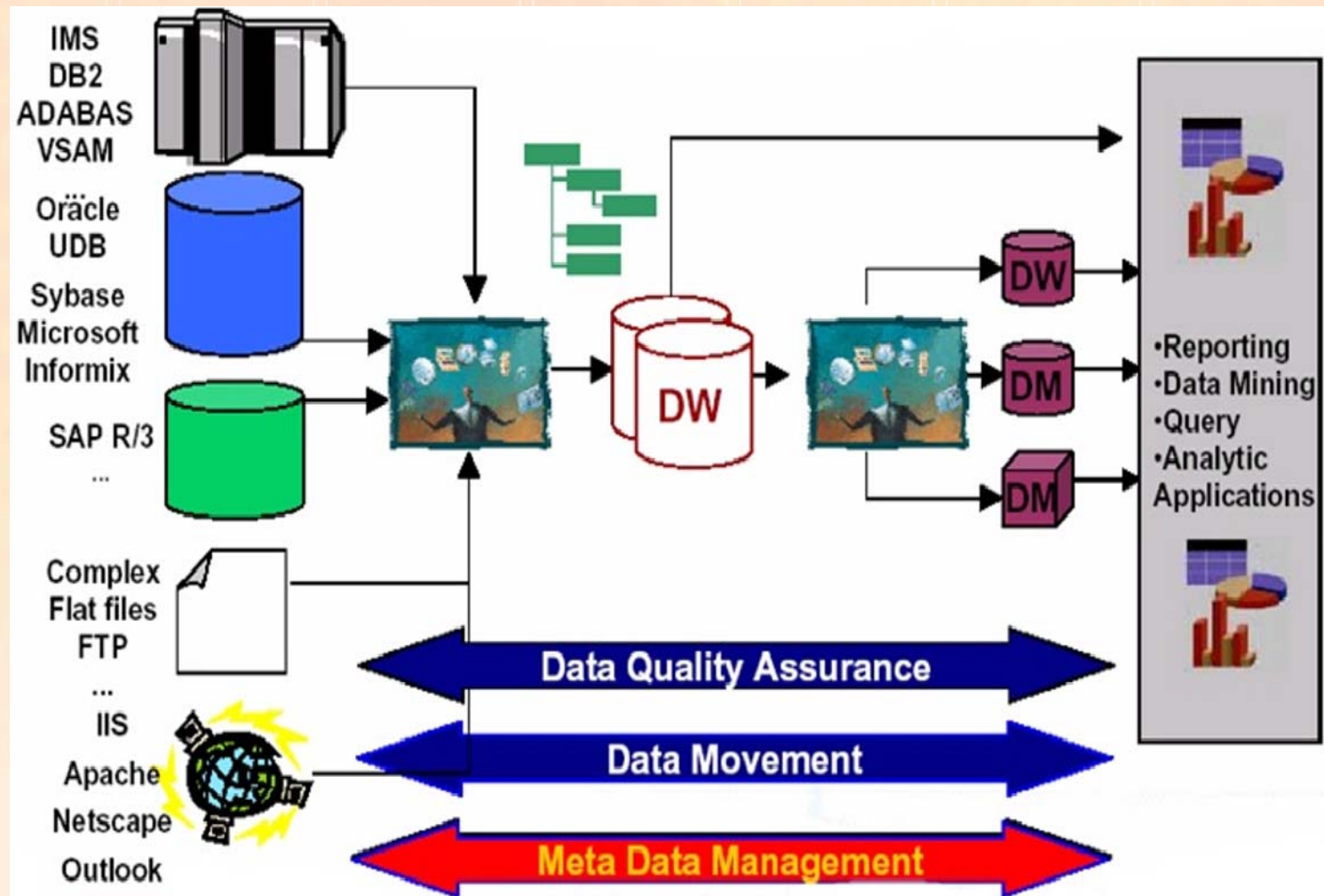
Status of jobs: 14 entries Server time: 11/5/98 01:15 PM

# DataStage Process Stages

Sources	Process	Targets
Informix	Aggregation	Informix
Oracle	Sort	Oracle
Sybase	Merge	Sybase
DB2	Cleaning	DB2 UDB
MS SQL Server	Route	MS SQL Server
Teradata	400+ Transforms	Teradata
FTP	Custom Stages	Red Brick
Flat Files		Informix XPS
XML		MQSeries
IMS & VSAM		XML
Named pipes		FTP
EMC InfoMover		
ODBC		
SAP		
Siebel		
Peoplesoft		
MQ Series		



# The Data Warehouse Environment



# Metadata Management

- Provide *meaning* to all aspects of your data warehouse application
  - What does the data mean?
  - What is the data's structure and format?
  - Where did the data come from?
  - How was the data calculated?
  - When was the data loaded?
  - Who owns the data?
  - Where is the data used?



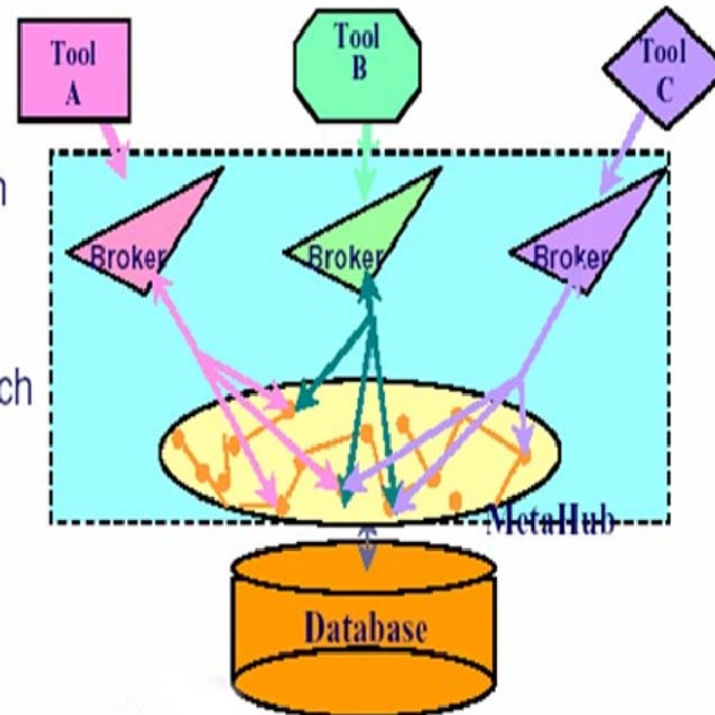
# Managing Metadata using Metastage

Enables exchange of meta data among applications with different underlying meta data stores

Patented technology with three key elements:

- ◆ **Transformation** technology decomposes data concepts of each tool into simple semantic units and recomposes them as needed
- ◆ **Brokers** map the data model of each tool to the hub
- ◆ **MetaHub** is a schema that represents the essential semantic units for a shared environment

*Major advances:  
decomposition/re-composition process;  
Logical, not physical integration...*

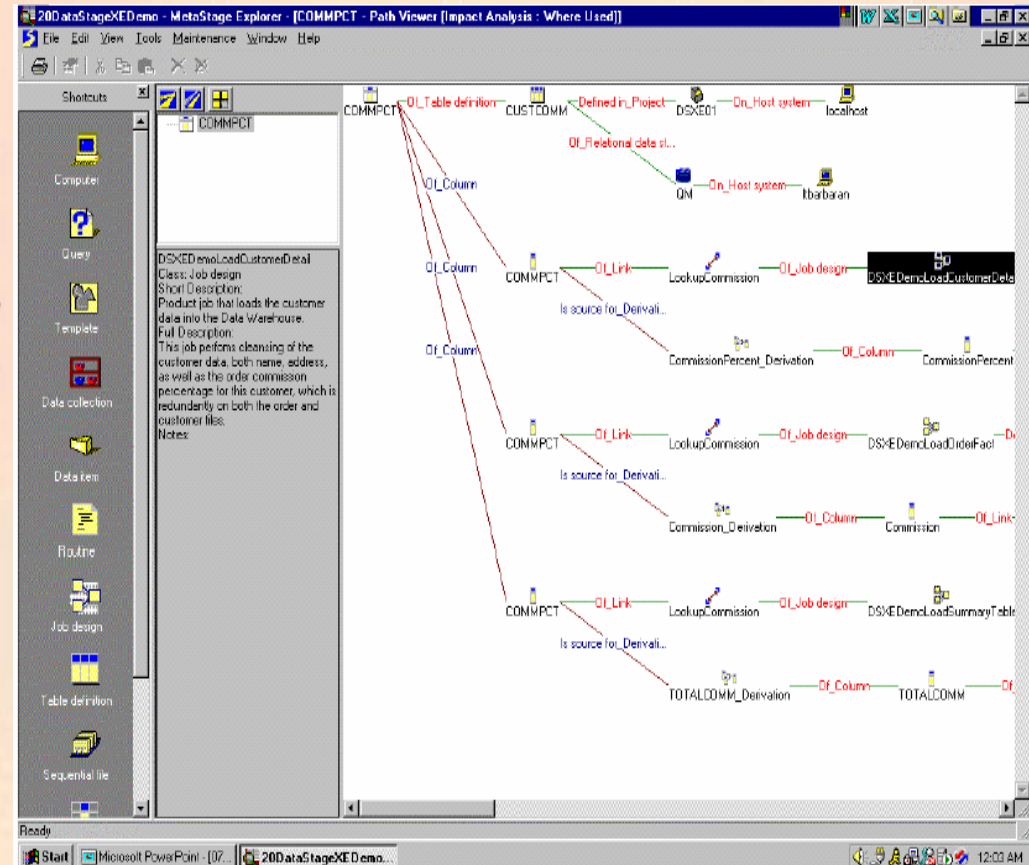


# Impact Analysis

Example questions:

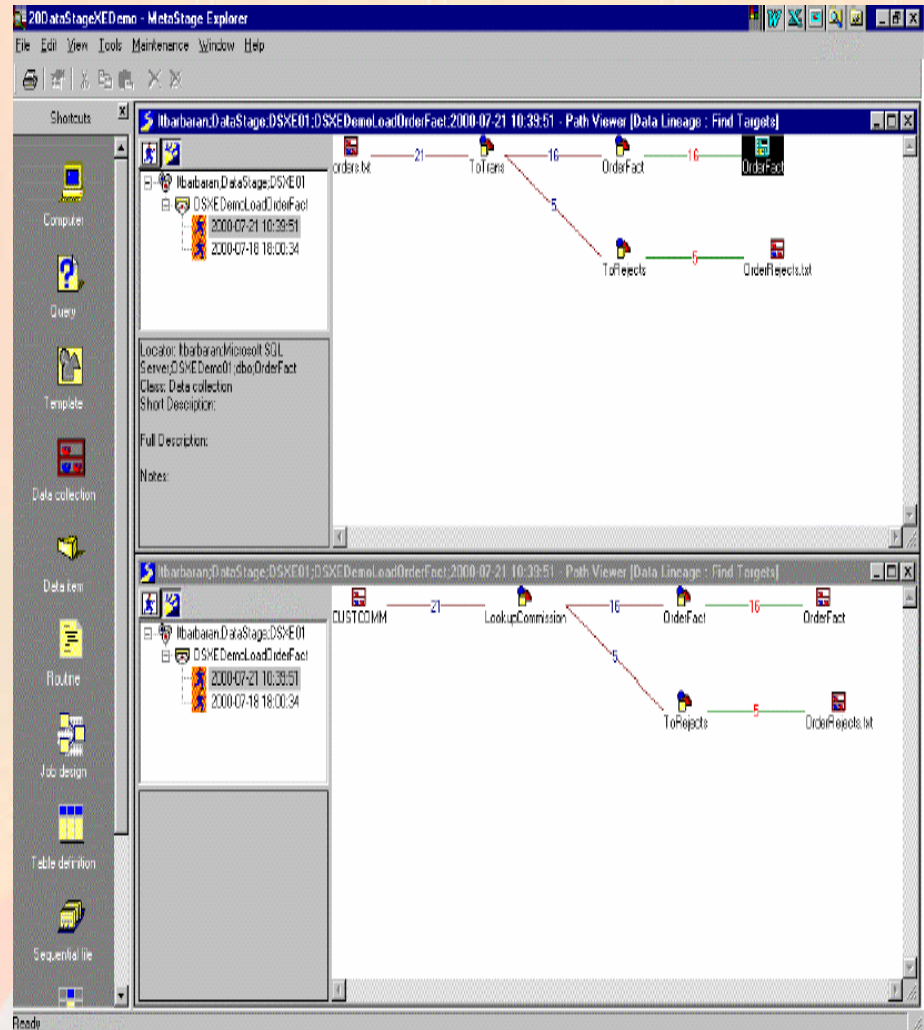
- What else gets affected if I make a change?
- What else am I dependent upon?
- What tables and columns are in the Data Warehouse data model?

MetaStage answers these questions and more by following containment and dependency relationships within the metadata in a directory.



# Data Lineage

Data lineage analysis on the data warehouse table allows to identify the sources of data written to a table and what target data is read from a given table



# Metadata Management

