

# NCR Vorgehensmodell, Data Warehouse Architektur

Technische Universität Wien, Wissenschaftlerinnenkolleg Internettechnologien  
Institut für Softwaretechnik und Interaktive Systeme, Hörsaal EI 3A

Roland Pfeffer

Data Warehouse Consultant

Teradata Austria

[roland.pfeffer@teradata-ncr.com](mailto:roland.pfeffer@teradata-ncr.com)

You've never seen your business like this before.

# Agenda

- Unternehmensprofil
- Database: OLTP vs. DWH
  - Technische Unterschiede
  - Unterschiede Projektablauf, Projektorganisation
  - DWH: Datengetrieben vs. Prozessgetrieben
- Geschäftsanforderungen, Methodologie
  - NCR Vorgehensmodell
  - Phasen des Vorgehensmodells
  - Vorgänge des Vorgehensmodells
- DWH Architektur
  - Logische Komponenten
  - Source Systeme, ELT vs. ETL
  - Staging Area, Integrationsbereich
  - Aggregate, Dimensionale Strukturen, Anwendungsschicht
  - Technische Aspekte / Anforderungen
  - Warum Parallelität und Skalierbarkeit ?
  - Data Warehouse Workload
- Aspekte der Datenmodellierung

# Company Profile

You've never seen your business like this before.

# Teradata all over the world

NCR Corporation, acting in more than 130 countries

**Teradata**, a division of NCR

- fully dedicated **ONLY** on Data Warehousing
- with over 20 years experience in DW
- Over 2000 DW business and technical consultants with average 15 years of experience

# Teradata Experience

## *Why Guess... We Know.*

*Here are some key facts about Teradata customers:*

### 1. The biggest number of successful DW implementation

**Over 1200 Data Warehouses Worldwide**

### 2. Cooperation with the leading companies

- 6 of the world's top 10 retailers
- 7 of the world's top 9 telecommunications companies
- 6 of the world's top 7 airlines
- Teradata customers manage over 50% of the assets from the top 15 global banks.

*„The 3+TB DW market belongs to NCR Teradata.”*

**Mark Shainman, Meta Group**

### 3. Unique experience in Large Data Warehouses

- 300+ customers with data warehouses > 1 Terabyte
- 100+ customers with data warehouses > 3.5 Terabytes

### 4. Extensive Telco industry knowledge

**124 DW in Telco sector implementations**

### 5. Customers satisfaction

*„The NCR Data Warehouse project was the most successful IT project we ever had.”*

**Pannon GSM, Hungary**

# Database:

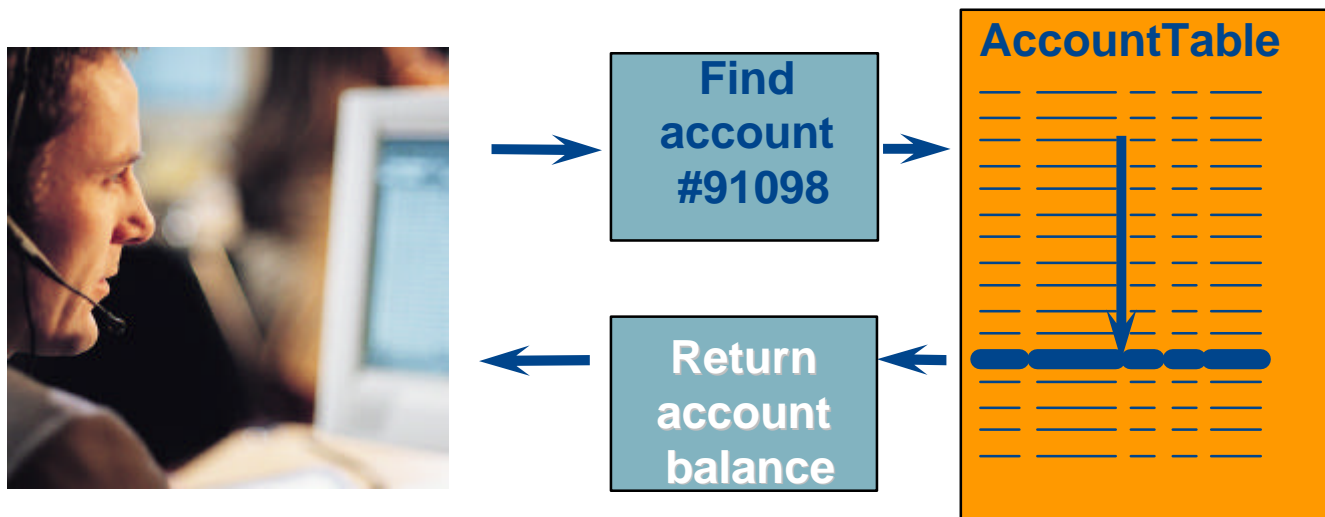
Online Transaction Processing (OLTP) vs.  
Data Warehouse (DWH)

You've never seen your business like this before.

# Database: OLTP vs. DWH

## *Typical OLTP Access*

- What's the balance in account 91098 ?
- Update Status\_Married of Customer\_Id 896677
- Record Claim of Customer\_Id 983246
- Record Order for Customer\_Id 654667
- What was the last Order of Customer\_Id 476878 ?



→ Single Table Accessed  
→ Short, Quick Queries

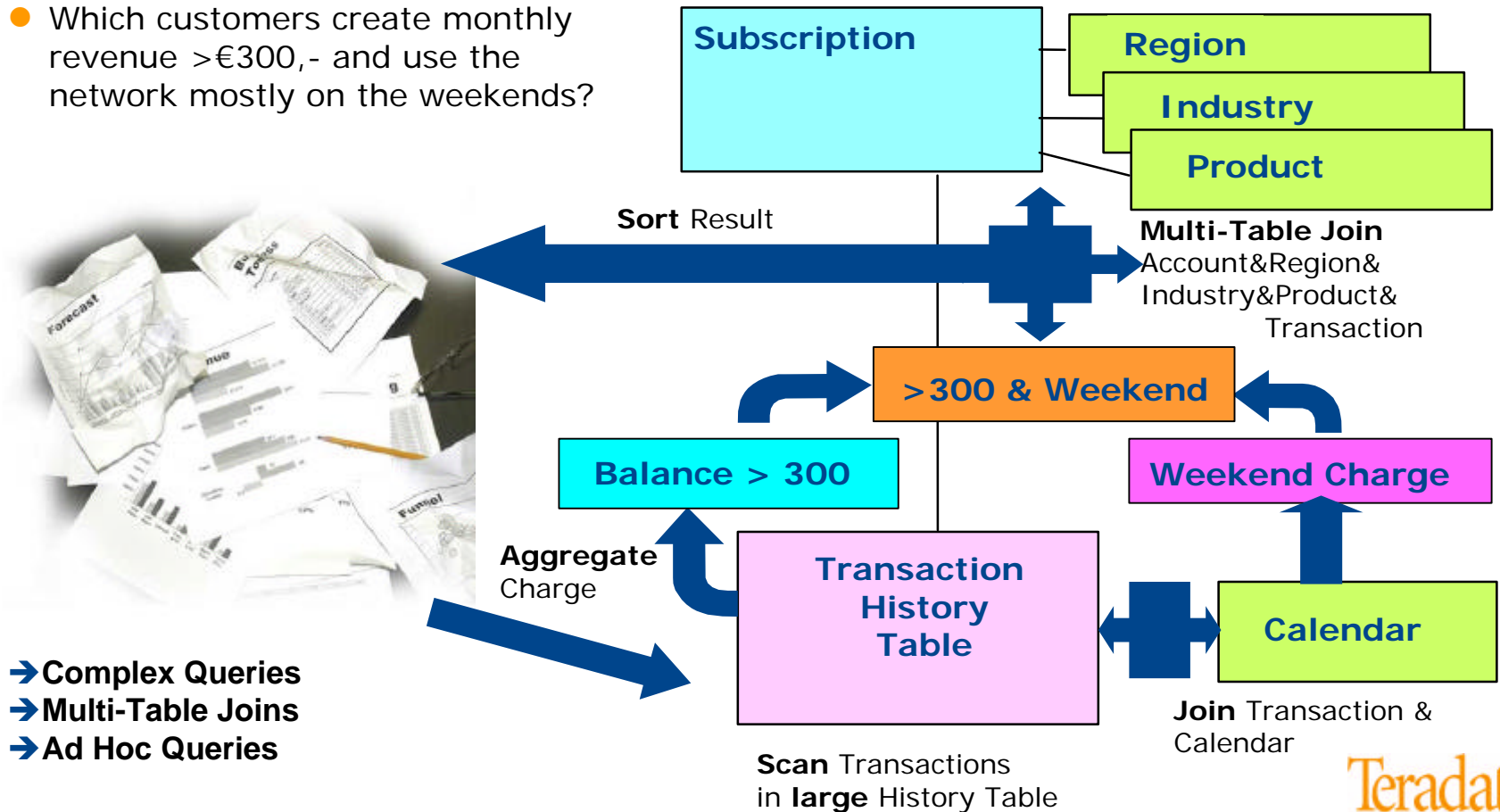
→ Prime Key Retrieval  
→ Known Queries

# Database: OLTP vs. DWH

## Typical DWH Access

- Identify potential churn customers based on attributes of customers who already churned (e.g. call behaviour)
- Which customers create monthly revenue >€300,- and use the network mostly on the weekends?

- How much revenue is generated by our top 10% customers ?
- How much impact has a potential decrease of our tariffs for the product X ?





# Database: OLTP vs. DWH

*Why to use special Technology for Data Warehousing ?*

Billing, Accounting,...

***Small # Records***

***Retrieve via Key***

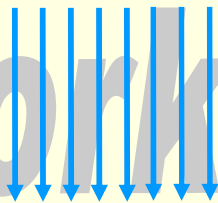
***Predictable***

***Coded programs***

***Record Updates***

***Sub-second Response***

***Short, Update Transactions***



**Departmental,  
Operational Data**

Data Warehouse

***Large # Records***

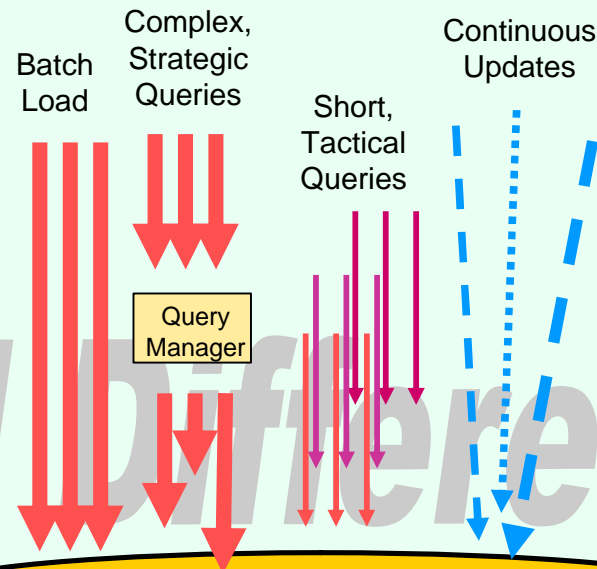
***Complex Joins***

***Unpredictable***

***Adhoc Tools***

***Batch Updates***

***Various Response times***



**All Decision-Making Data Integrated**

Workload Differences

# Database: OLTP vs. DWH

## *Relevance of Processes and Data*

- IT Projects in an operational environment are process driven
  - Business processes define requirements or business processes are aligned to new definitions: e.g.: „*We need a new application to bill new services which are used by our customers* “ (Billing)
  - Design a functional specification including a detail description of the end user interface, processes and functions
  - Implementation according to the functional specification: data model, functions, reports
  - Test and acceptance
  - End of the project

# Database: OLTP vs. DWH

## *Relevance of Processes and Data*

### ● Data Warehouse projects are data driven

- Business questions, which cannot be answered using existing tools and where the answer possibly contributes to revenue growth, margin optimization or cost reduction, define the requirements (data subject areas) for the Data Warehouse like e.g.: *We need to identify customers, who are valuable to our organisation and have a high propensity to churn to address this customers proactively and propose a appropriate offer to reduce our customer churn rate.* (Churn – Prediction)
- How the data areas are analyzed can (e.g.: in the case of generating Key Performance Indicators so called KPI´s or reports) but does not need to be defined in the design phase (e.g.: impact analysis, data mining,...)
- To generate a Logical Data Model it is necessary to identify the data areas, data entities and attributes which are required for the analysis, and the relationships between these items. In addition entity and attribute definitions, transformation rules, cleansing processes and data integration rules define the scope of the Data Warehouse project.

# Database: OLTP vs. DWH

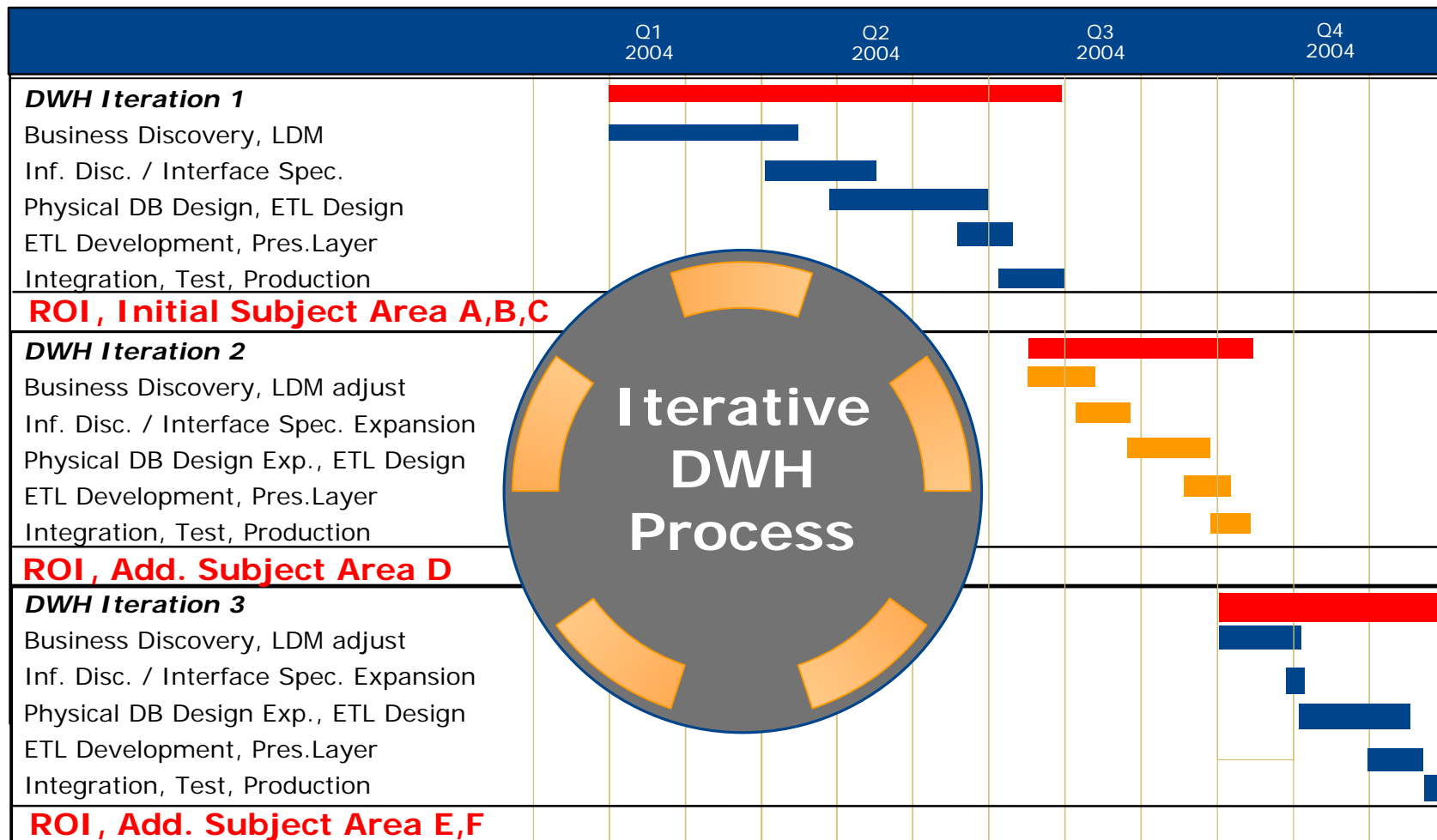
## *Relevance of Processes and Data*

### ● Data Warehouse projects are data driven

- Implementation: Loading the database data schema from the required sources according to the defined rules for transformation, cleansing and integration. Also the logical (views) or if required physical presentation layers (derived tables) need to be generated based on the base physical data schema (which is the representation of the Logical Data Model) to support Reports and Online Analytical Processing (OLAP) Analysis.
- Goal: The focus of the Data Warehouse is the structured and integrated representation of the enterprise business data for analysis opposed to the implementation of application specific functions in an OLTP environment.
- New perceptions of the analysts and business experts based on analytic results from the DWH rise new questions and requirements for the DWH (requirements to add new data areas to the DWH) → **next DWH Iteration**

# Project cycle DWH

(simplified project plan)



# Database: OLTP vs. DWH

## *Project cycle: Risk & Success Factors*

Because of the dynamic nature of the DWH environment a effective process chain is needed to run a successful DWH project: Source Systems – Extract/Load/Transform – Data Integration – Analysis and reporting of the data by analysts and business experts.

- „We need to have a detailed functional specification before we can do anything“ or „The first business group, which requires additional data in the DWH must pay for the effort“ *does not work well in a DWH project environment*. → *End Users get offended*: „Under these circumstances it is easier for us to implement our own data mart which exactly siutes our requirements “ → results in a Data Mart Dilemma, many decentral Data Marts arise, which require much maintenance in summary and generate different results, when compared.

# Database: OLTP vs. DWH

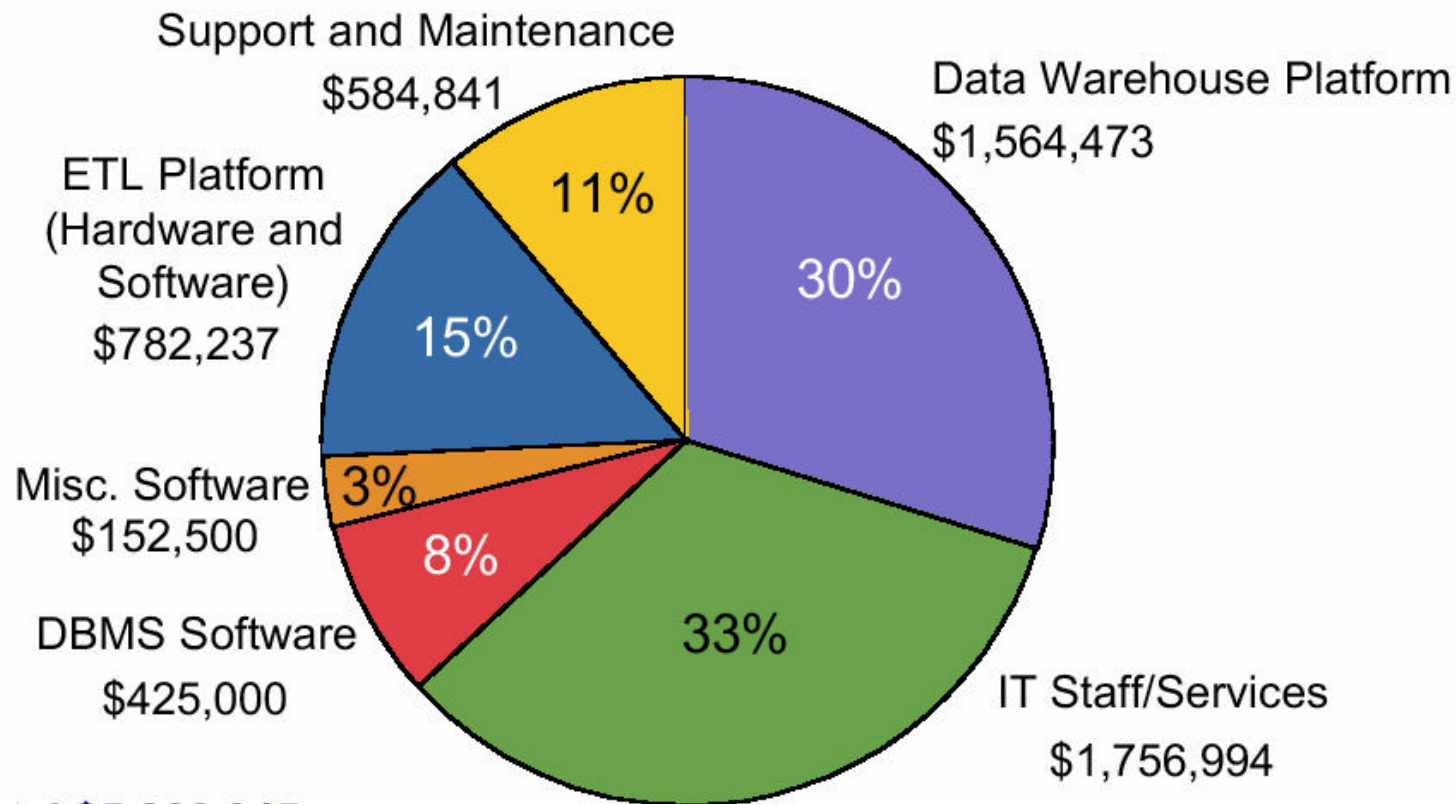
## *Project cycle: Risk & Success Factors*

Because of the dynamic nature of the DWH environment a effective process chain is needed to run a successful DWH project: Source Systems – Extract/Load/Transform – Data Integration – Analysis and reporting of the data by analysts and business experts.

- „We need to sell the DWH services within our organisation actively to attract as many users as possible, to let their valuable business questions answer by the DWH“. If the DWH is the most attractive tool for the end user to satisfy their information request, then *Data will be stored only once in an integrative centralized repository*. → averts a Data Mart Dilemma, saves costs and improves quality.
- Integrative collaboration between IT – business unit A – business unit B is required
  - business units: Implementation competence to IT
  - IT: regular active and innovative new development of the DWH, flexible implementation to new requirements

# Data Warehouse Sample Budget

**1TB of Raw Data With 3TB Disk Storage**

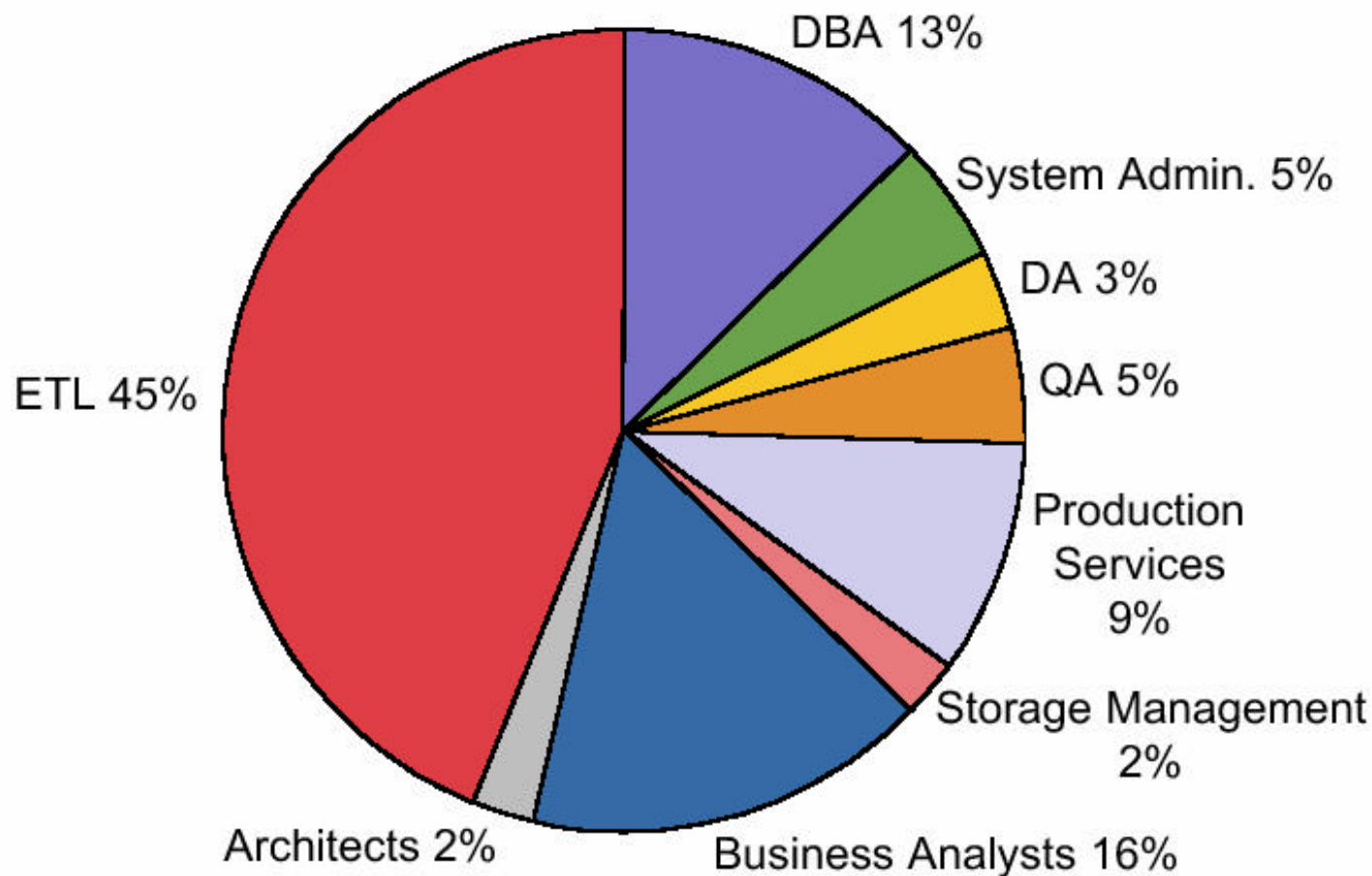


**Total \$5,266,045**



# Data Warehouse Staffing

---



# Business Requirements, Methodology

You've never seen your business like this before.

# Methodology

- **What is a Methodology?**

- Methodology is based on communication principles - accurate, clear, precise, and complete communication between the many individuals who participate in the building of a system. It is a sequence of procedures dictated by the needed inputs required to produce the next set of deliverables.

- **Why does Teradata have a methodology?**

- **Market Needs:**

Teradata Clients expect a strong, proven methodology that Teradata actively uses with each project. This is a reasonable assumption and most knowledgeable Clients ask about this.

- **Communication:**

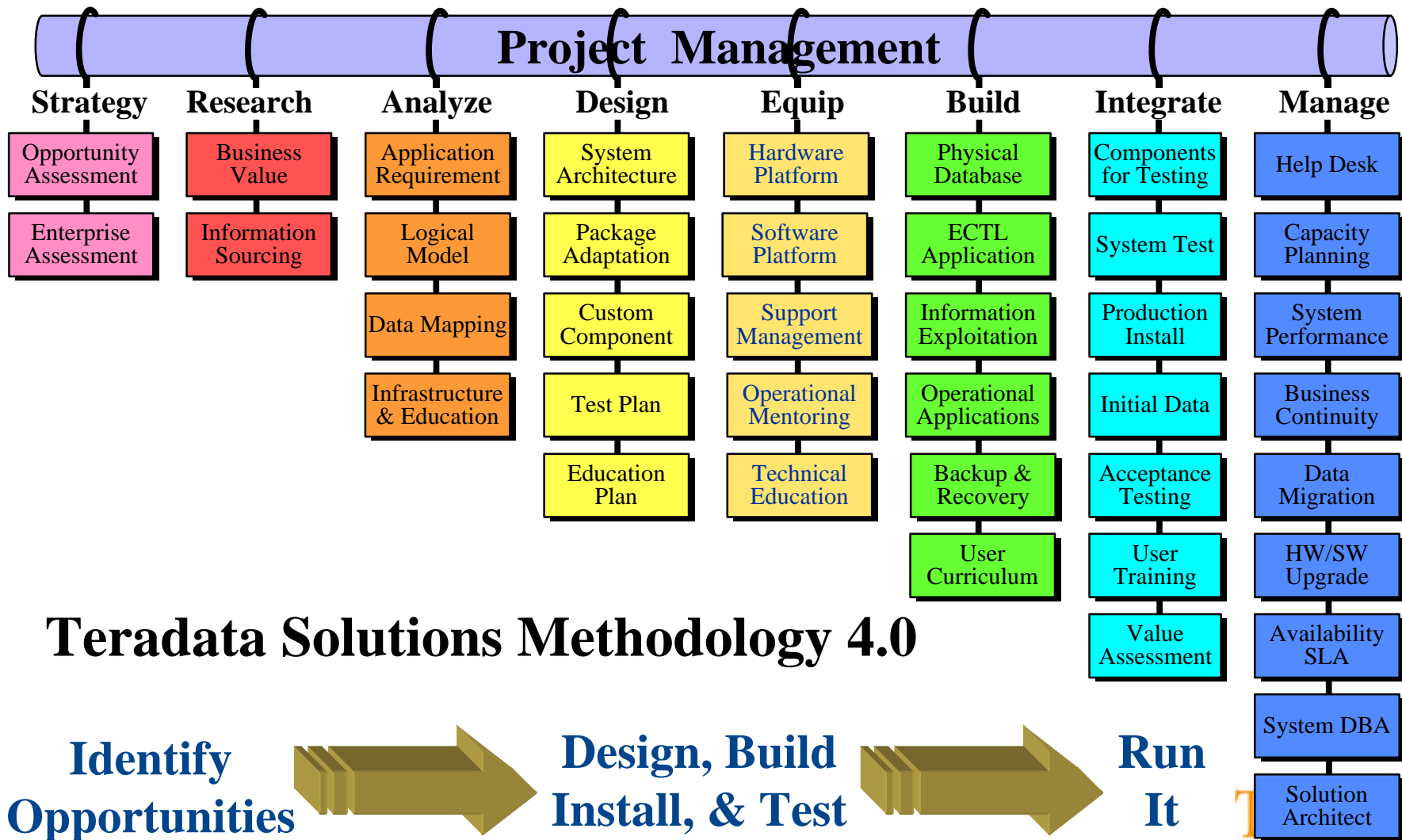
The Teradata methodology facilitates good communication within the Teradata project team.

- **Project management:**

A strong methodology fosters effective project management. TSM 4 is based on an engineering approach - not an art that is not repeatable or predictable.

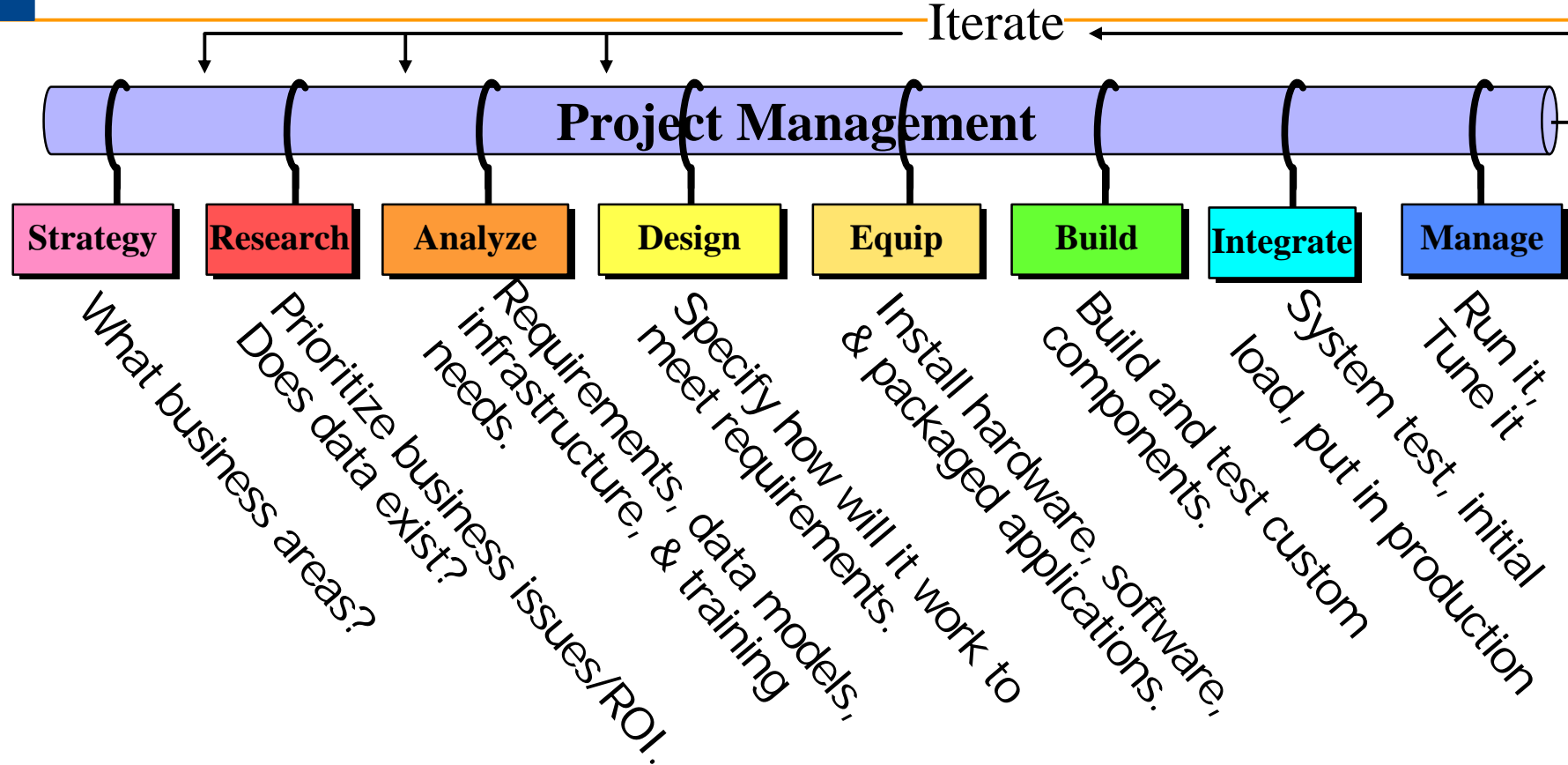
# Business requirements / Methodology

## Teradata Methodology



# Business requirements / Methodology

*Teradata Methodology: High-Level View Phases*



*DBMS Neutral Services*

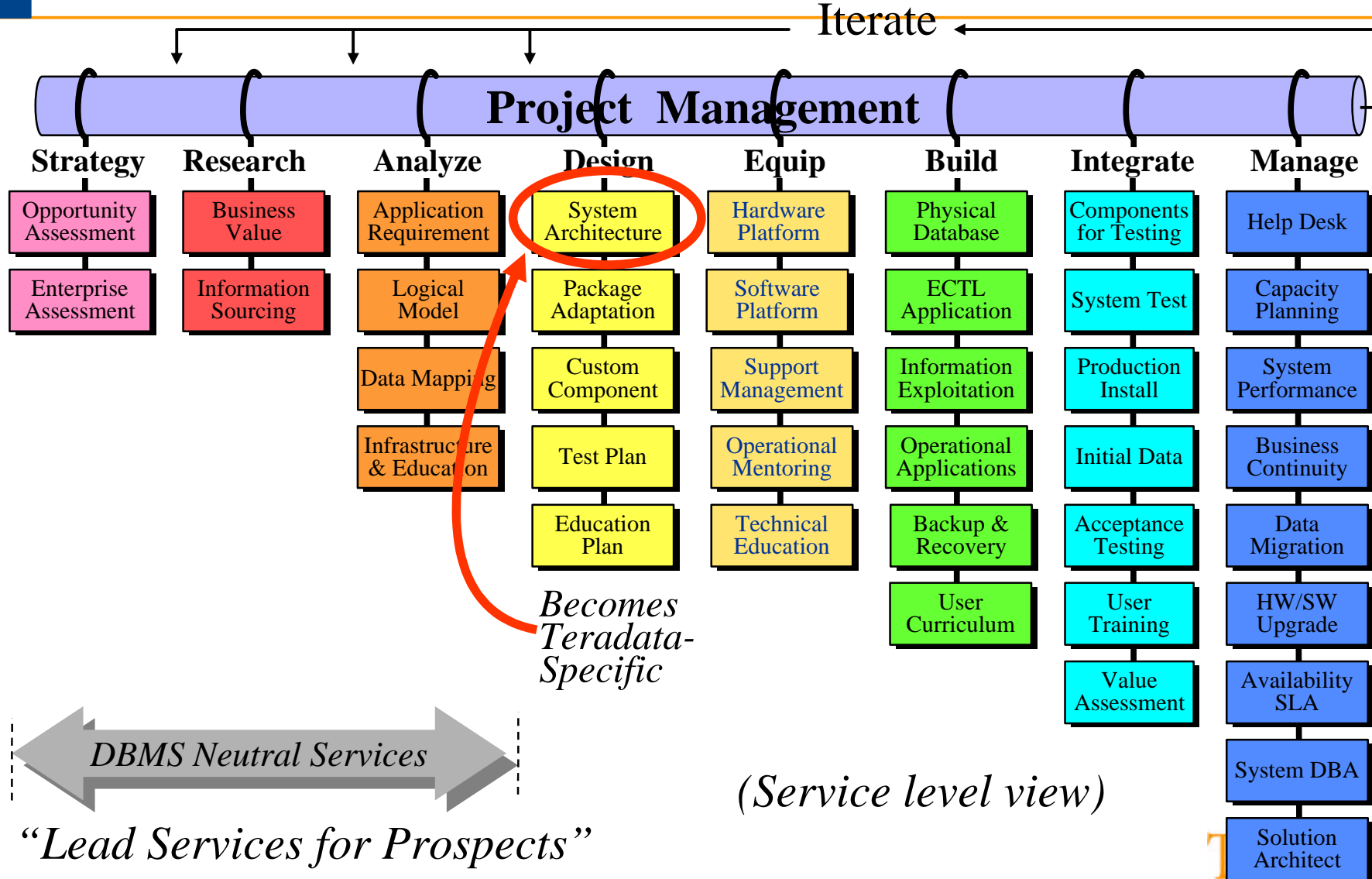
*“Lead Services for Prospects”*

***TSM 4.2: An Iterative Methodology***

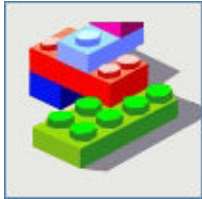
**teradata**  
a division of **NCR**

# Business requirements / Methodology

## Teradata Methodology: Detail of TSM Services



# TSM Features



- Modular. Services consist of tasks that have defined inputs and work products. (Tools described later allow for analysis of the impact of removal of a particular service or task.)
- Structured. Allows partnering with other service providers and Clients. (Tools described later will allow customizing the list of services in the Statement of Work and adding lists of work products that Clients or other service providers must deliver to the project team to complete its work.)
- Standardized Work Breakdown Structures (WBS) across projects. This allows analysis of improvement opportunities based on Formal Technical Reviews within each service and definition of a set of supporting files for every work product to provide alternate templates, examples, and guides by solution or industry.
- Solution Modifiers. Makes it easy to customize the methodology for different solutions by specifying which tasks, services, and phases are appropriate to the work to be done, the work products that must be provided by the Client or third parties, and the work products that will be deliverables.

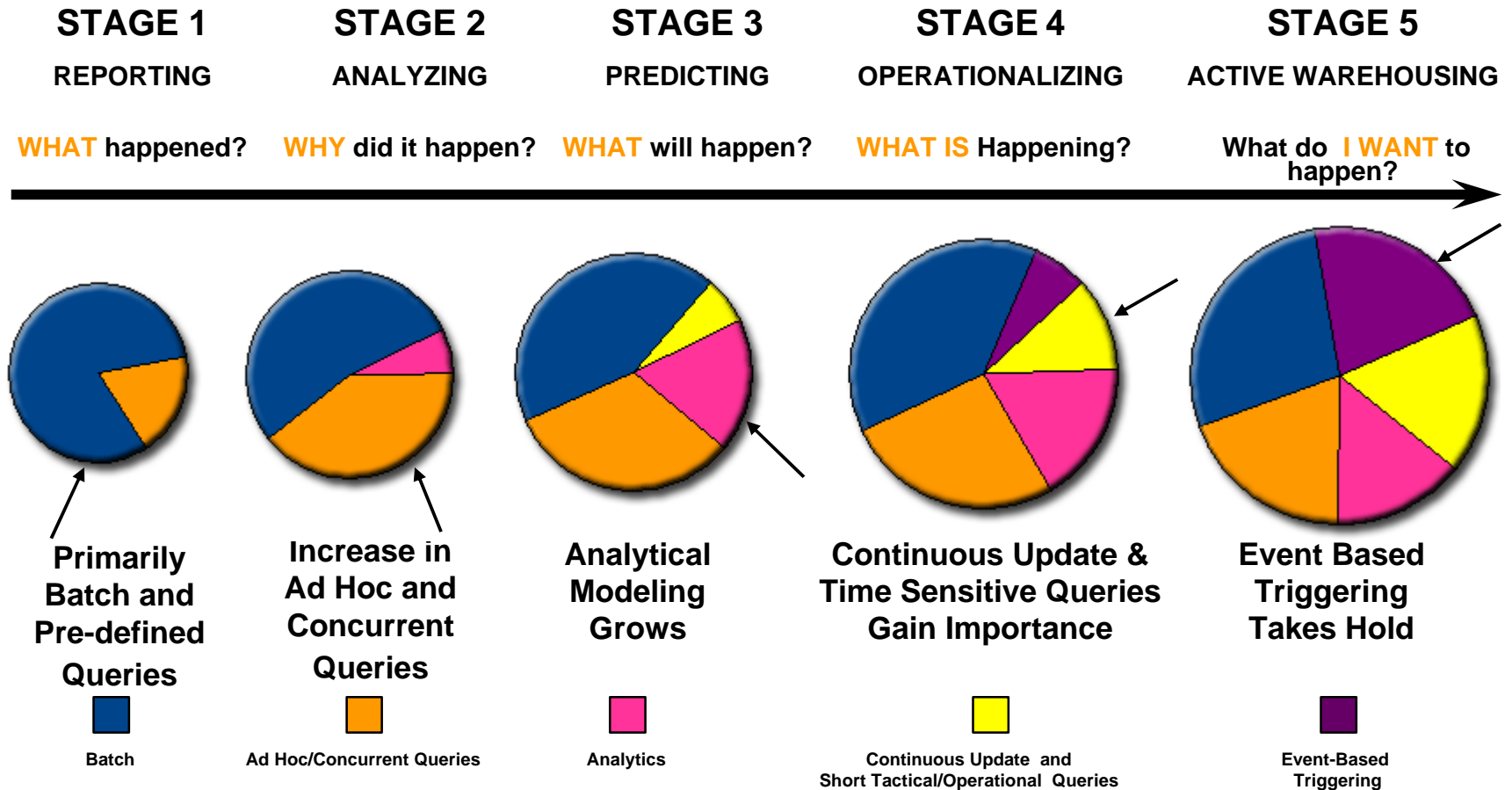
# DWH Architecture

You've never seen your business like this before.



# Data Warehouse Architecture

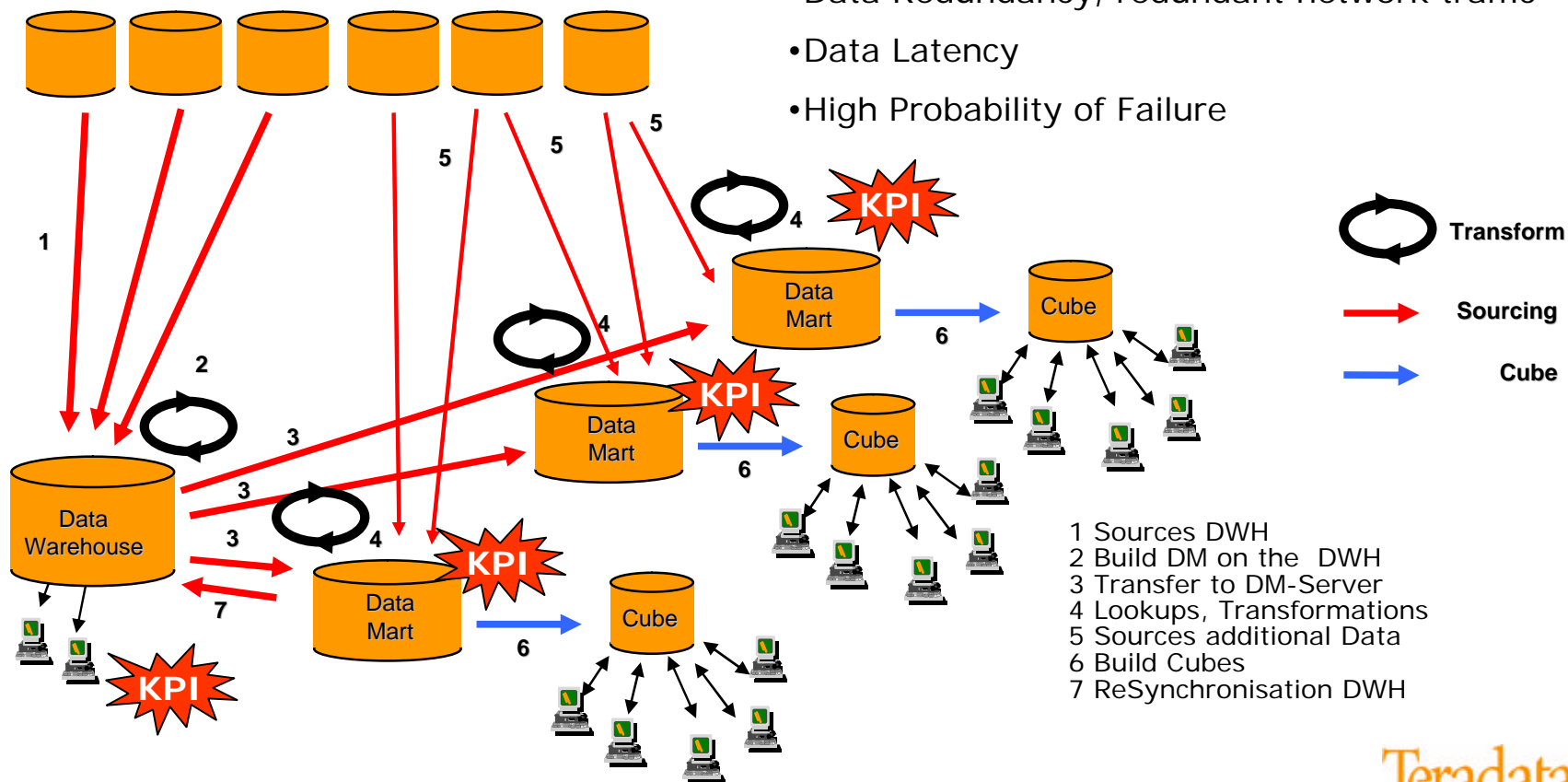
## Information Evolution in a Data Warehouse Environment



# Decentralized, distributed DWH Architecture

Information Processing (ODS not included)

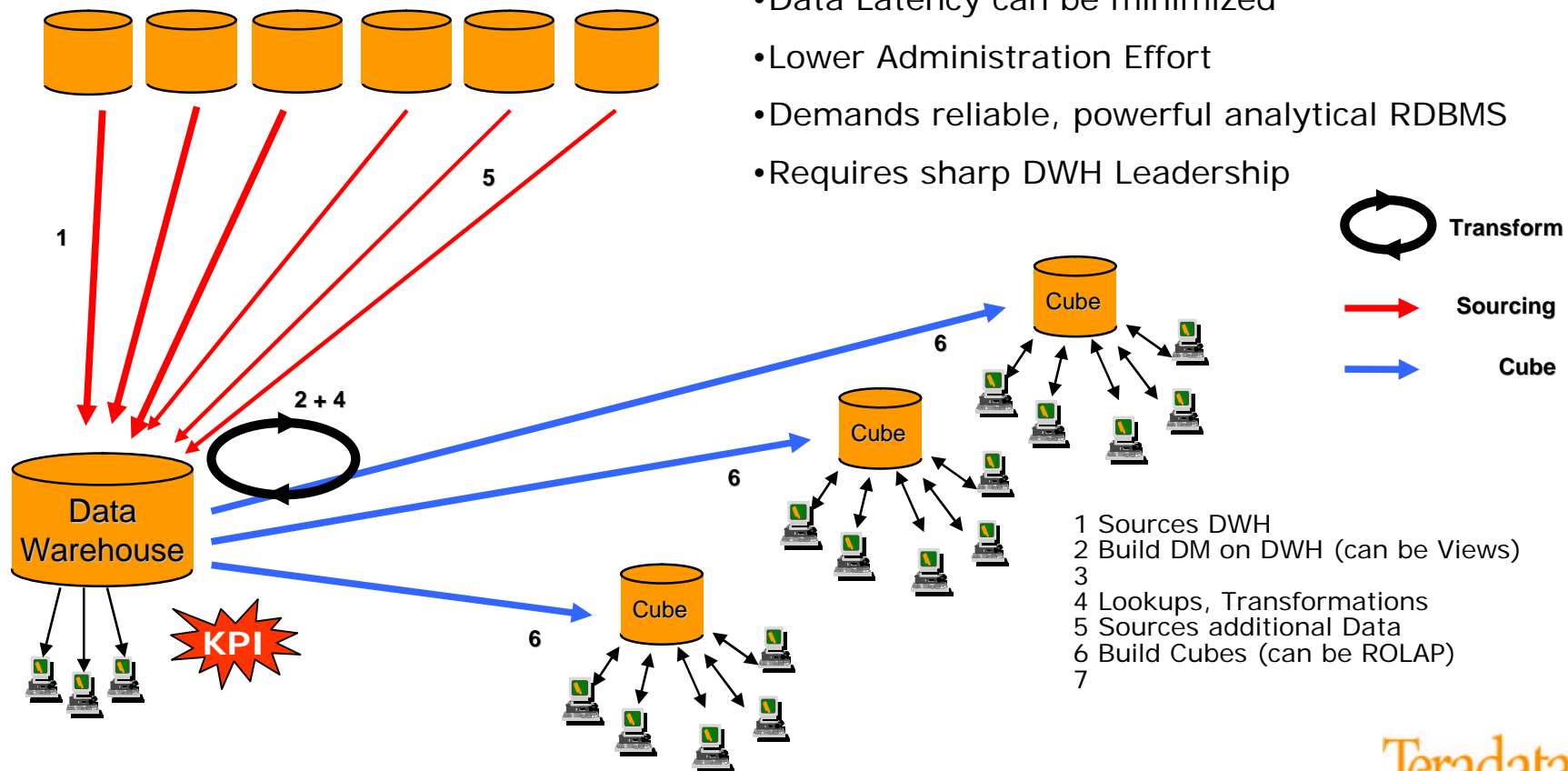
## Source Systems



# Centralized DWH Architecture Approach

*Information Processing (ODS not included)*

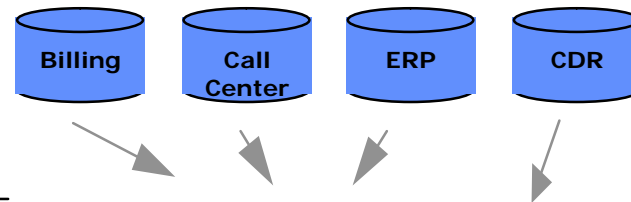
## Source Systeme



# NCR Data Warehouse Architecture

Source Data

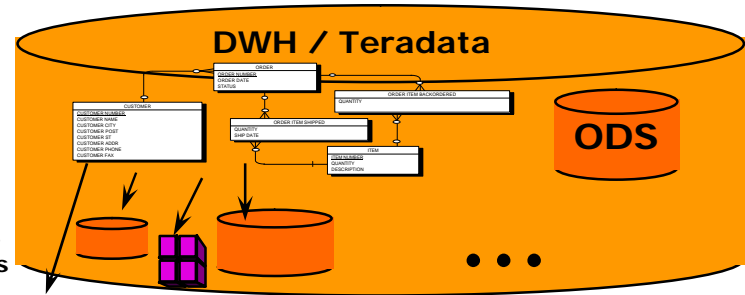
Operational Data



Data Transformation

„Single Version of truth“

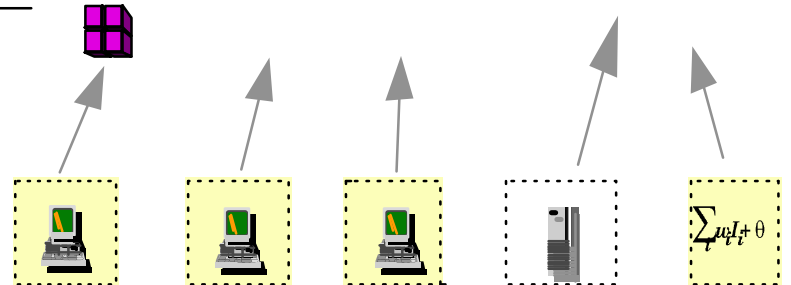
The „HEART“  
of the Data Warehouse



Logical Data Marts  
Physical Data Marts  
ROLAP

MOLAP

Use INFORMATION  
And support  
PROCESSES



Business Users

End User

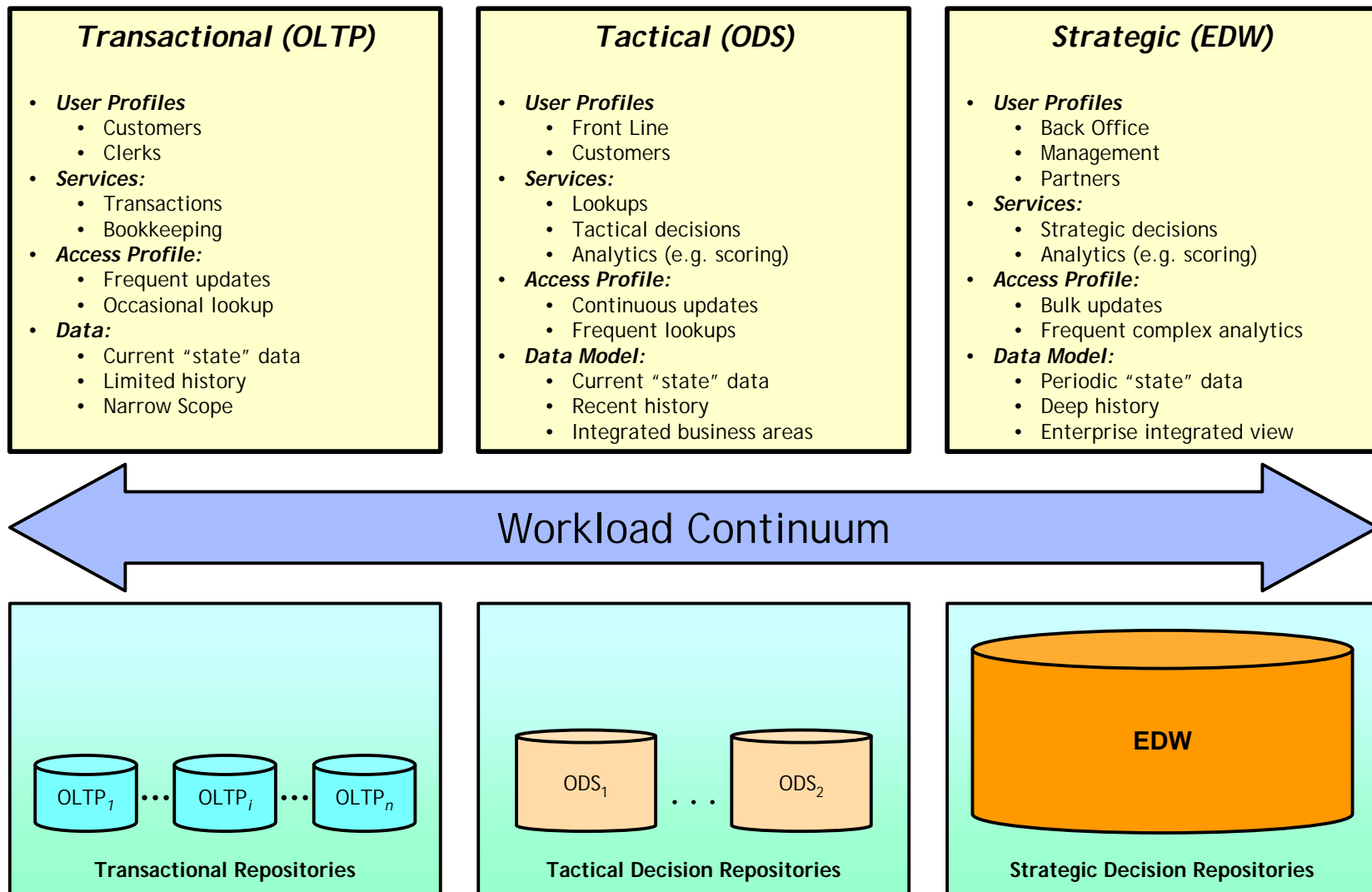
Poweruser Environment

Standard-reports

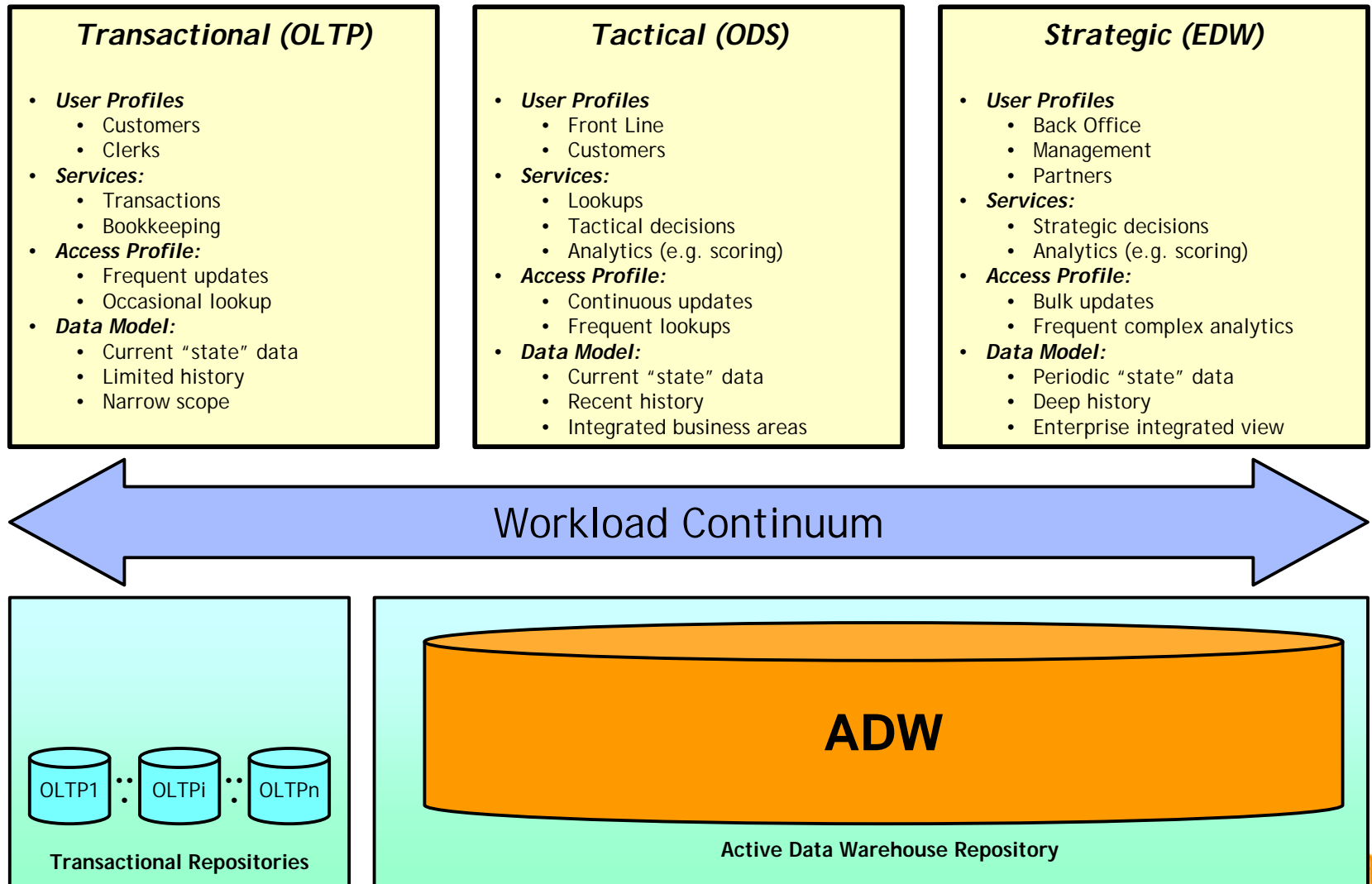
DWH-Operative Application

Data Mining

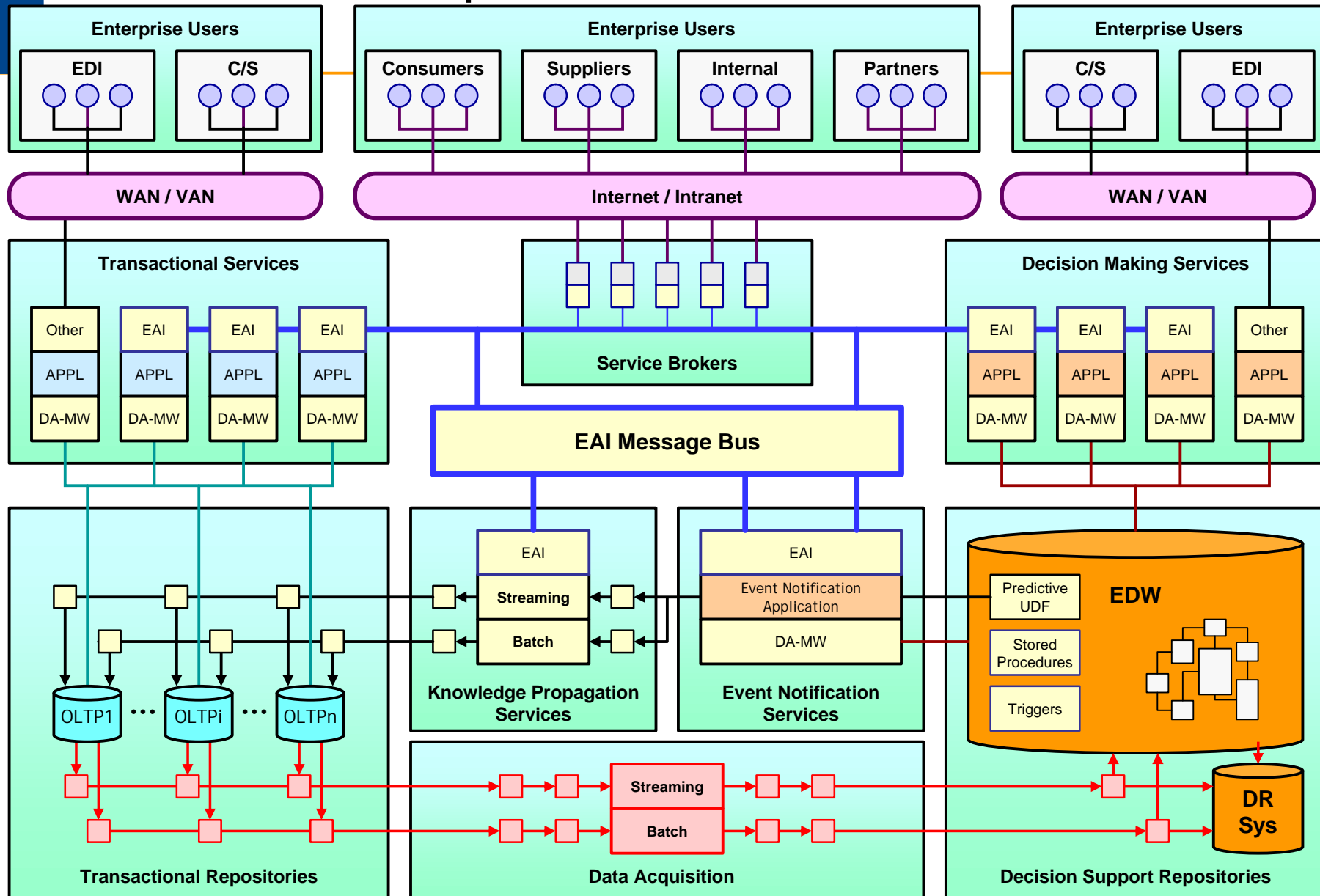
# IT Workload Continuum



# Active Data Warehouse Range

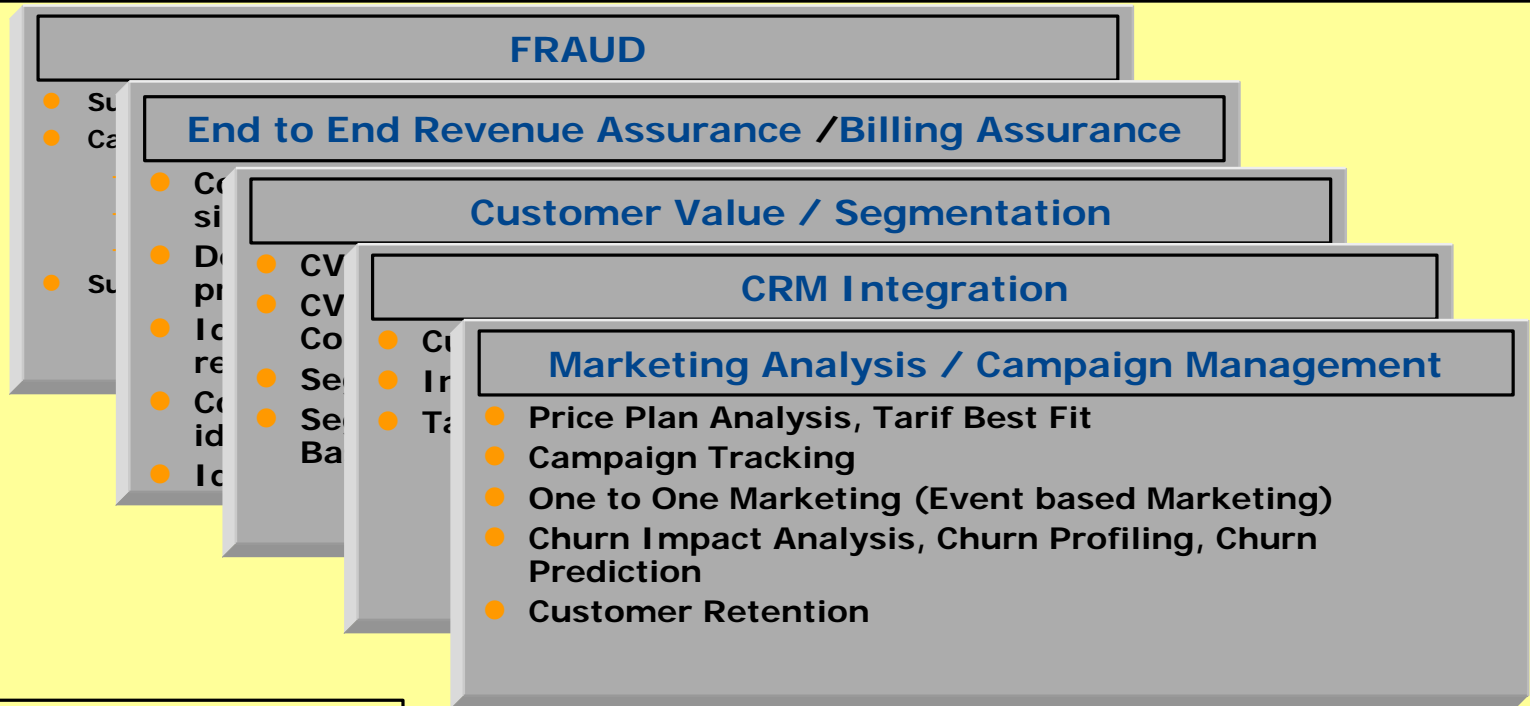


# Real-Time Enterprise Reference Architecture



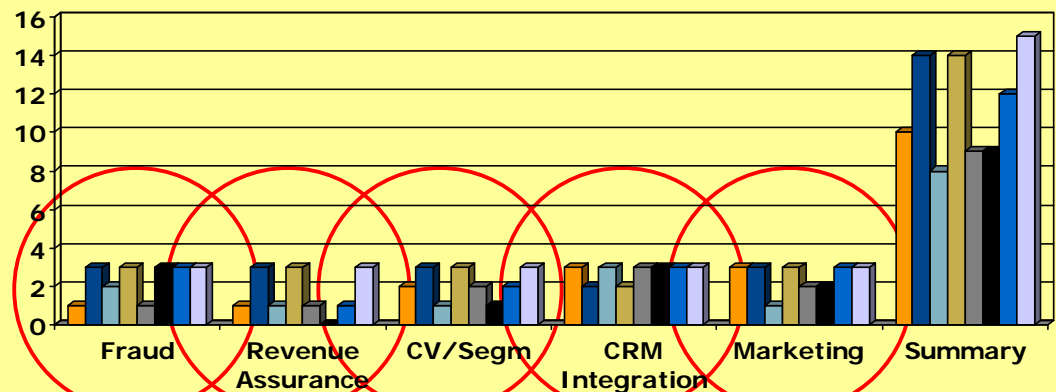
# DWH Requirements to support Business Activities

(Examples Telecom Industry)



## DWH Requirements

- Channel Integration
- Complex Analysis
- Tactical Queries
- High Data Volume
- High User Volume
- 7x24, SLA
- Actual Data
- Direct User Access to Detail Data



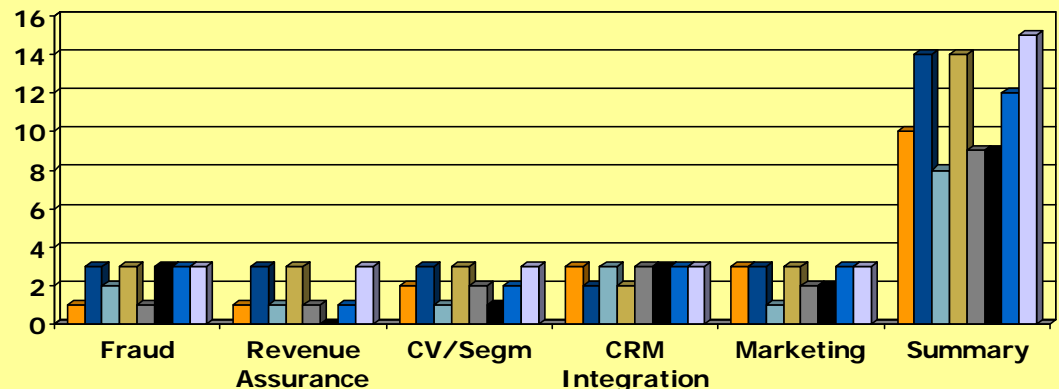


# DWH Requirements to support Business Activities

- Enable Mixed Workload, SLA, **Guarantee System Resources to different User Groups (dynamically !)** (e.g.: Analysts can run Queries, when also Data Load Processes are active)
  - *Priority Scheduler*
  - *TDQM*
- Effective Use of System Resources to **provide Complex Analysis, AdHoc Analysis and Performance**
  - *Teradata All Parallel Architecture (parallel join, sort, aggregate,...)*
  - *Teradata Optimizer*
  - *Partitioned Primary Index*
  - *Value List Compression*

## DWH Requirements

- Channel Integration
- Complex Analysis
- Tactical Queries
- High Data Volume
- High User Volume
- 7x24, SLA
- Actual Data
- Direct User Access to Detail Data

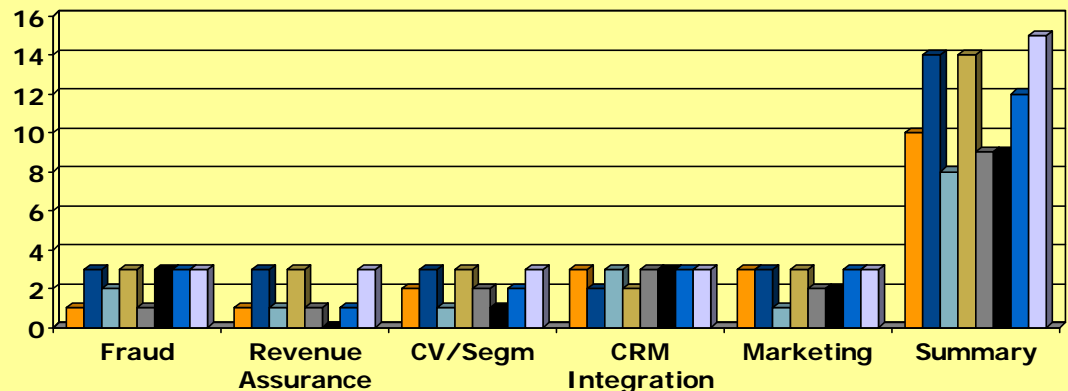


# DWH Requirements to support Business Activities

- Low System & Database Administration, parallel aware cost based Optimizer to **support Complex Analysis, AdHoc Queries, Direct User Access on Detail Data** (e.g.: relieve IT support for Complex Queries)
  - *Teradata Automatic Management* (Space Mngmt, Data Placement, Data Partitioning, Workspace Management)
  - *Teradata Parallel Aware Optimizer*
- High availability and reliable Tools to **support 7x24 access** (e.g. lower downtimes for expected, unexpected downtime and data maintenance)
  - *Availability Functions built in Teradata*
  - *Checkpoint Restartable Parallel Load Utilities*

## DWH Requirements

- Channel Integration
- Complex Analysis
- Tactical Queries
- High Data Volume
- High User Volume
- 7x24, SLA
- Actual Data
- Direct User Access to Detail Data

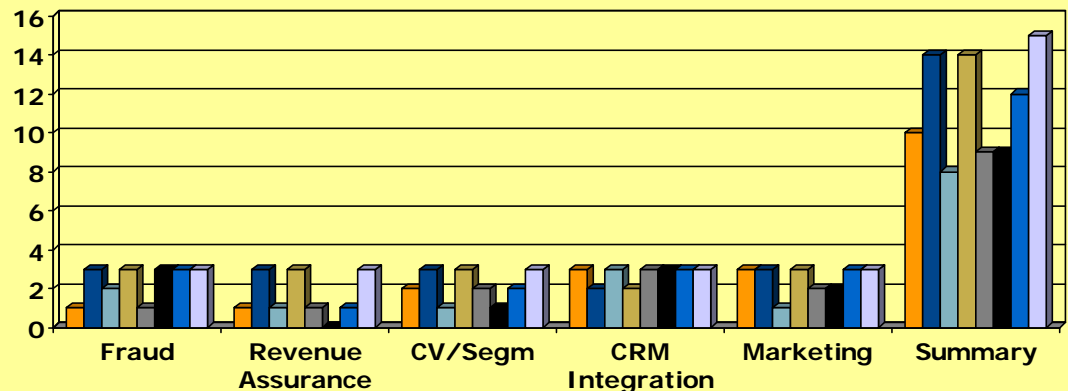


# DWH Requirements to support Business Activities

- Easy grow of the environment in all aspects to **support more Data, more users, additional subject areas, higher performance, query complexity** (enable IT to support new DWH requirements fast)
  - *Teradata Scalability* (Teradata Database, Teradata Hardware, Loading Tools, Backup)

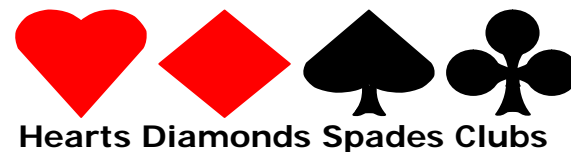
## DWH Requirements

- Channel Integration
- Complex Analysis
- Tactical Queries
- High Data Volume
- High User Volume
- 7x24, SLA
- Actual Data
- Direct User Access to Detail Data



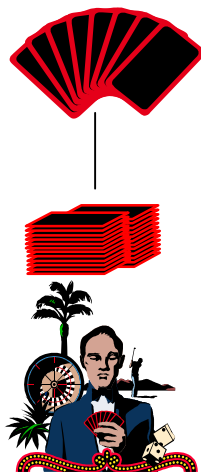
# Parallel versus Single Workload

## Counting Aces



### Single work load

40 cards

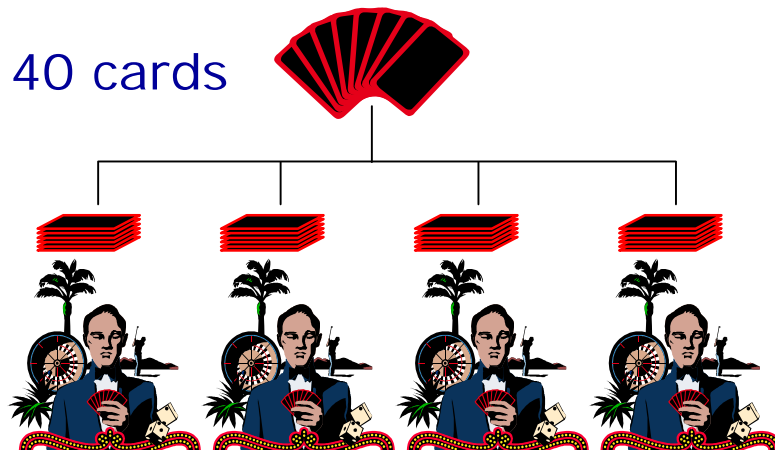


40 cards

Time of a single worker scanning  
all 40 cards

### Parallel work load

40 cards

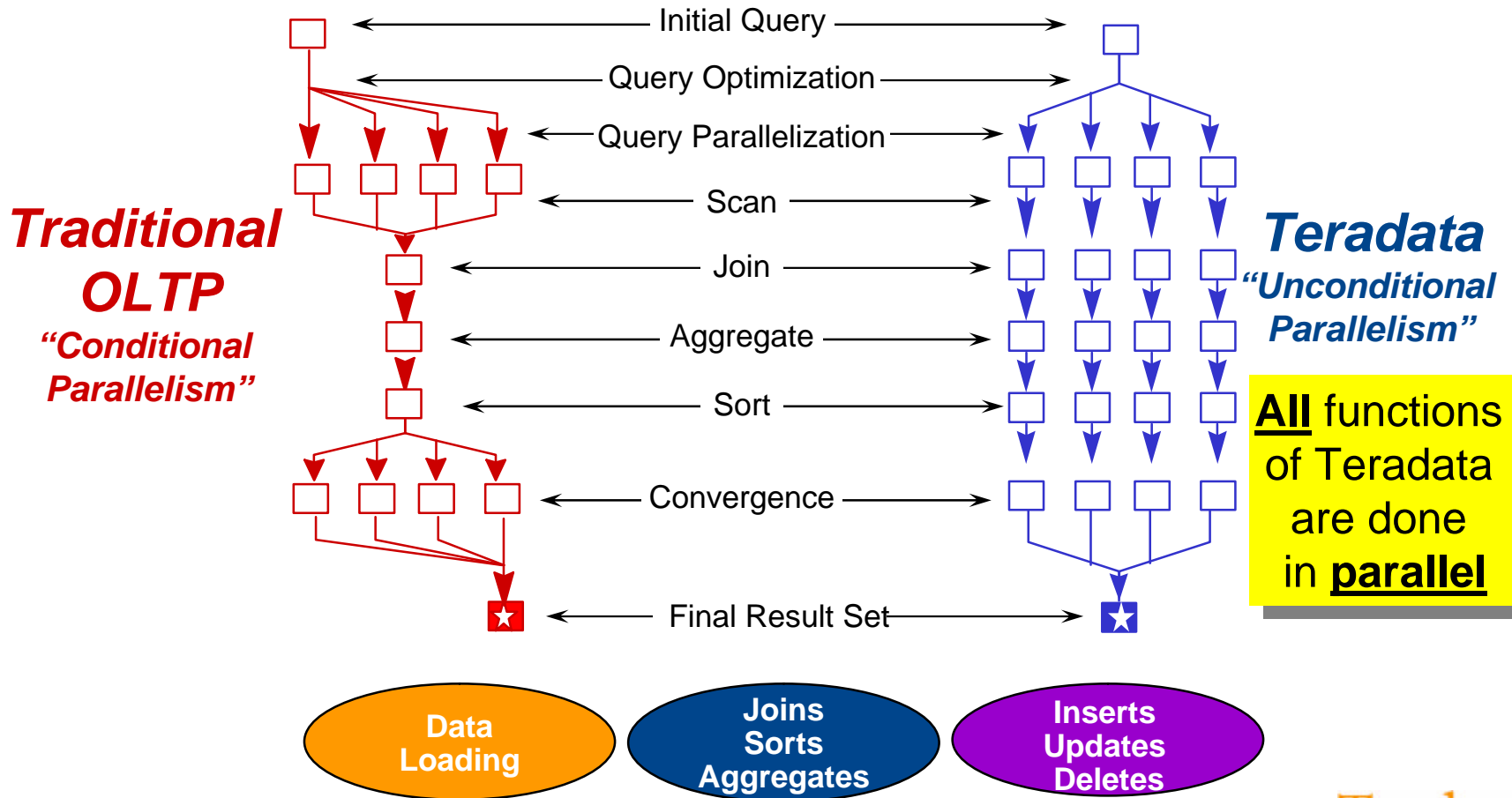


10 cards 10 cards 10 cards 10 cards

Time of 4 workers scanning just  
10 cards in parallel

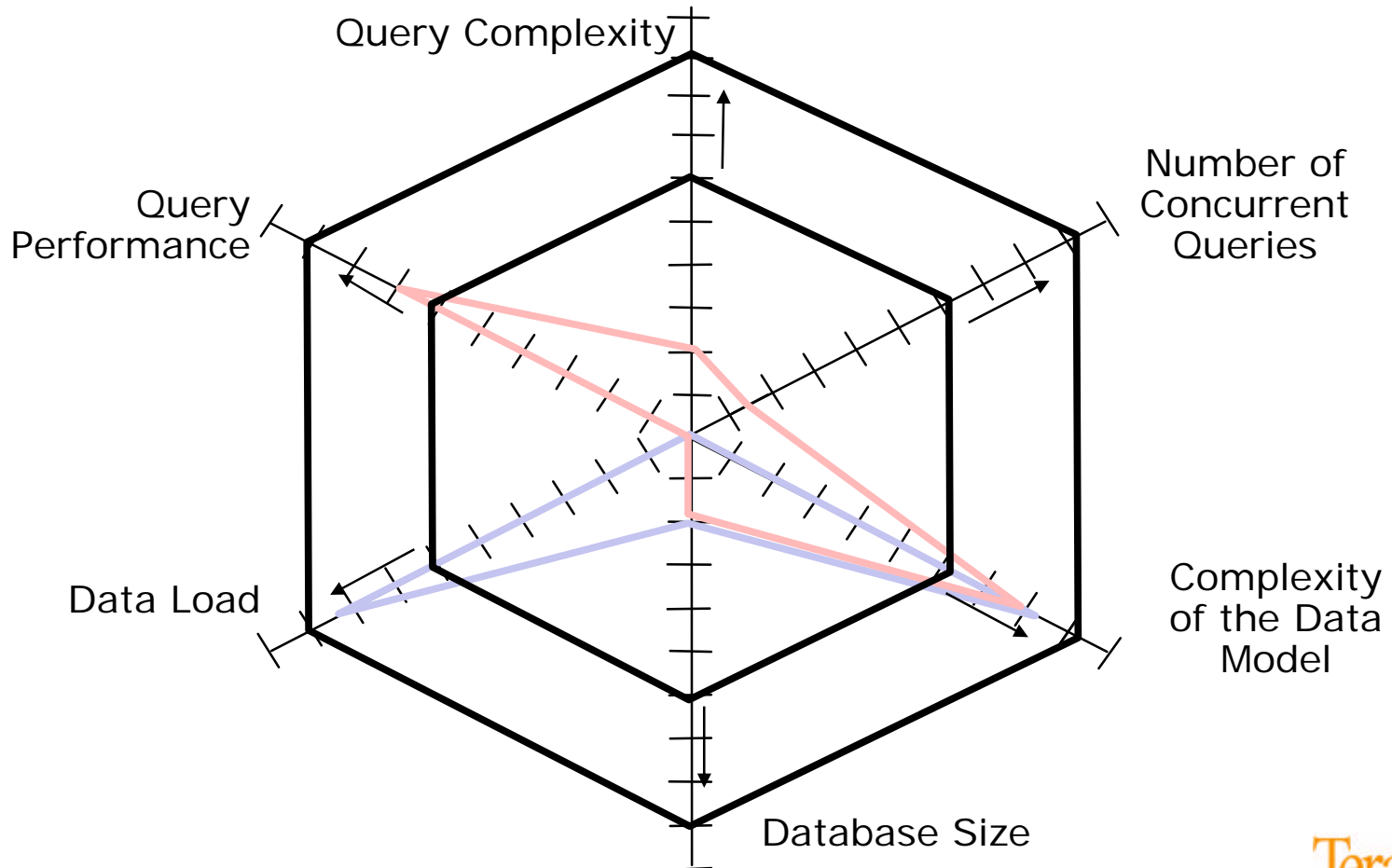
# Technical Issues of Parallelism

**Each UoP (Unit of Parallelism) does an equal amount of work**



# Definition of Scalability

## Scalability across dimensions



# Availability and Manageability

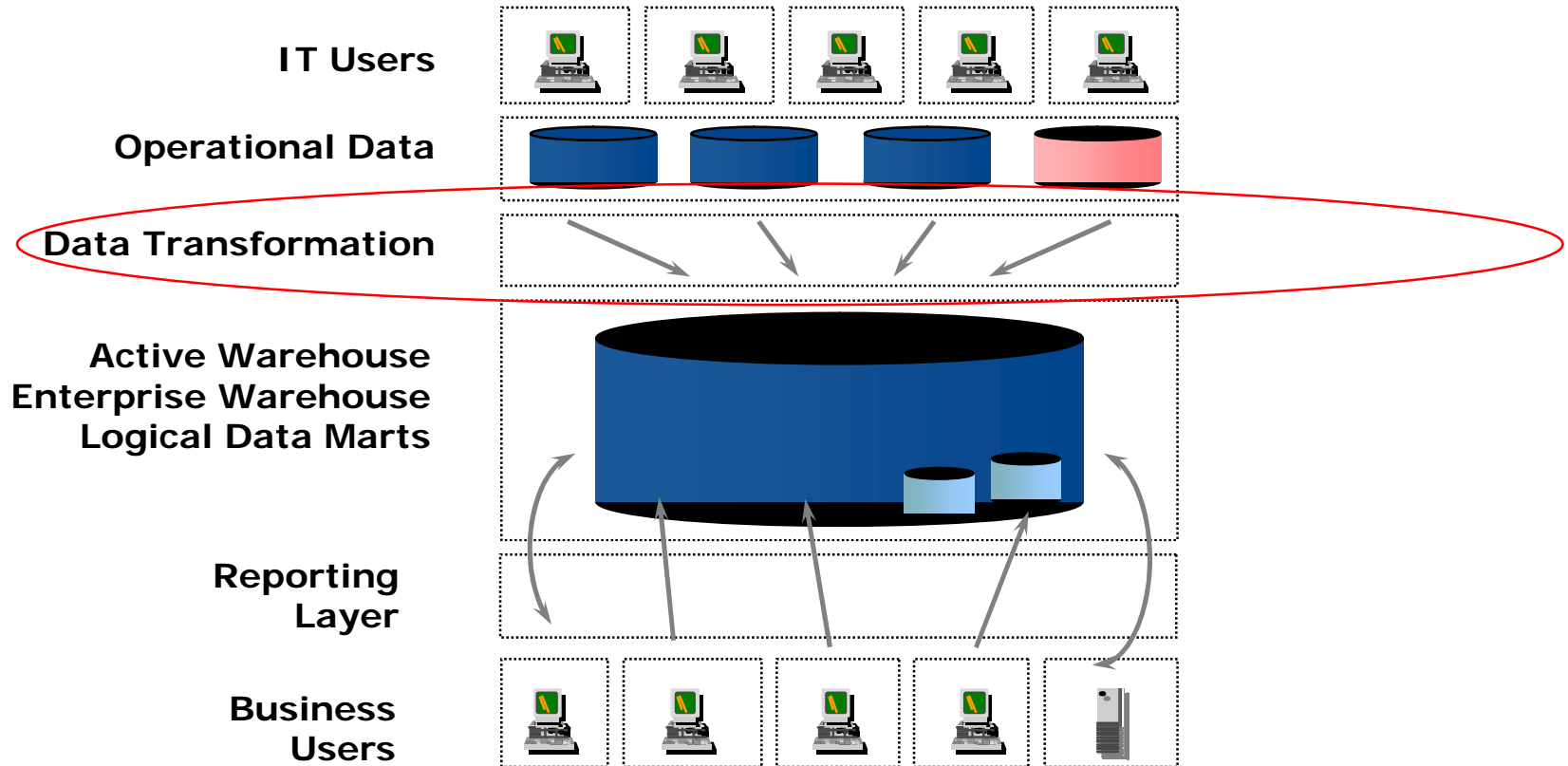
- Availability

- Hardware Availability
- Database Availability (Data Maintenance)
  - End User Data Availability
  - Response time
  - SLA

- Manageability

- ETL Process
- Database Administration
- OLAP Administration
- End User Query Support

# Data Warehouse Architecture





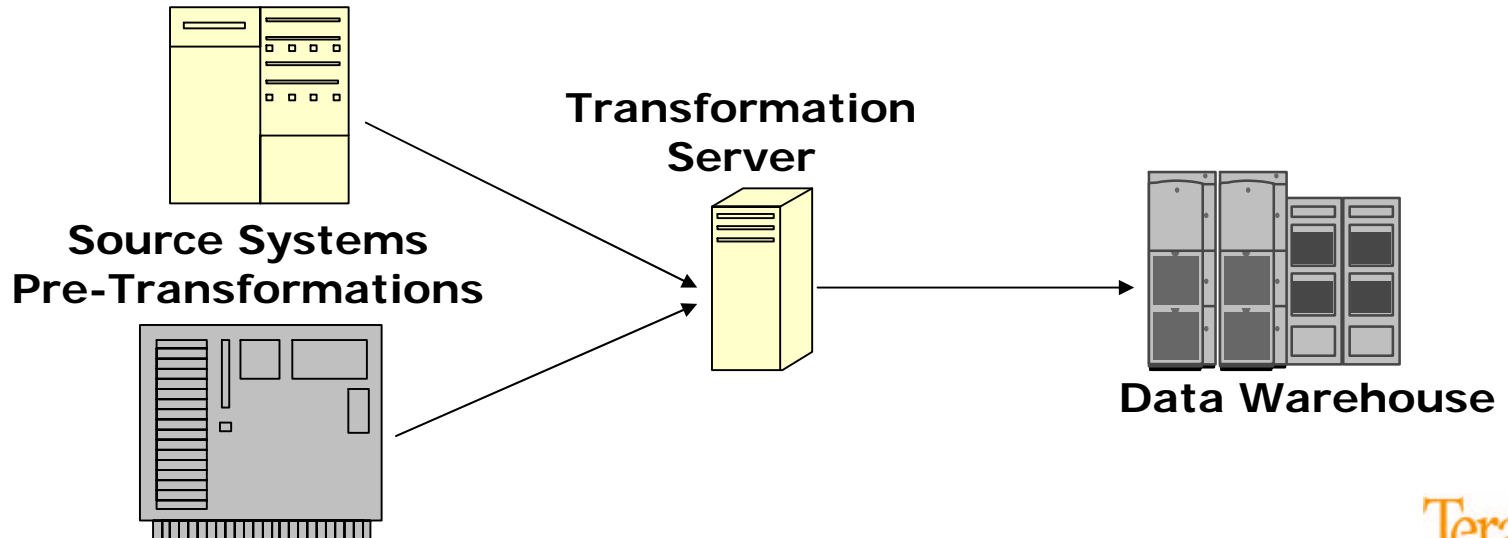
# DWH Load Strategies ETL vs. ELT

- ETL - all transformation occur prior to being loaded to database.
  - Most ETL tools are designed to do transformation prior to loading to DBMS.
  - Simpler processing design particularly for non-Teradata trained people who tend to think in record at a time mode.
- ELT - data loaded to database where some transformations occur using SQL and the database.
  - Power of Teradata utilized to perform potentially massive transformations by using SQL, Insert Selects, Stored Procedures, and Triggers
  - Way of moving data into tables in stages to allow result consistency for query operations or to avoid locking issues.
  - May be less expensive for customers to add capacity to Teradata system for this workload than to have resources needed for a large SMP ETL server.
  - Hummingbird and ETI are the ETL vendors that support these types of processes.

# DWH Load Strategies – ETL

## ETL : Extract - Transform - Load

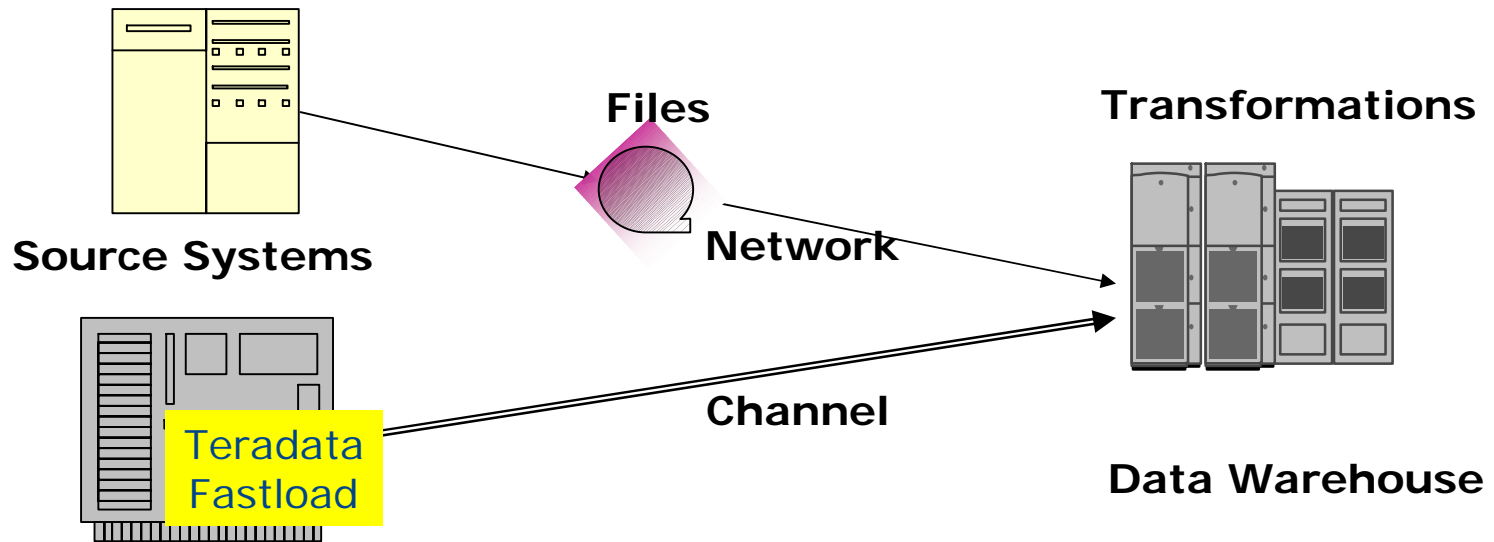
- Extract with pre-transformations from source system
- Transformation on a transformation server
- Load one-to-one into target database (TD) without further transformations
- Example: Informatica, ETI, AbInitio



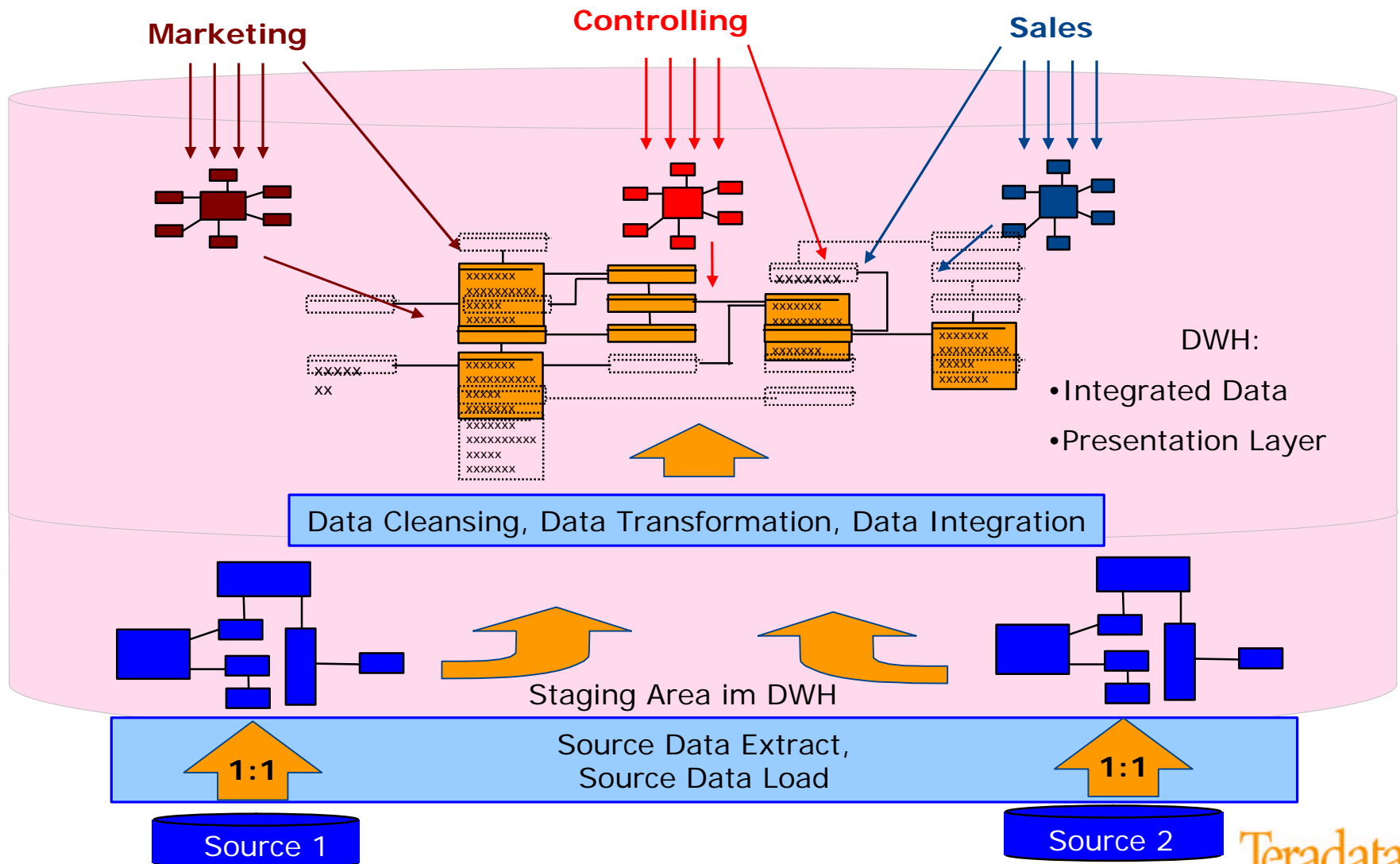
# DWH Load Strategies – ELT

## ELT : Extract - Load - Transform

- Extract one-to-one from source system
- Load one-to-one into Teradata staging area
- Transform on Teradata
- Load/Insert into Target Teradata Tables



# DWH Load Strategies – ELT



# DWH Load Strategies +/-

## ETL

- + less space on DWH
- + DWH resources free for queries
- + single point of transformation definitions
- transformation space on source system or powerful transformation server
- transformation server idle during off batch hours
- potential bottleneck

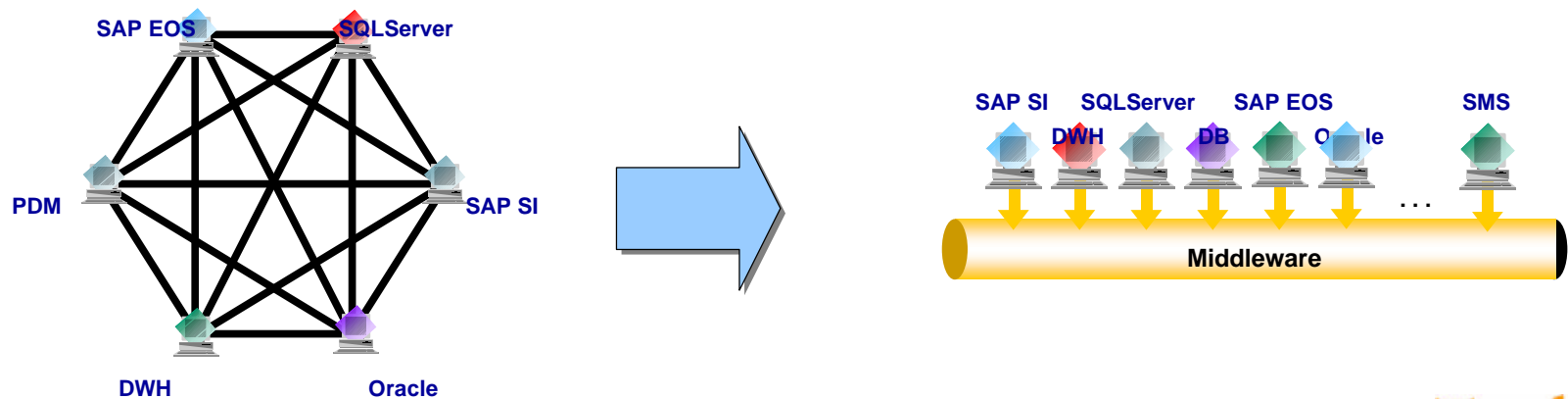
## ELT

- + transformations on all AMPS (fast !)
- + optimal resource usage
- + transformation power scales with DWH
- more space on DWH
- transformation may slow down queries during batch load hours

# DWH Load Strategies – EAI

## EAI : Enterprise Application Integration

- Supports the delivery and/or retrieval of data between applications/DWH.
- Supports the harmonisation of transformation data content.
- Supports the process flow of data between applications
- EAI will reduce the interfacing efforts.



# Definition - EAI

- **Enterprise Application Integration (EAI)** refers to messaging architectures that acquire, store, transform and intelligently deliver business information to other applications, processes, data bases and data stores, in real time.  
EAI is the unrestricted sharing of data and business processes among any connected applications and data sources in the enterprise.  
'Enterprise Application Integration', David S Linthicum
- **"EAI-style integration technology** is focused on process automation. This type of solution assumes that the goal is to complete automation of a process, parts of which are currently automated in applications that are specific to business functional domains. The application must also extract some data (representing a business event) from an application, move it, transform it, and ultimately deliver it to some other application to continue the automation." (Source: E-Business Integration: Differentiating EAI and ETLM, Daniel Sholler, Meta Group, 16 October 2000)
- **Teradata EAI interfaces**
  - **Tibco** adaptor, partnership agreement is in process
  - **Vitria** adaptor, partnership in place
  - Teradata Access module for **IBM MQ Series**

# ETL vs. ELT vs. EAI

- ETL and ELT

## Typical DWH Batch Loading Concepts

- Source Data complete
- Start Loading and Transformation Batch
- Load Finished -> User Access

- EAI

## Enterprise data and process flow concept

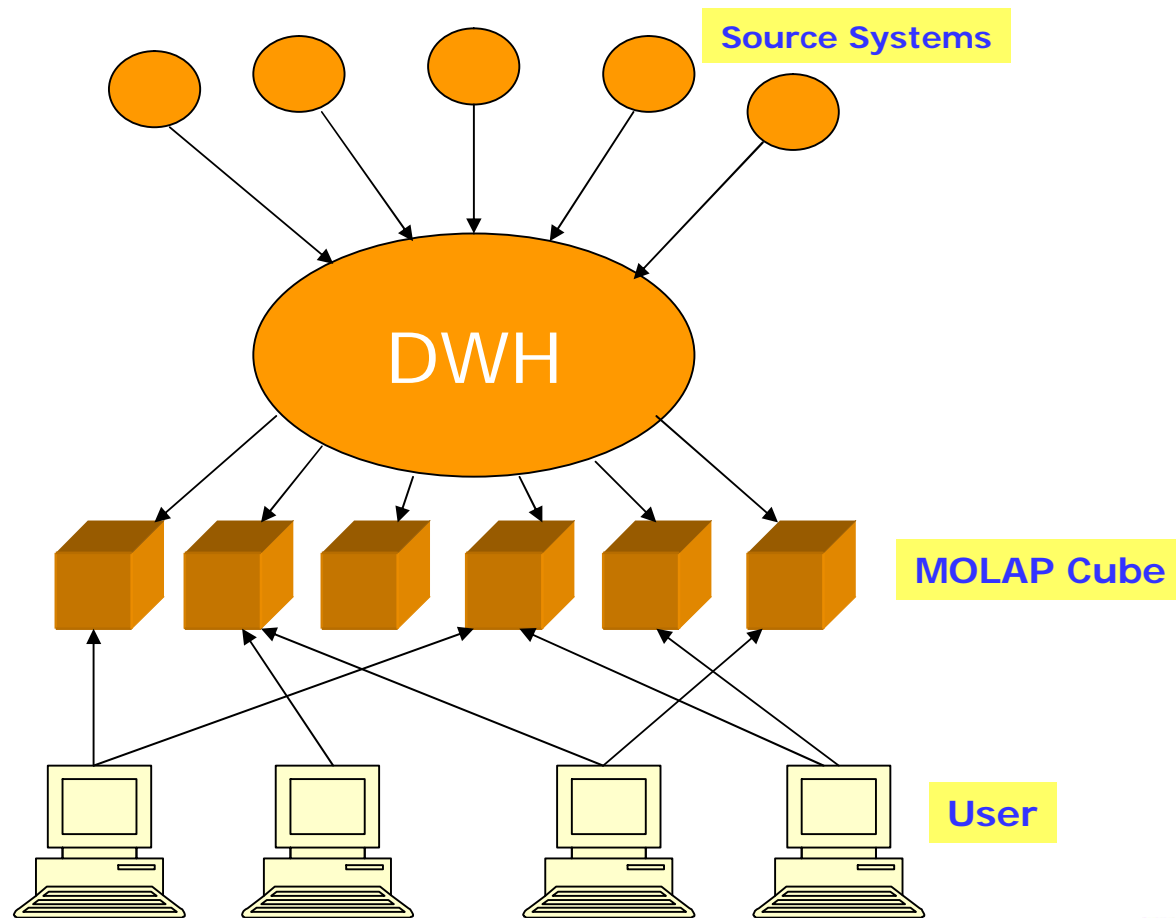
- DWH is one component of the whole concept
- Data is delivered continuously
- Loading is done with continuous load streams. (Trickle feed)
- Loading and User Access is done at the same time.



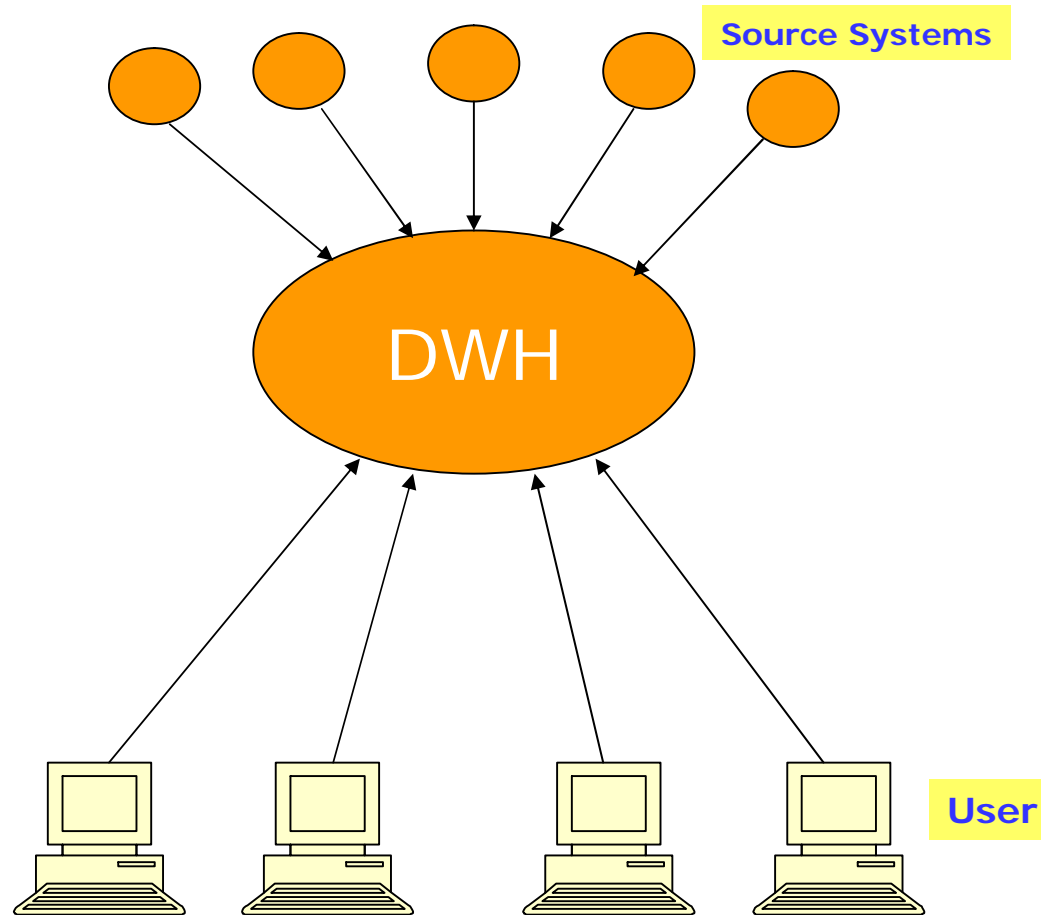
# OLAP, ROLAP, MOLAP

- OLAP – On-Line Analytical Processing
- ROLAP – Relational OLAP
- MOLAP – Multidimensional OLAP

# MOLAP Architecture



# ROLAP Architecture



# ROLAP vs. MOLAP

## ● ROLAP

- + Analysis to detail data (Drill-Down)
- + Analysis across all data areas within the DWH (Drill-Cross)
- + All Data can be linked
- + Data can be written back to the DWH (Segmentierung, etc.)
- + Always DWH-actual
- Indices und Aggregates necessary for performant access

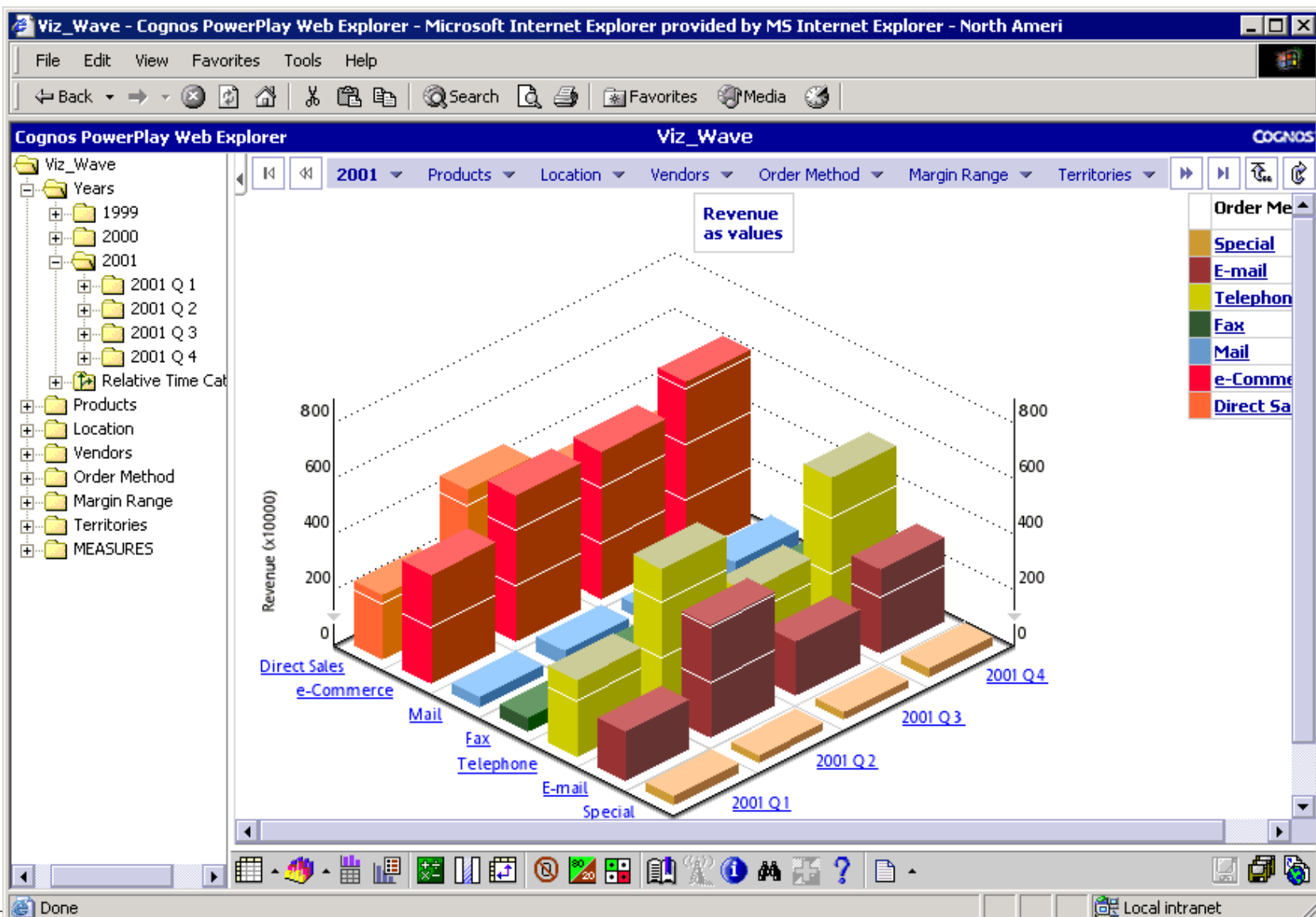
ROLAP: Free Analysis  
across all DWH data

## ● MOLAP

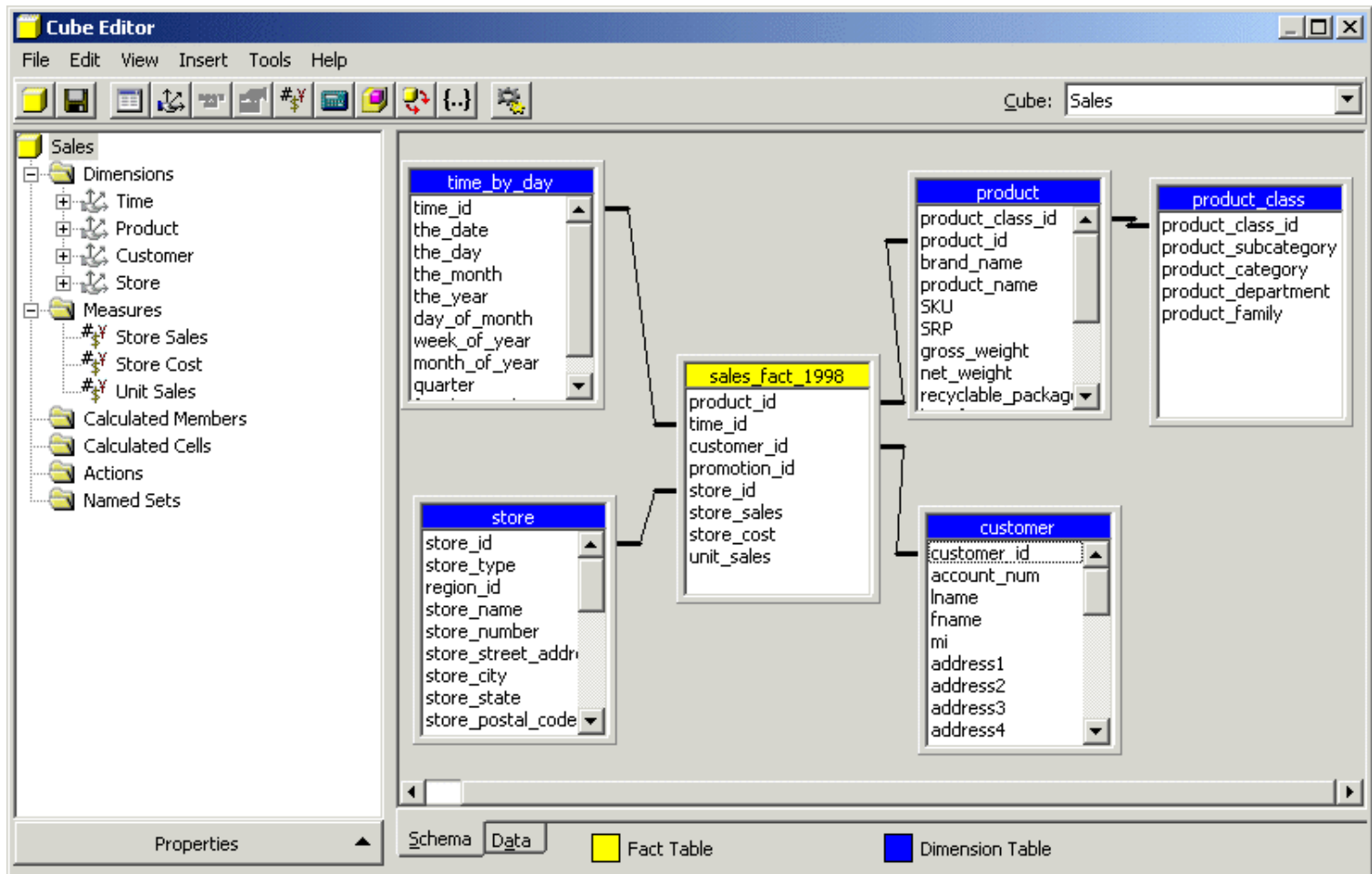
- + High Performance
- + Data is available Off-Line of the DWH
- Analysis only within defined Cubes possible
- Only defined data links
- High Maintenance Effort

MOLAP: Performance for  
defined Analysis

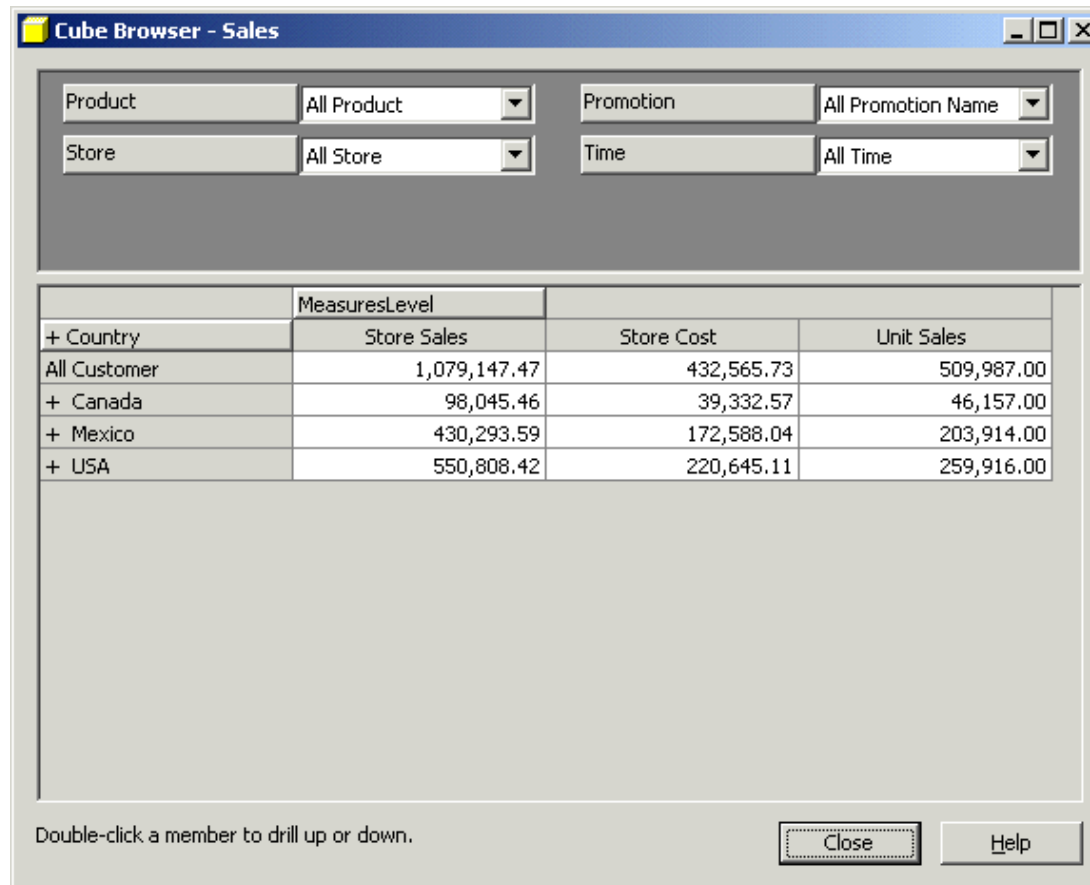
# OLAP



# OLAP: Facts and Dimensions



# OLAP: Drill Down, Drill Across



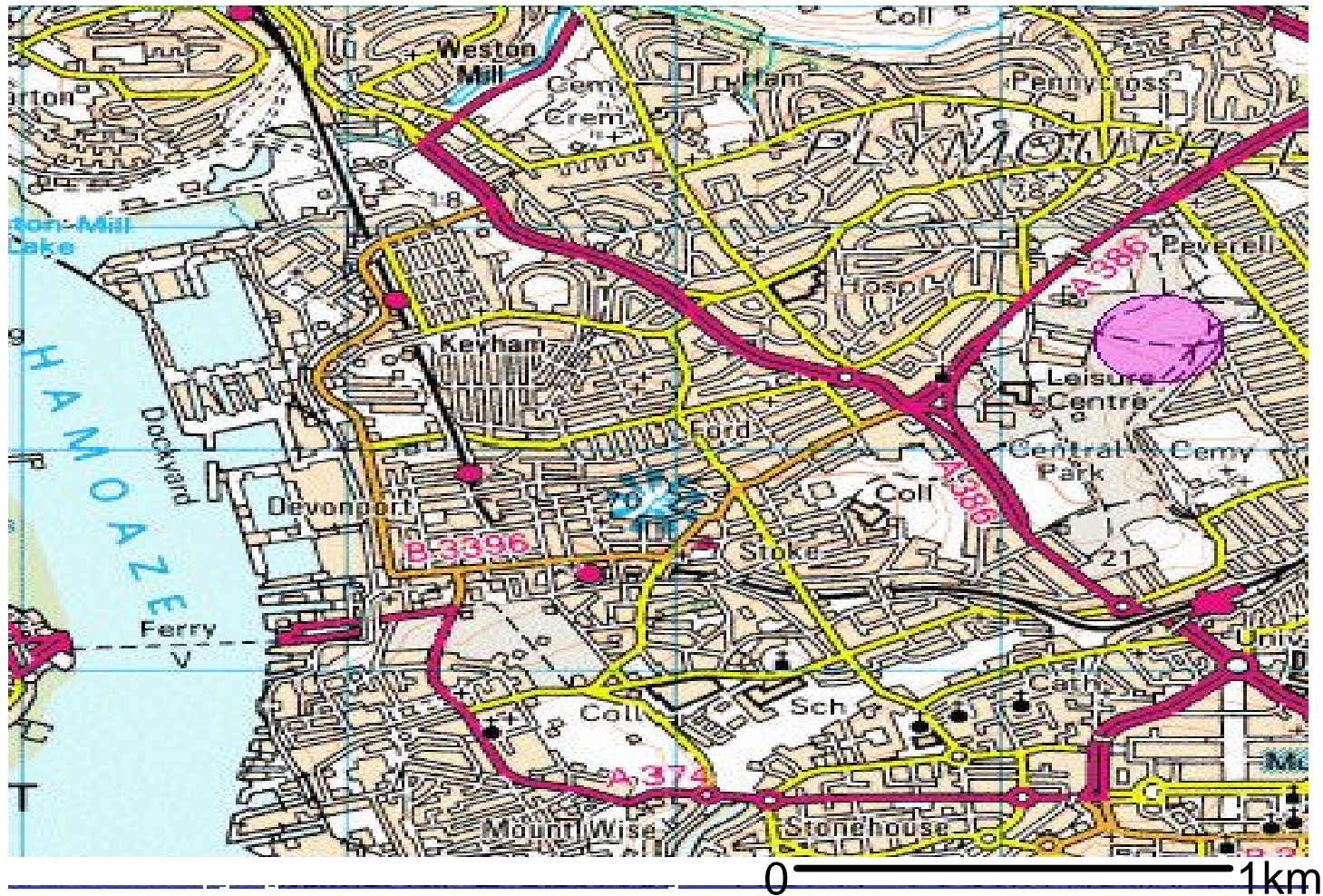
Product: All Product   Promotion: All Promotion Name  
Store: All Store   Time: All Time

	MeasuresLevel		
+ Country	Store Sales	Store Cost	Unit Sales
All Customer	1,079,147.47	432,565.73	509,987.00
+ Canada	98,045.46	39,332.57	46,157.00
+ Mexico	430,293.59	172,588.04	203,914.00
+ USA	550,808.42	220,645.11	259,916.00

Double-click a member to drill up or down.

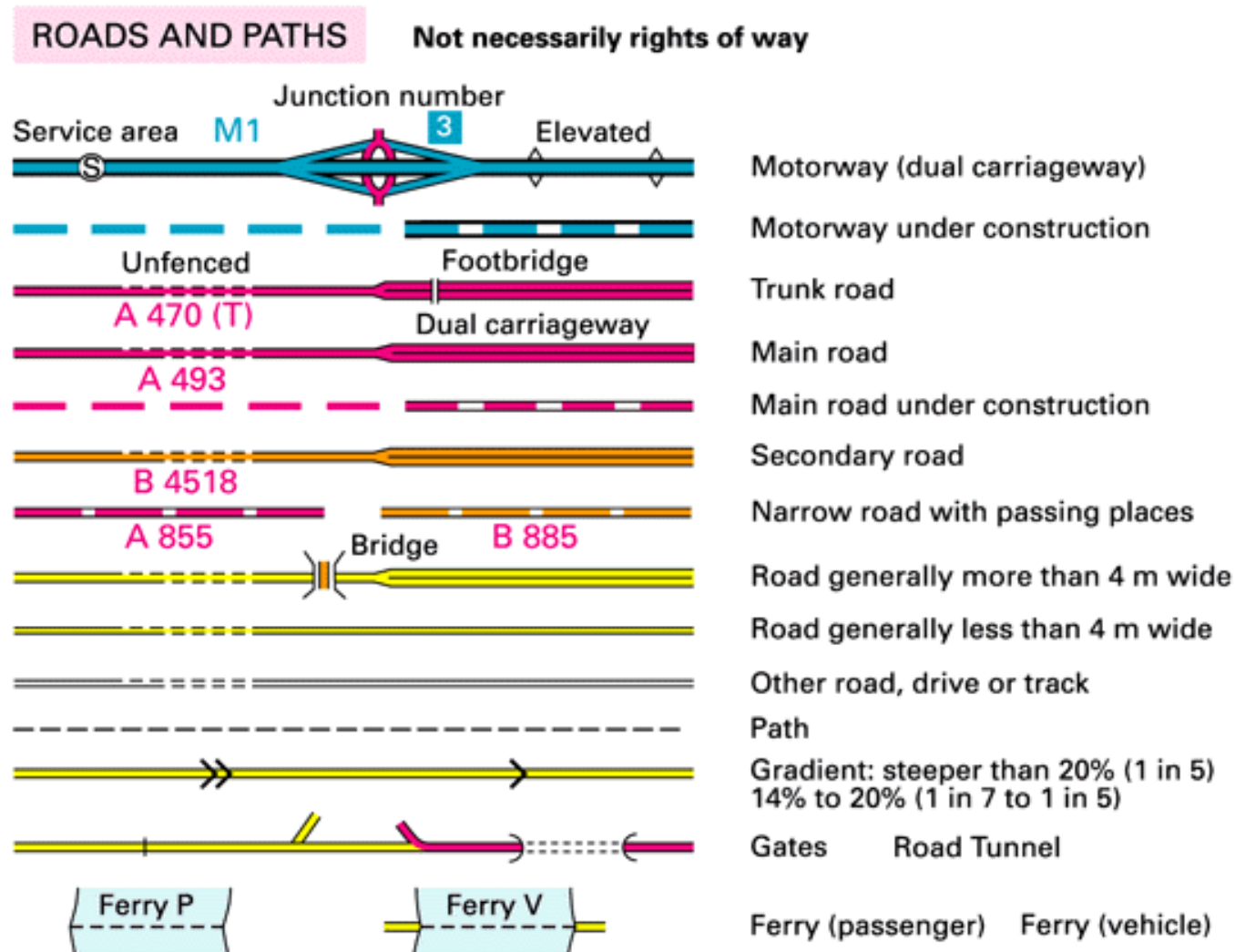
Close   Help

# DATA



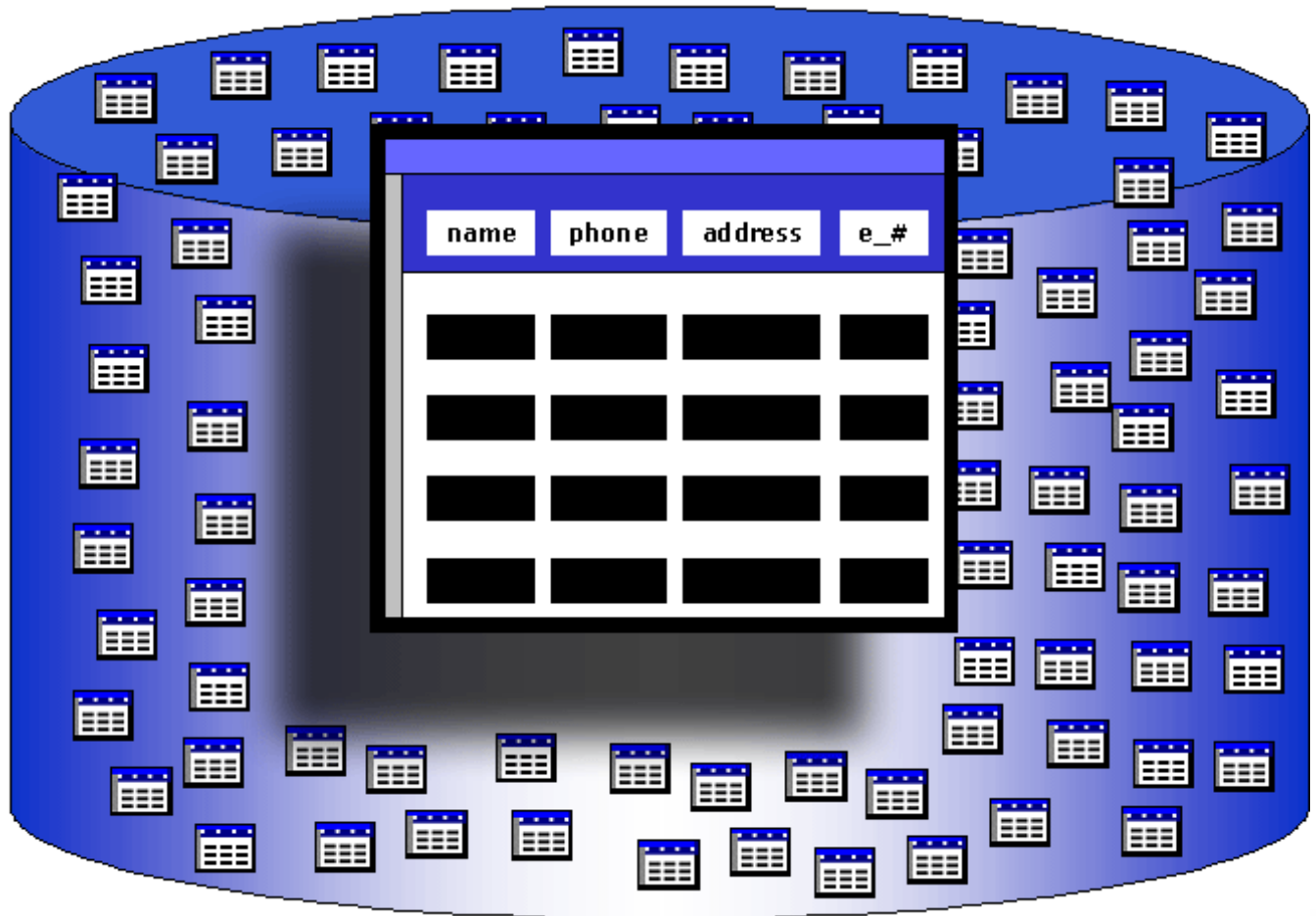


# METADATA



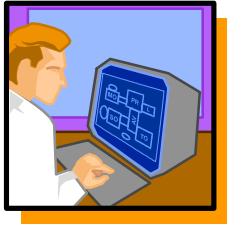
# Warehouse Meta Data

You...



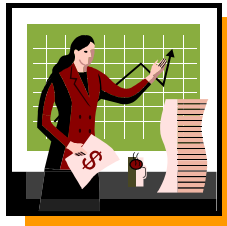
Looking for information.

# Types of Metadata



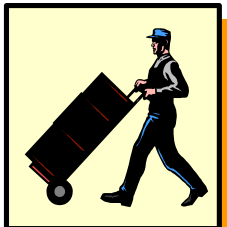
## 1.) Technical Metadata

Descriptions of databases, tables, columns, views, indexes, and constraints



## 2.) Business Metadata

Definitional Information used in locating, understanding, and accessing information.



## 3.) Data Movement Metadata

Descriptions of when, from where, and how data was moved into the Warehouse

# What is the Value?

## Values



**Reduce IT Related Problems**

**Increase System Value**

## ROI Measures



**IT staff is less likely to make programming errors when making system enhancements, since all impacted programs, tables, and fields are identified.**

**The Data Warehouse has greater relevance for the business users, letting them do their jobs more efficiently.**

# What is the Value? (cont'd)

**Values**



**ROI Measures**



**Reduce Redundant Data**

**Enables identification of redundant data. In addition, this ability reduces the likelihood of building systems containing redundant data in the first place.**

**Improve Business Decision Making**

**Users are able to access and specifically understand the information they need to make business decisions.**

# What is the Value? (cont'd)

## Values



## ROI Measures



**Reduce Impact of  
Employee Turnover**

**Documents the knowledge that may otherwise be known only by the developer who built the programs, and makes it available to the entire IT staff.**

**Improve System Performance**

**Removes redundant data and processes thereby improving system performance.**



- [MDS Overview](#)
- [Help](#)
- [New Search](#)
- [Sample Glossary](#)

The following Models are in your repository. To access data in a specific Model simply click on the Model name:

DatabaseModel

Logoff

## Database Information:

**Database Name :** E-Business

**Description :** This Database contains the tables associated with the FS E-Business LDM. The ERwin data was loaded through the Meta Integration Model Bridge (MIMB).  
[Click Here](#) for a quick view of the Model

**System :** [FSLDM](#)

Property	Value
----------	-------

CommentString

### Tables

Table Name	Description
1. <a href="#">T0000_CHANNEL_DEMOGRAPHICS</a>	<p>The Channel Demographics derived table was developed to provide the user with demographic_value_id's for each channel_id/channel_type_code. The demographic items included are listed in the CHANNEL_DEMOGRAPHICS view. Below is the SQL to populate this table:</p> <pre>insert into T0000_CHANNEL_DEMOGRAPHICS select * from CHANNEL_DEMOGRAPHICS</pre> <p>Note: The Channel_Demo_Start_Date and Channel_Demo_End_Date columns need to be updated after the table is populated from the CHANNEL_DEMOGRAPHICS view.</p>
2. <a href="#">T0001_PARTY_DEMOGRAPHICS</a>	<p>The Party Demographics derived table was developed to provide the user with demographic_value_id's for each customer. The demographic items included are listed in the PARTY_DEMOGRAPHICS view. Below is the SQL to populate this table:</p> <pre>insert into T0001_PARTY_DEMOGRAPHICS select * from PARTY_DEMOGRAPHICS;</pre> <p>Note: The Party_Demo_Start_Date and Party_Demo_End_Date columns need to be updated after the table is populated from the PARTY_DEMOGRAPHICS view.</p>

# DWH Data Modelling

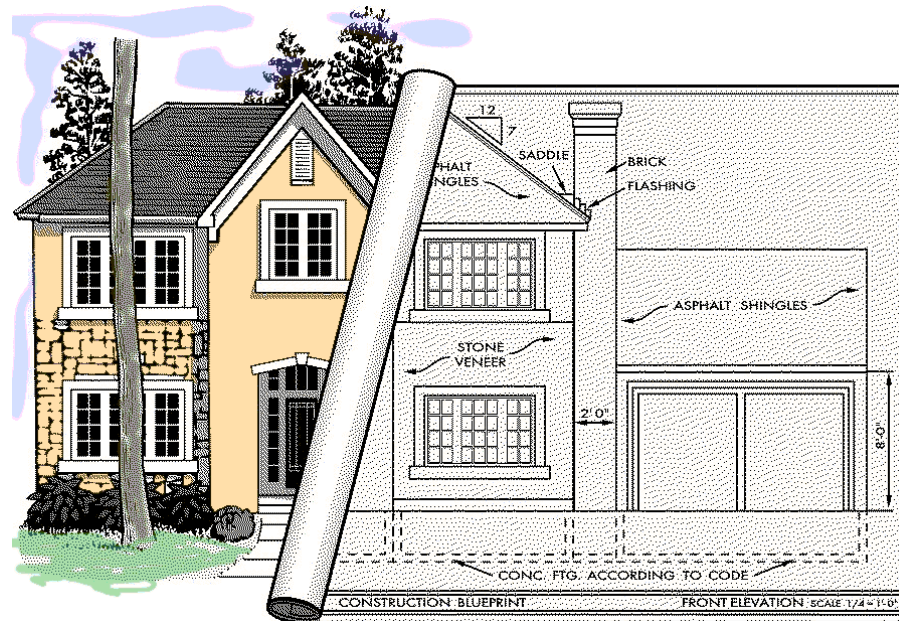
You've never seen your business like this before.



# What is a Logical Data Model?

- Specification of the data structures and business rules needed to support the business
- Includes data about those things, events and concepts of interest to the bank
- A representation of the data organization independent of physical storage or presentation format.
- A neutral cross functional view - not Accounting's view or Marketing's view
- Non-redundant - a fact is represented in one place - in the right place

# LDM: Basis for structured DWH development cycles, Roadmap



# Physical or Logical (OR BOTH?)

**Physical** - Designed for efficiency. May denormalize, may aggregate, may replicate. May (almost certainly will) bias efficiency toward certain users

**Logical** - Designed for effectiveness of extracting *sharable* information from diverse sources of data

# LDM/PDM Differences

## Logical Data Model (LDM)

- model of the business
- document business rules and relationships, use as a tool with business users to confirm understanding
- use to confirm we can answer business questions
- keep as the blueprint - the clean business view before physical changes
- database independent

## Physical Data Model (PDM)

- model of the database
- include physical data attributes and requirements - datatype, size, indexes
- modify for performance after testing - depends on usage - types of queries, frequency, volume (might add derivable fields, merge subtypes, separate history, add summaries)
- database dependent

# LDM/PDM Differences - Terms

## Logical Data Model (LDM)

- Entities, Attributes
- Definitions
- Relationships, Keys (PK, FK)  
(how data is related, business perspective)
- Domains
- Business Names
- Normalized, Third Normal Form (3NF)
- Business View

## Physical Data Model (PDM)

- Tables, Columns
- Descriptions
- Join Paths, Indexes (UPI, NUPI, USI, NUSI, join indexes) (how to access data)
- Domains, Data Types and Size
- Names may be abbreviated
- Denormalized (summary tables, derived data, current vs. history)
- IT view

# What, no LDM ?!?

*(or, Why would a customer skip the LDM?)*

- Case 1: “Fast start”, rehosting, etc. from current systems, datamarts, or data warehouses to Teradata
  - shorter term, quick solution
  - usually followed by “Phase 2”, to include an LDM, especially necessary when integrating and consolidating
- Case 2: “We don’t need one.”
  - time pressures - part of the process, can do work in parallel - don’t need to wait until entire LDM is complete before continuing (though changes may be made)
  - physically-oriented decision makers
  - no experience with LDMs or relational models
  - don’t want to change how it’s always been done
  - more likely on IT-run project

# Risks

- Without LDM
  - database design tends to be query-based - depends on IT's understanding of what the business user needs
  - changes required for new or altered queries, different views of the business
  - start getting duplicate fields, inconsistent data
  - PDM gets cluttered, harder to predict effects of changes
  - no master plan leads towards separate data marts
  - short-term thinking, harder to grow and change

# Data Schemas used for Data Warehousing

*Dimensional / 3NF*

- The fundamental objective for a relational database management system (RDBMS) is to *keep data independent of the applications* or analysis that use it.
  - Different applications require different views of the same data.
  - Data must be extendable and new applications and analysis introduced without having to modify existing applications.
- The key to data independence is data normalization
- Normalized Models are called to be complex
- Normalized data schemas are the most performance demanding
- Normalized Models easily adapt to new requirements without redesign



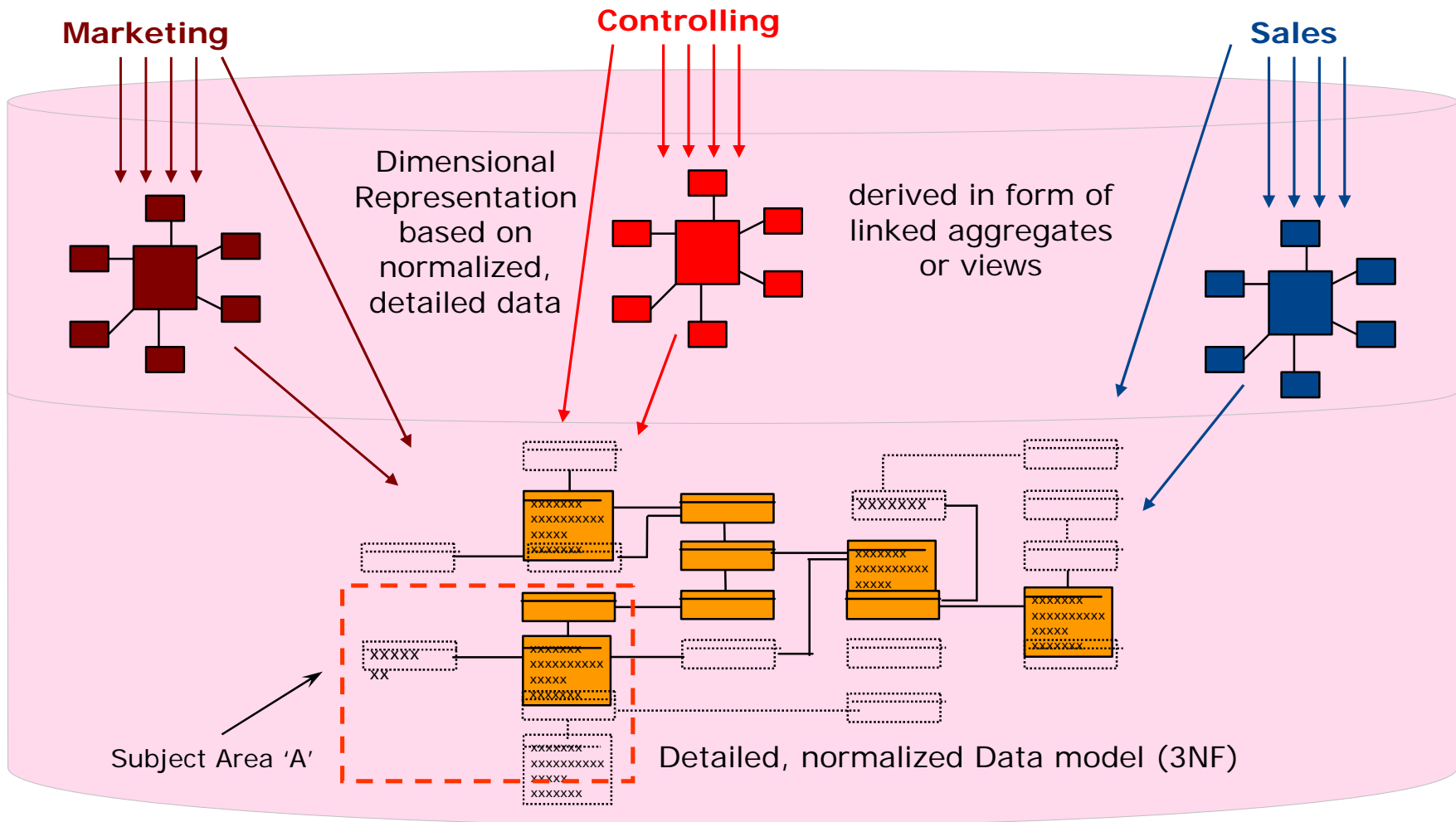
# Data Schemas used for Data Warehousing

*Dimensional / 3NF (cont.)*

- To address the issue of performance in many environments, data independence was abandoned and denormalized schemas were created to address particular analytical needs.
- Dimensional schemas are called to be easy to use
- Dimensional schemas keep redundant data
- Complex ETL procedures
- Analysis is restricted to dimensions
  - meets requirements for the majority of end user analysis
  - extended analysis (additional datasources) requires redesign

# Data Schemas used for Data Warehousing

*Use the best of both: 3NF & Dimensional*



**Access to dimensional structures or detailed data according to requirements**

**(Ad hoc, on demand) !**

# References

---

- <http://www.billinmon.com/>
- <http://www.ralphkimball.com/>
- <http://www.teradatamagazine.com>
- <http://www.teradata.com>
- Teradata University Network

**Vielen Dank !**