

Heisenberg and the uncertainty laws of BI

Zoltan Vago, Senior DWH Consultant

zoltan.vago@teradata.com

03-June-2015

$$H = \sum_q \cos q \cdot C_q^\dagger C_q$$

$$H = \frac{1}{2} \sum_{\langle ij \rangle} (S_i^+ S_j + h.c.)$$

$$C_i = (-1)^{\sum_{j < m} n_j} S_m$$

x-y-model

The uncertainty principle



$$\Delta x \Delta p \geq \frac{\hbar}{2}$$

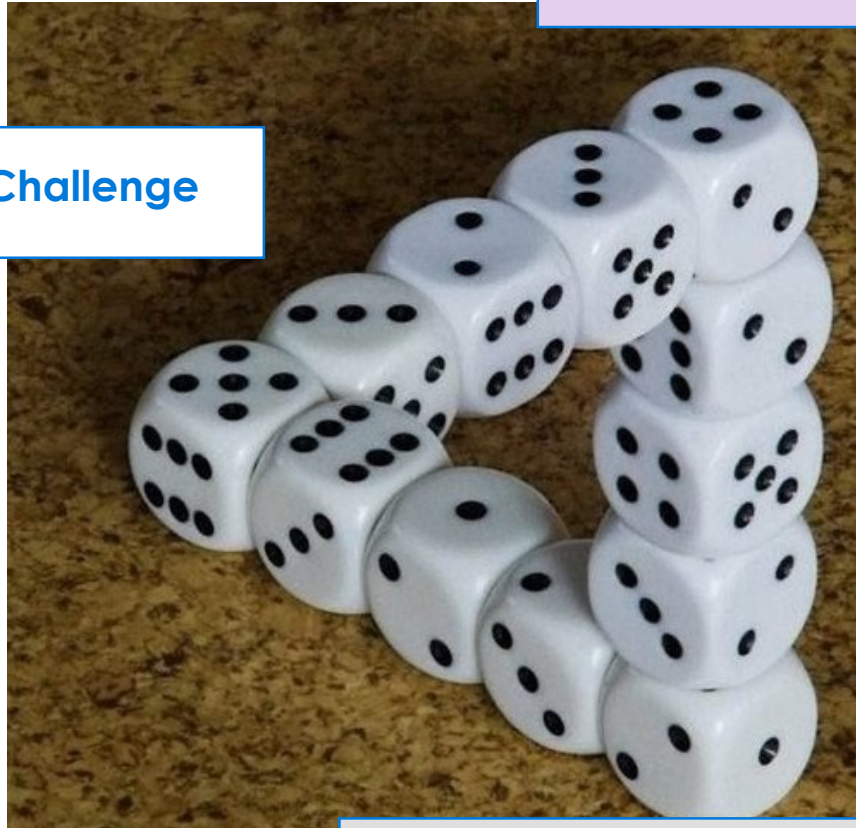
The more precisely the position of some particle is determined, the less precisely its momentum can be known, and vice versa.

Werner Heisenberg, 1927

Three BI conflicts

The IT's BI vs Business' BI Challenge

The Data Modelling Challenge



The Traditional DWH vs Big Data Challenge



Teradata serves
2,600+ customers
in **77** countries



The Data Modelling Challenge

How can we create a data model which is general, flexible AND simple and business user friendly and providing very fast response times at the same time?

The more flexible and general a data model is, the less it supports simple and user friendly querying, and vice versa.

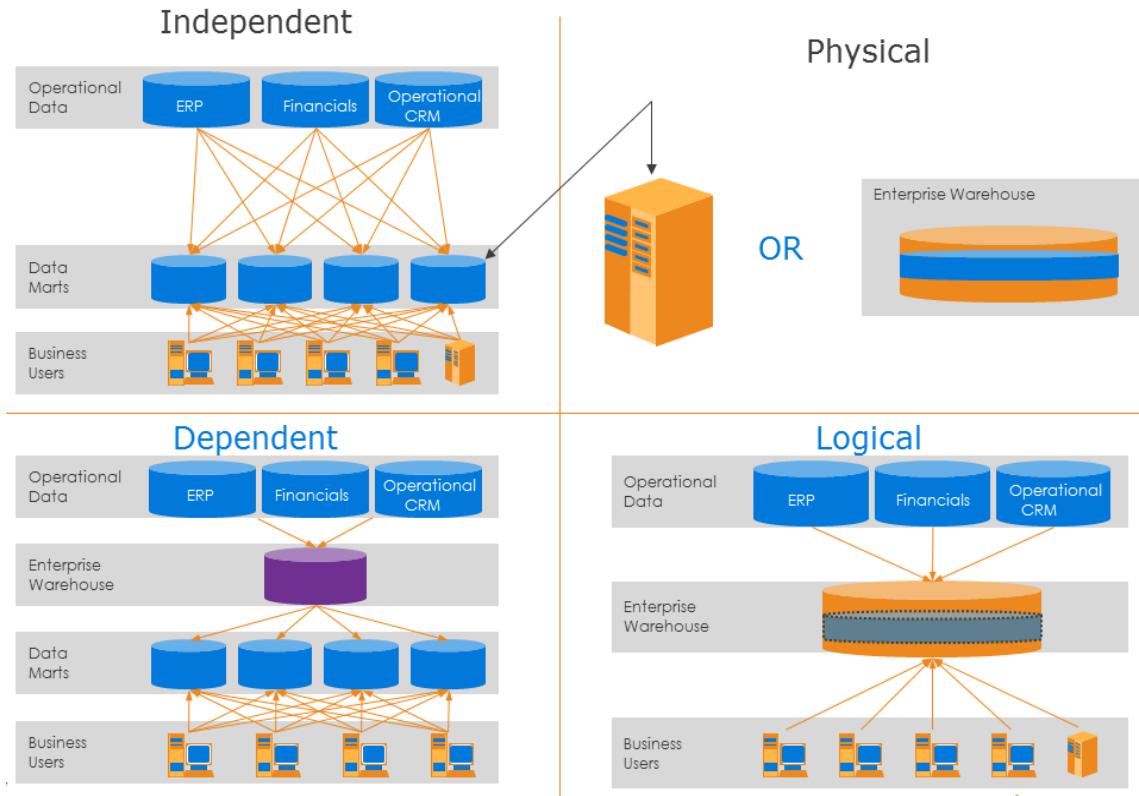
The Data Modelling Solution

Two data models instead of one

- 3NF
 - „Business in its whole complexity”
- Dimensional
 - „Business performance from different views”

New problems

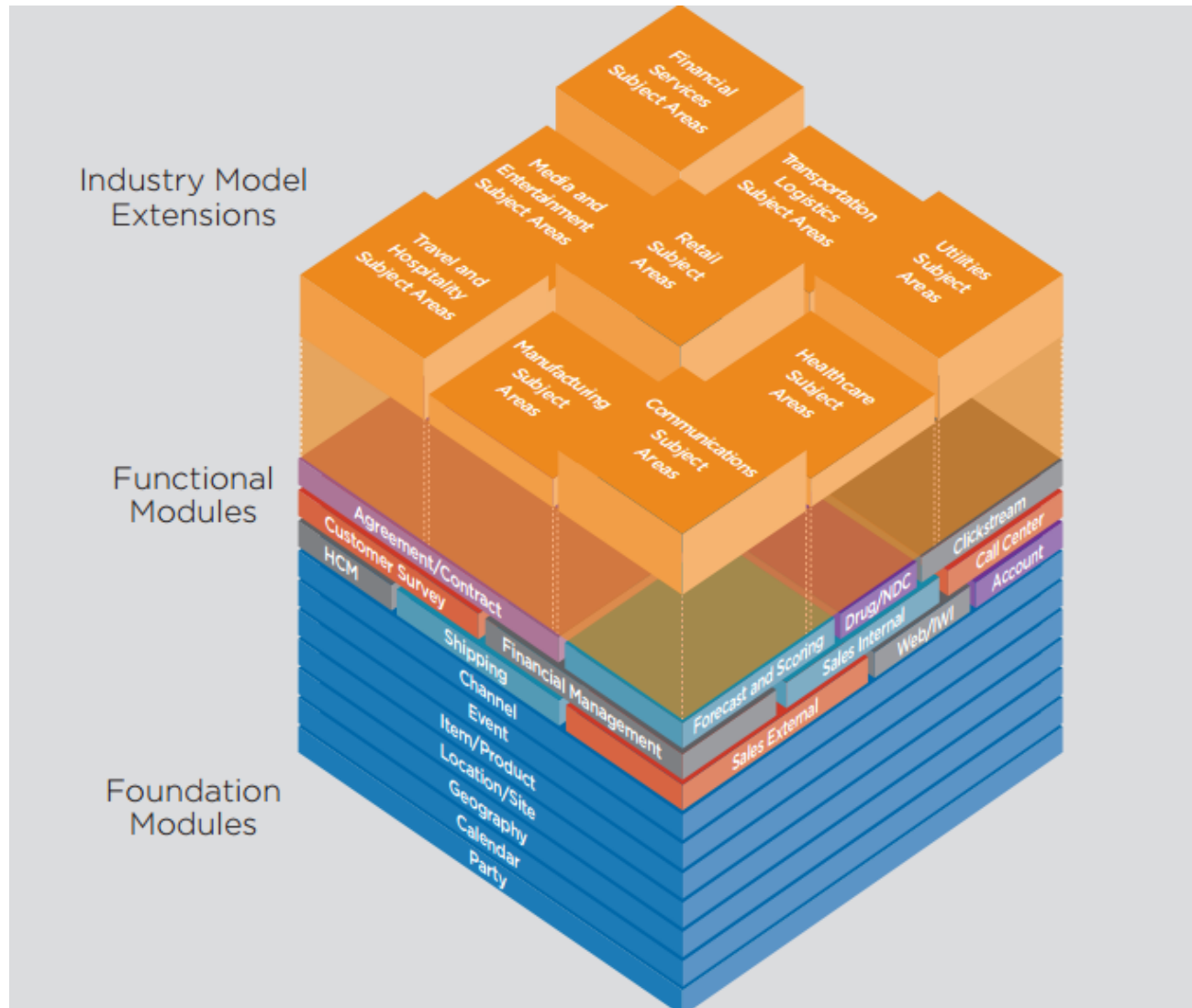
- Doubled ETL
- More space required
- Integrity



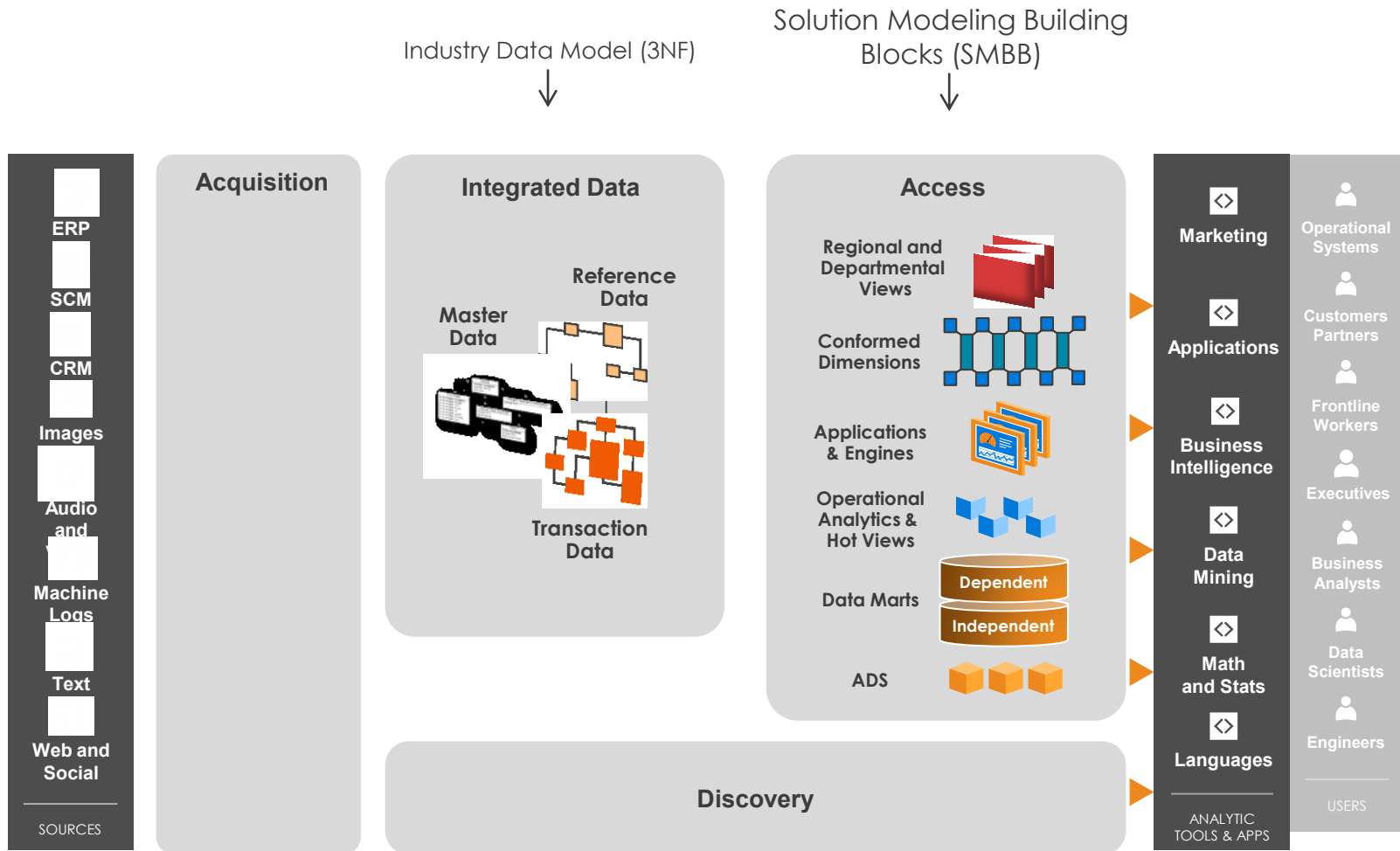
What really makes a data model is not it's type but it's content and quality

Teradata answer: iLDMs and Access Layer Methodology

Teradata Unified Data Model Framework = iDM + Modules + Features from other iDMs



Reference Information Architecture



Access Layer Approach

Teradata Access Layer Practices

Processes

Checklists

Design Patterns

Plan

Elicit

Scope

Design

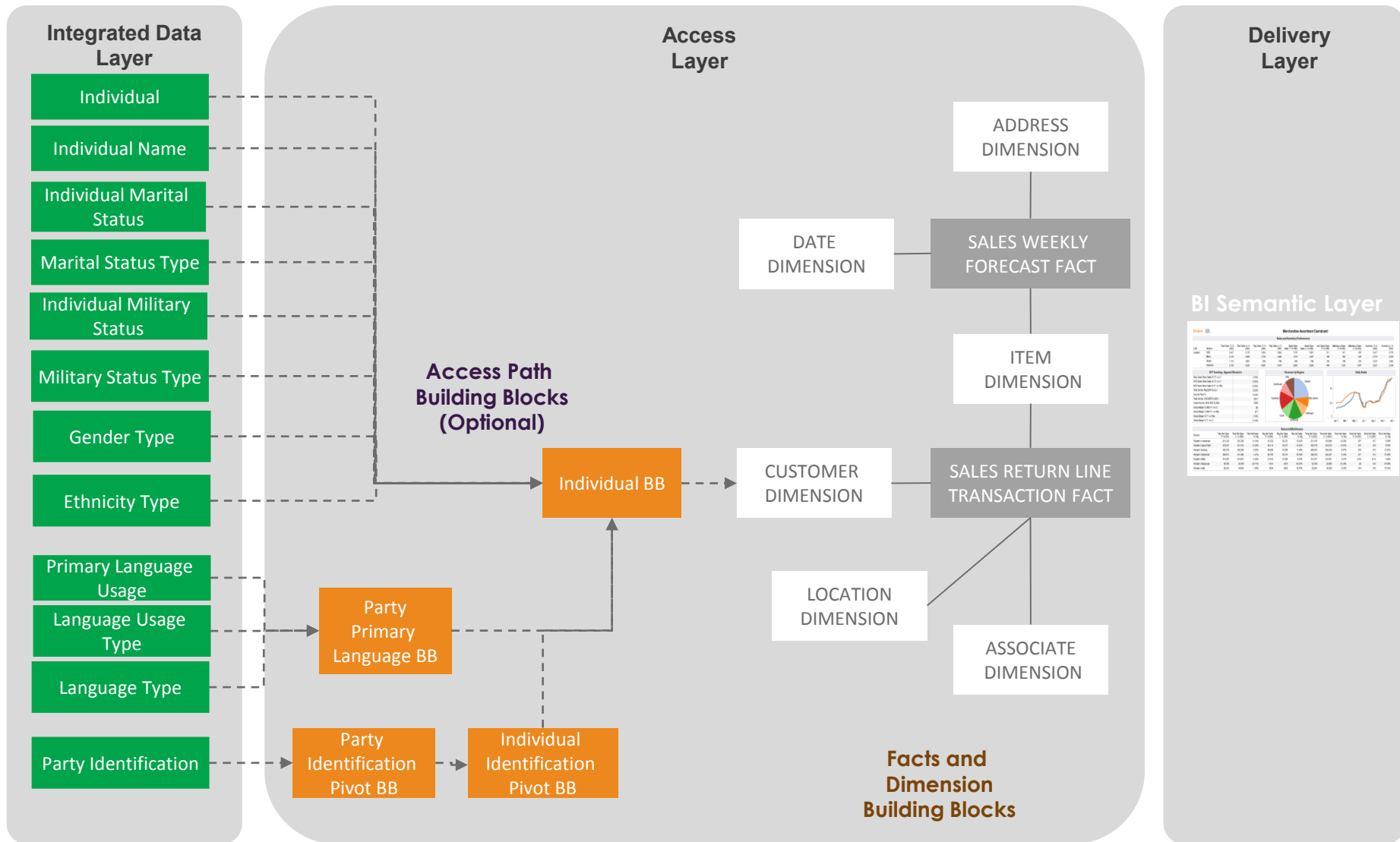
Implement

Validate

Teradata Predefined Semantic Data Models

Solution Modeling Building Blocks

Access Layer Design Example



BI Development Ownership Challenge

How can we solve high quality, fail safe and easy to operate BI development which is agile, follows business changes instantly and supports fail fast business experiments at the same time?

The higher operative quality a single IT environment has, the less it supports ad-hoc developments, and vice versa.

BI Development Ownership Solutions

Political war?



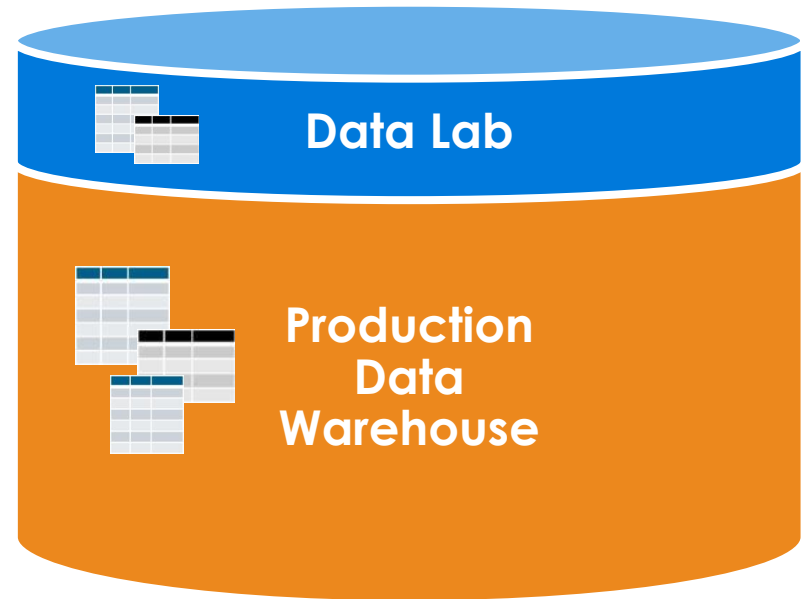
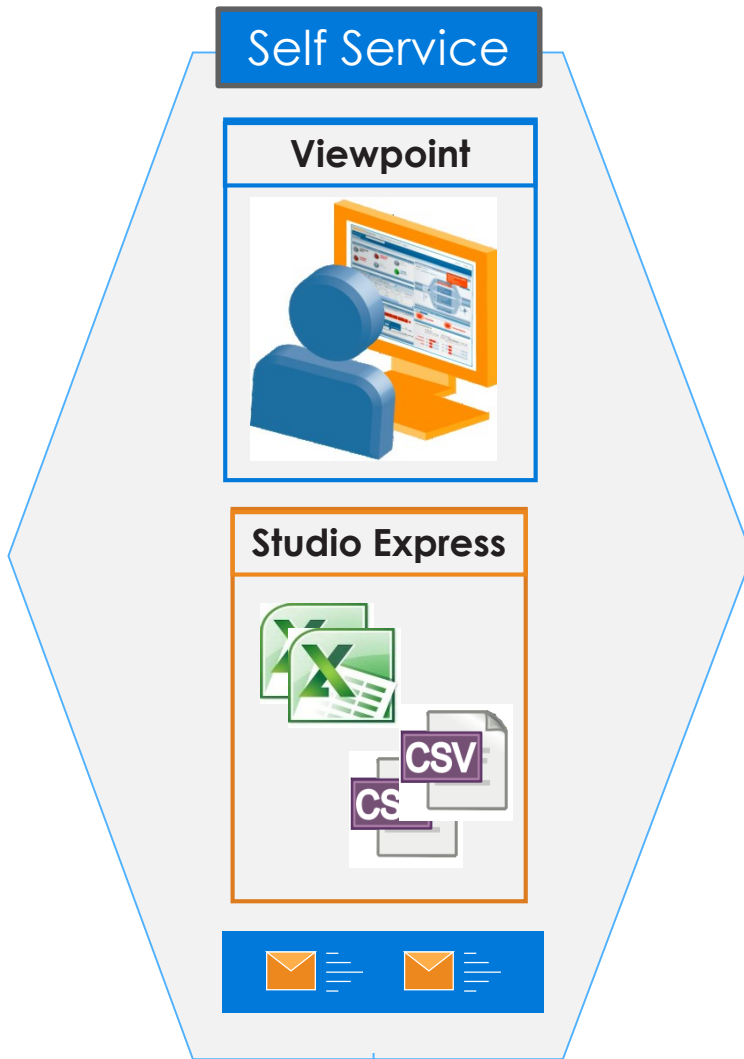
BI Governance

Sandboxing



Teradata Data Lab

Production Analytic Sandboxes



- ✓ Load experimental, untested data from external sources
- ✓ Rapid prototyping, exploratory and experimentation analysis
- ✓ Easily join to production data

The Data Platform Challenge

How can we create a data processing environment which serves big data and traditional BI requirements at the same time?

The more general a data platform is (in terms of processing profiles e.g. OLTP/BI/Image processing /Transaction streaming/etc.) the less it can compete with profile specific platforms, and vice versa.

The Data Platform Challenge – Best Practices

Do nothing

- Keep your existing analytic environment as long as business requirements don't force you to change

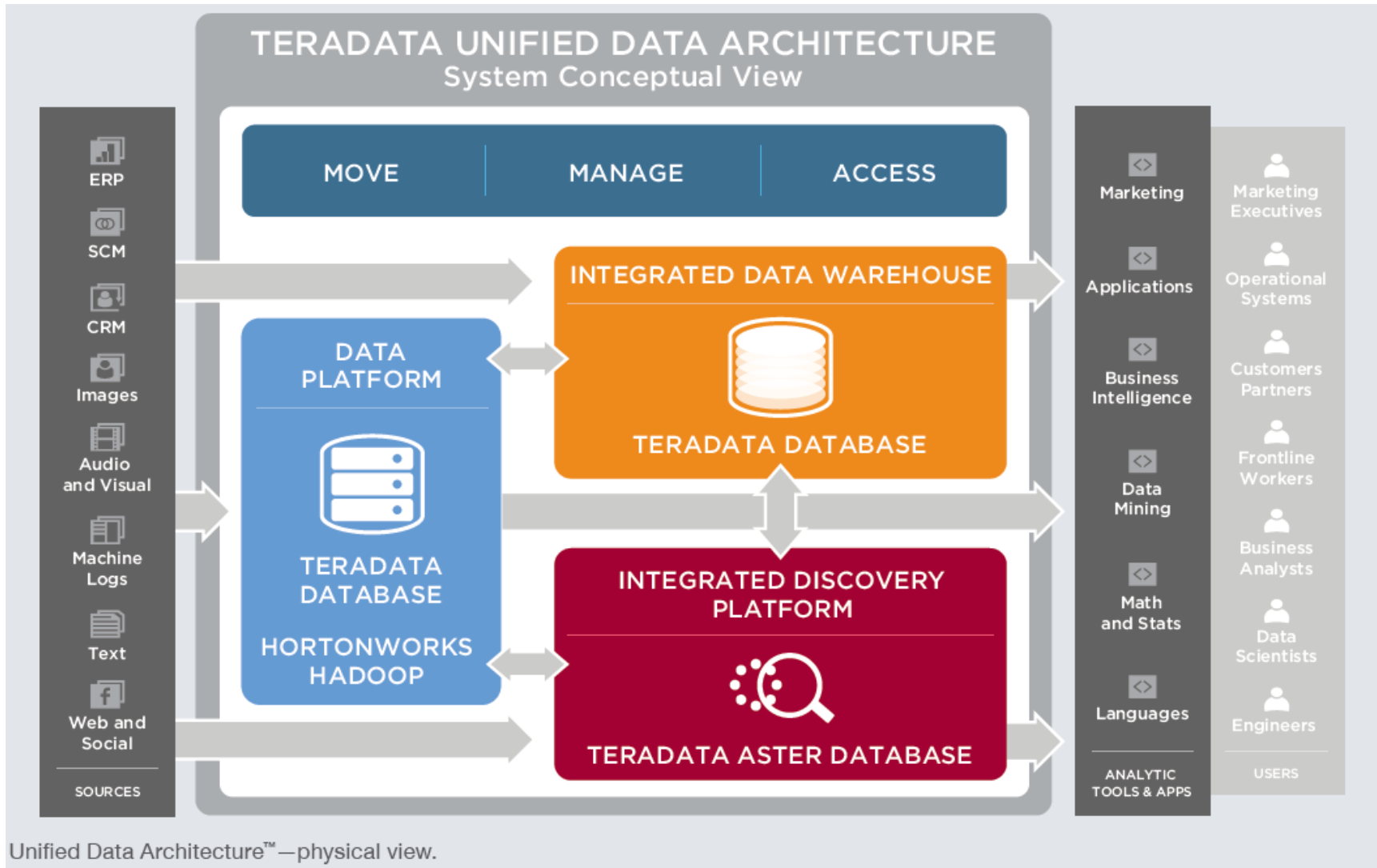
Set up workload specific environments

- ETL, EDW, Data Marts, Data Mining, Real-Time Campaign Management, Big Data, Ad-hoc Analytics (Sandbox), Storage (Data Lake)
- Large number of interfaces, data movement requirements, governance and data quality problems

Theory: „Logical Data Warehouse”

- Gartner's term
- There is an umbrella covering all data management functions
- Different workload specific platforms are hidden and integrated into an orchestrated ecosystem

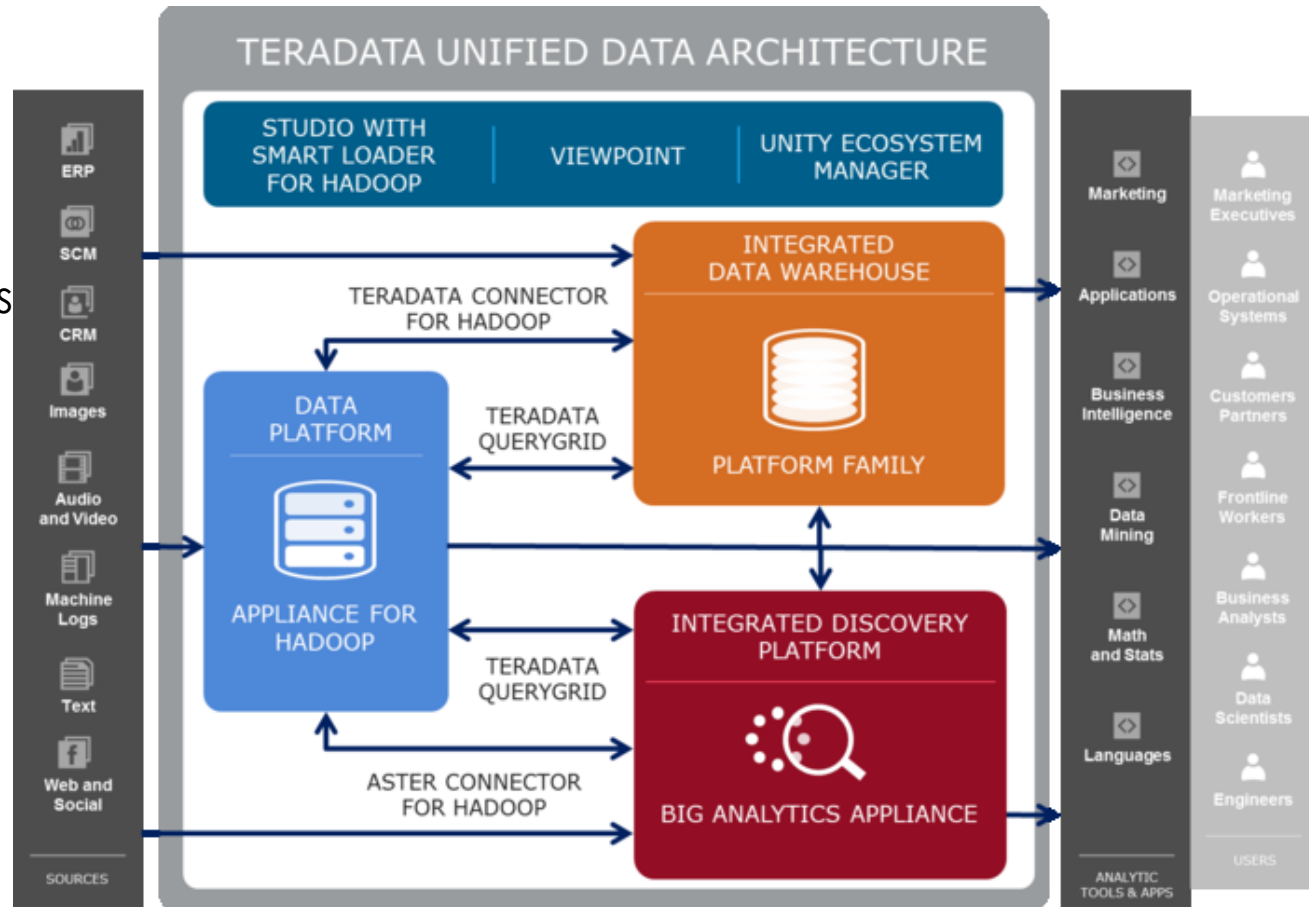
Teradata answer: Unified Data Architecture (UDA)



The Teradata Unified Data Architecture

Simplify the complexity

- Teradata QueryGrid
 - Query execution automation and flexibility for users
- Hadoop Connectors
 - Data movement and access within the Teradata UDA
- Teradata Unity
 - Seamless environment management for administrators
- Teradata Viewpoint
 - GUI-based administrator tool

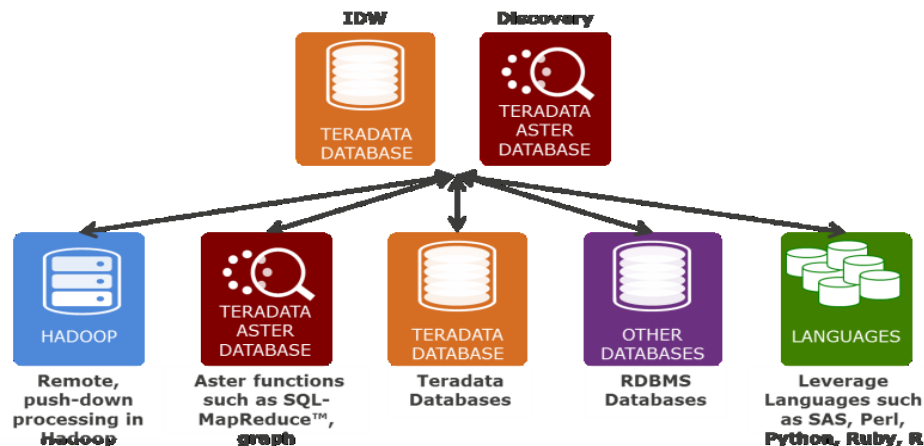


Teradata QueryGrid

Overview

Teradata solution to querying and analyzing data across the UDA

- Suite of connectors for query execution on a heterogeneous environment
- Queries can be run both within the Teradata ecosystem (UDA) or outside the Teradata ecosystem components
- The processing of the query takes place in the remote system – Push-Down processing – avoiding the need for data replication and movement, and generating results with lower data lags



Wrap-up

The IT's BI vs Business' BI Challenge

The Data Modelling Challenge

**iLDMs, SMBB,
and
Access Layer
Methodology**



Teradata Data Lab

The Traditional DWH vs Big Data Challenge

Teradata Unified Data Architecture

THANK YOU!

Teradata – more information

- **Teradata Hungary contacts**
 - István Magyar (Istvan.Magyar@Teradata.com) – general sales
 - Angela Kertész (Angela.Kertesz@Teradata.com) – product marketing
- **Sites, resources**
 - <http://www.teradata.hu>
 - <http://www.teradata.com>
 - <http://developer.teradata.com>
- **Programs**
 - **Teradata Connect 2015**
 - > June 9-10, 2015 – London
 - **Teradata CTO Roadshow 2015**
 - > May 27, 2015 - Warsaw
 - > June 2, 2015 – Prague
 - **Teradata PARTNERS Conference & Expo 2015**
 - > October 18-22 , 2015 – Anaheim, California

