

# IBM SOCIAL AND ANALYTICS CONFERENCE 2017

Redefine work with Watson

## Who needs data warehouse models?

Inmon vs. Kimball .. or hybrid and Why we need DWH models

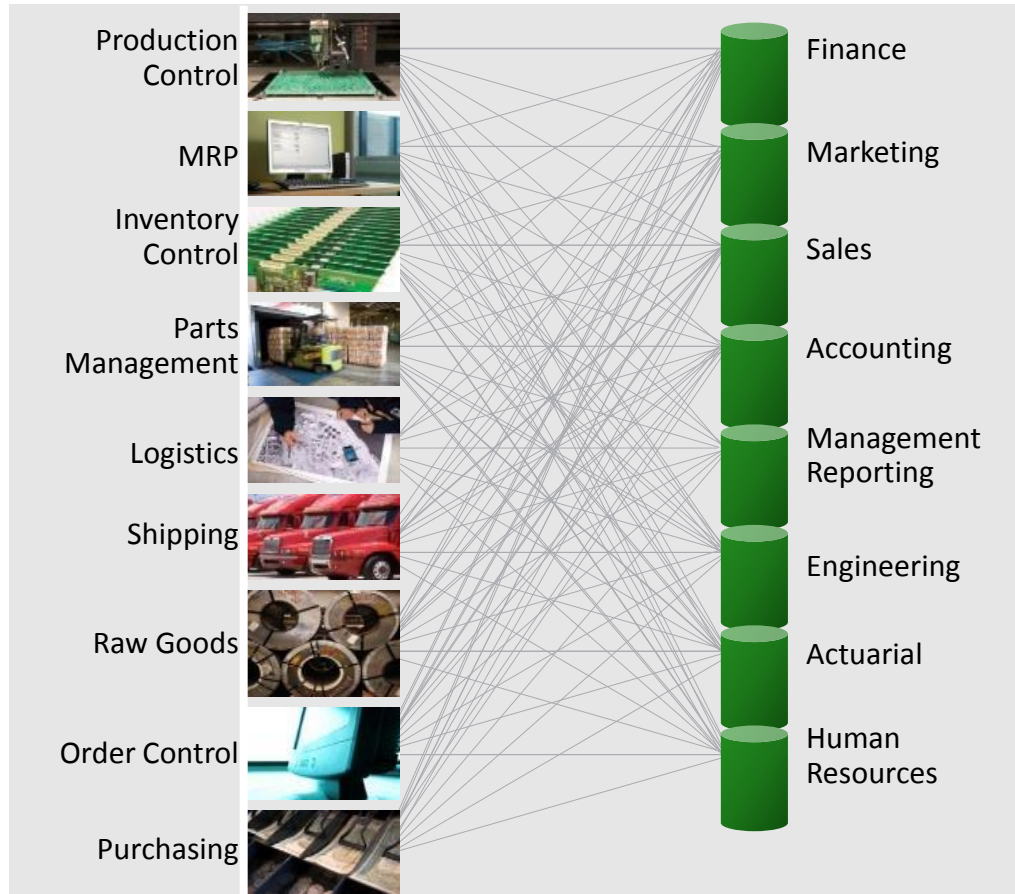




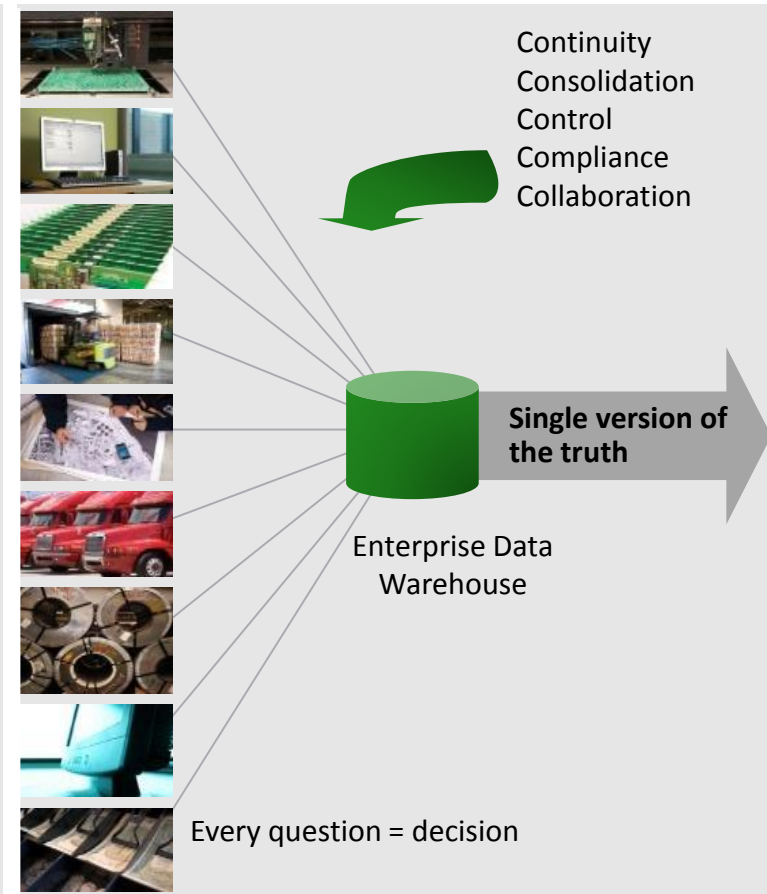


# Why use a Data Warehouse?

Applications + databases = chaos



Data warehouse = order



Two purposes of data warehouse: 1) save time building reports; 2) slice in dice in ways you could not do before





# Data Warehouse vs Data Mart

- Data Warehouse: A single organizational repository of enterprise wide data across many or all subject areas
  - Holds multiple subject areas
  - Holds very detailed information
  - Works to integrate all data sources
  - Feeds data mart
- Data Mart: Subset of the data warehouse that is usually oriented to specific subject (finance, marketing, sales)
  - The logical combination of all the data marts is a data warehouse

In short, a data warehouse as contains many subject areas, and a data mart contains just one of those subject areas





# Kimball and Inmon Methodologies

Two approaches for building data warehouses





# Kimball vs. Inmon

## Inmon

- Relational
  - Entity-Relationship (ER) model
  - Normalization rules
  - Many tables using joins
  - History tables, natural keys
  - Good for indirect end-user access of data

## Kimball

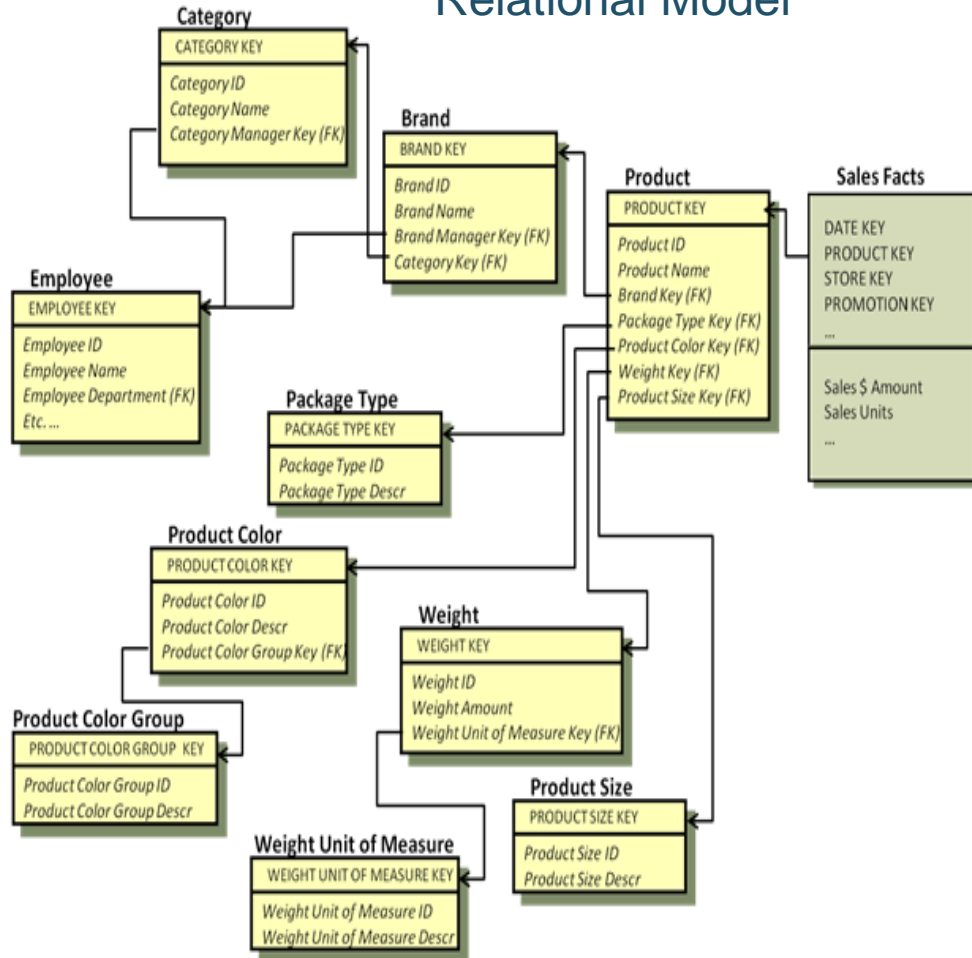
- Dimensional
  - Facts and dimensions, star schema
  - Less tables but have duplicate data (de-normalized)
  - Easier for user to understand (but strange for IT people used to relational)



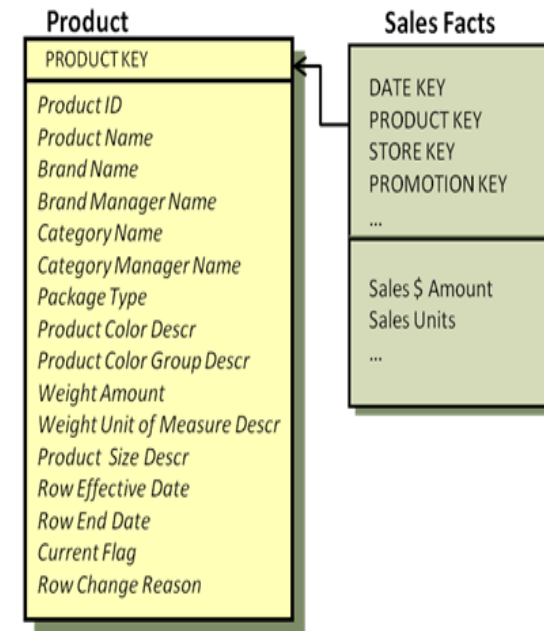


# Relational Model vs Dimensional Model

Relational Model



Dimensional Model



If you are a business user, which model is easier to use?





# Kimball vs. Inmon

## Inmon

- Enterprise data model (CIF) that is a enterprise data warehouse (EDW)
- IT Driven, users have passive participation
- Centralized atomic normalized tables (off limit to end users)
- Later create dependent data marts that are separate physical subsets of data and can be used for multiple purposes
- Integration via enterprise data model
- 3-tier (data warehouse, data mart, cube), duplication of data

## Kimball

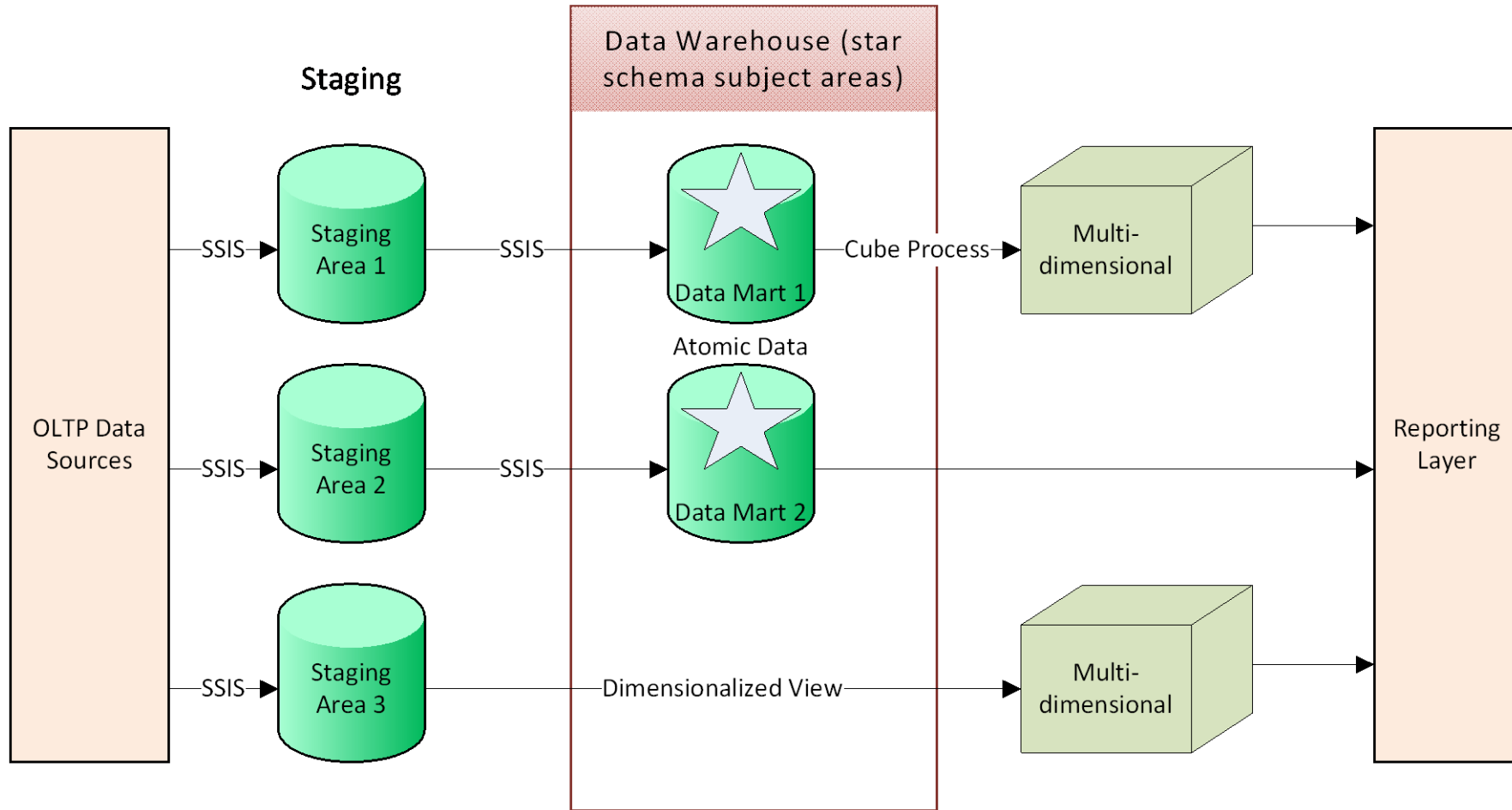
- Logical data warehouse (BUS), made up of subject areas (data marts)
- Business driven, users have active participation
- Decentralized data marts (not required to be a separate physical data store)
- Independent dimensional data marts optimized for reporting/analytics
- Integrated via Conformed Dimensions (provides consistency across data sources)
- 2-tier (data mart, cube), less ETL, no data duplication





# Kimball Model

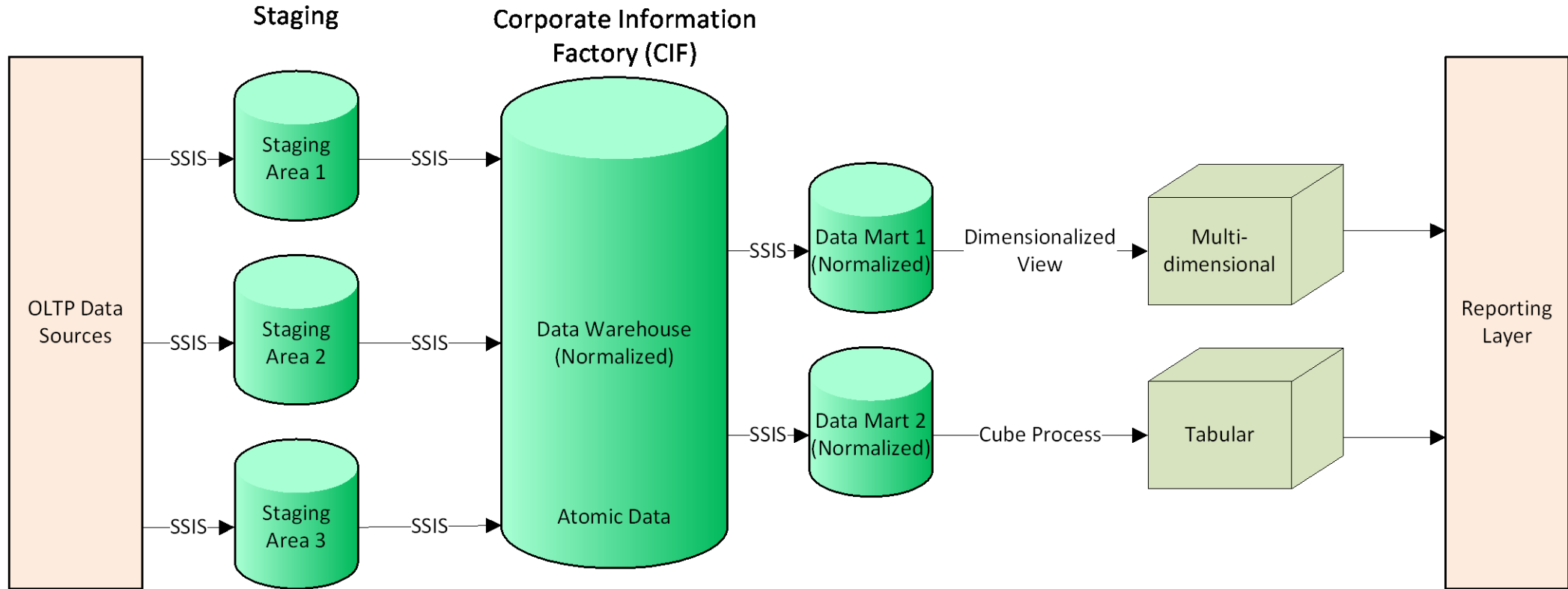
## DW Bus Architecture







# InmonModel







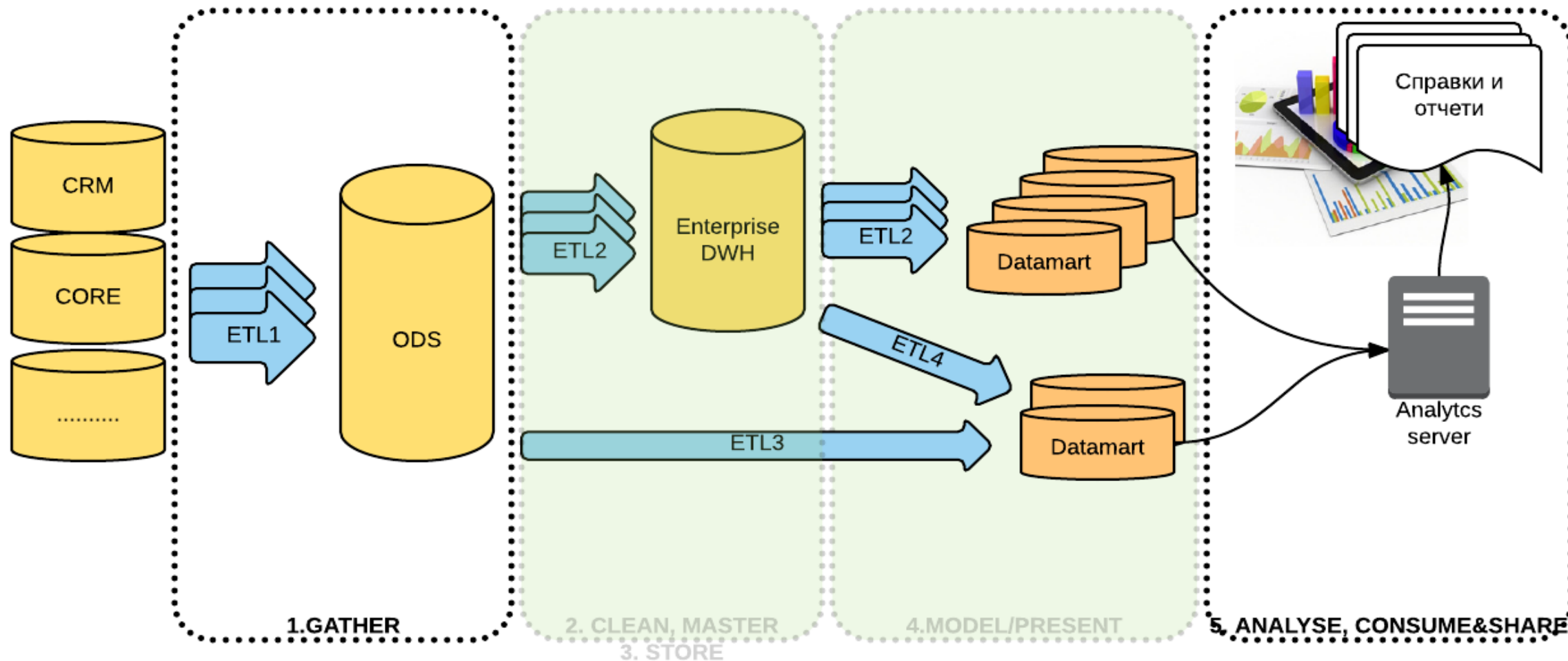
# Which model to use?

- Models are not that different, having become similar over the years, and can compliment each other
- Boils down to Inmon creates a normalized DW before creating a dimensional data mart and Kimball skips the normalized DW
- With tweaks to each model, they look very similar (adding a normalized EDW to Kimball, dimensionally structured data marts to Inmon)
- Bottom line: Understand both approaches and pick parts from both for your situation – no need to just choose just one approach





# Hybrid model







# Why we need DWH Data Models?



# Why do we need a model?

**without Model**



**with Model**





# What is the Purpose of a Data Model?

A visual business representation of how data is organized in the enterprise

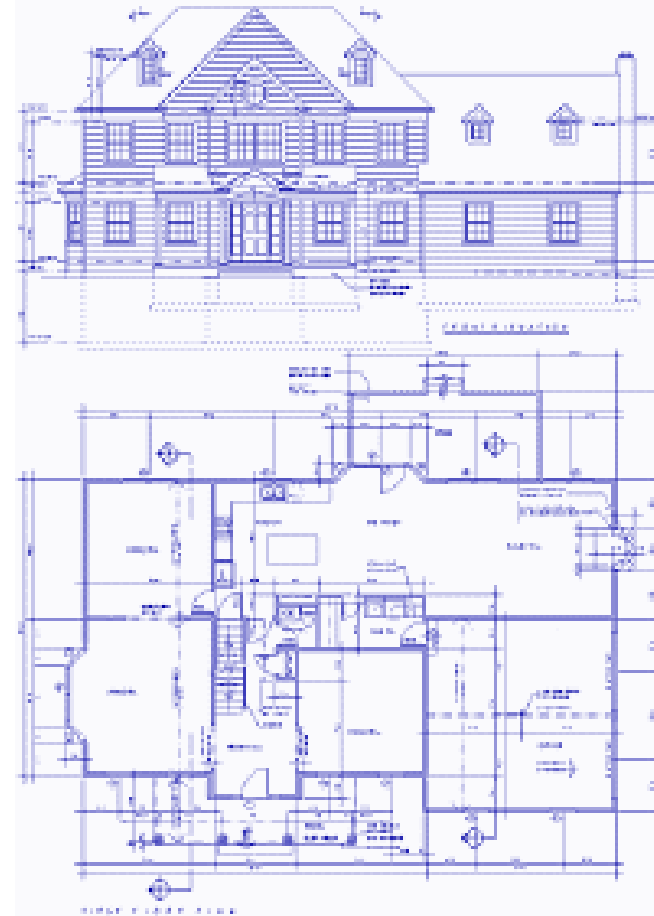
It provides discipline and structure to the complexities inherent in data management

Can you imagine building a house without a blueprint?

Or driving across the country without a map / navigation?

It facilitates communication within the business (e.g. within IT and between IT and the business)

It facilitates arriving at a common understanding of important business definitions (e.g. what is a customer?)



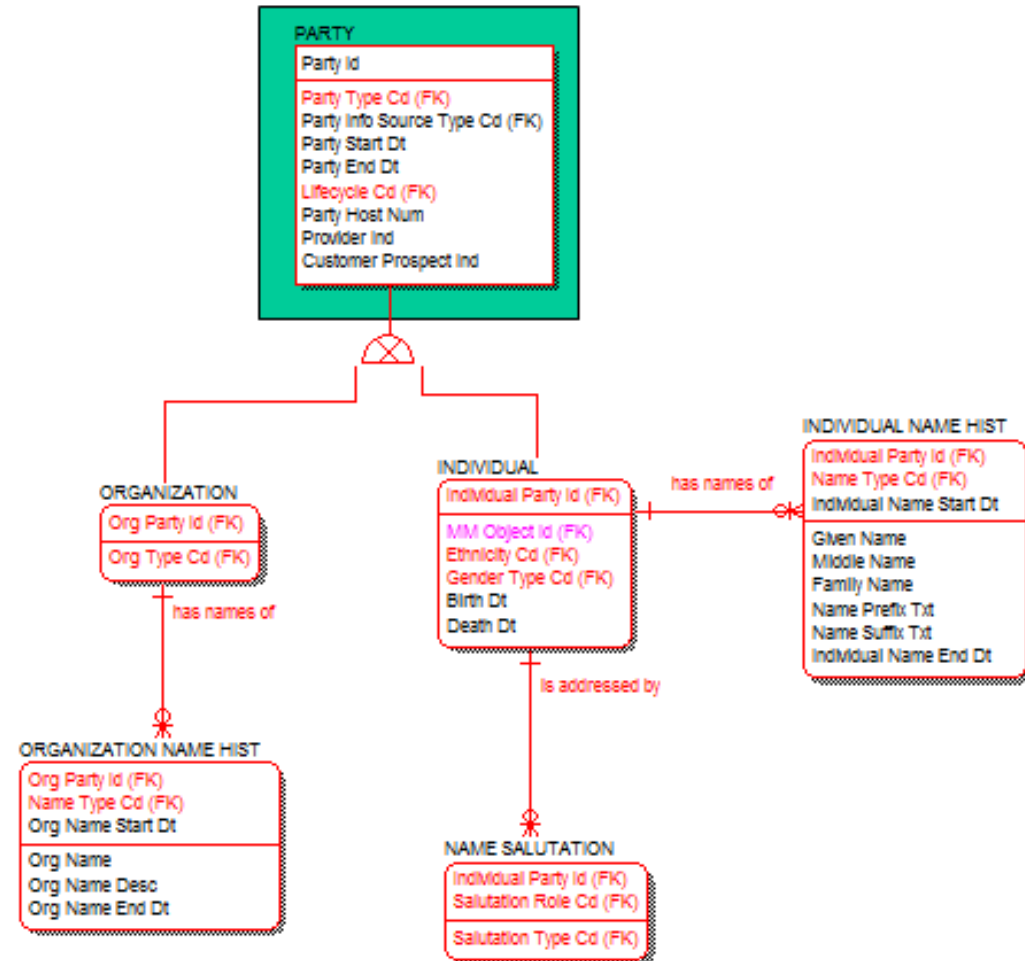




# What does a Logical Data Model do?

LDM Represents data requirements in a correct, sharable, stable and flexible format.

- **Stable** - Can service many application and business requirements over the long term
- **Correct** - Maps the data one-to-one to the business as it is in the real world
- **Sharable** - No bias toward a particular application or business requirement
- **Flexible** - Changes little or not at all as the business environment changes







# What is Physical model

- The Physical Data Model is the physical instantiation of the LDM, it is developed from the Logical Data Model, with changes to reflect restrictions imposed by the data that is actually available from the existing computer systems, from the data discovery and to improve the performance for the expected query paths.
- Tables may be combined and new relationships may be introduced to improve and simplify understanding, use and improve performance.
- During the data discovery and physical design phase new attributes and entities may be discovered, these should be validated with users and the Logical Data Model amended to include them where appropriate





# *Logical to physical*

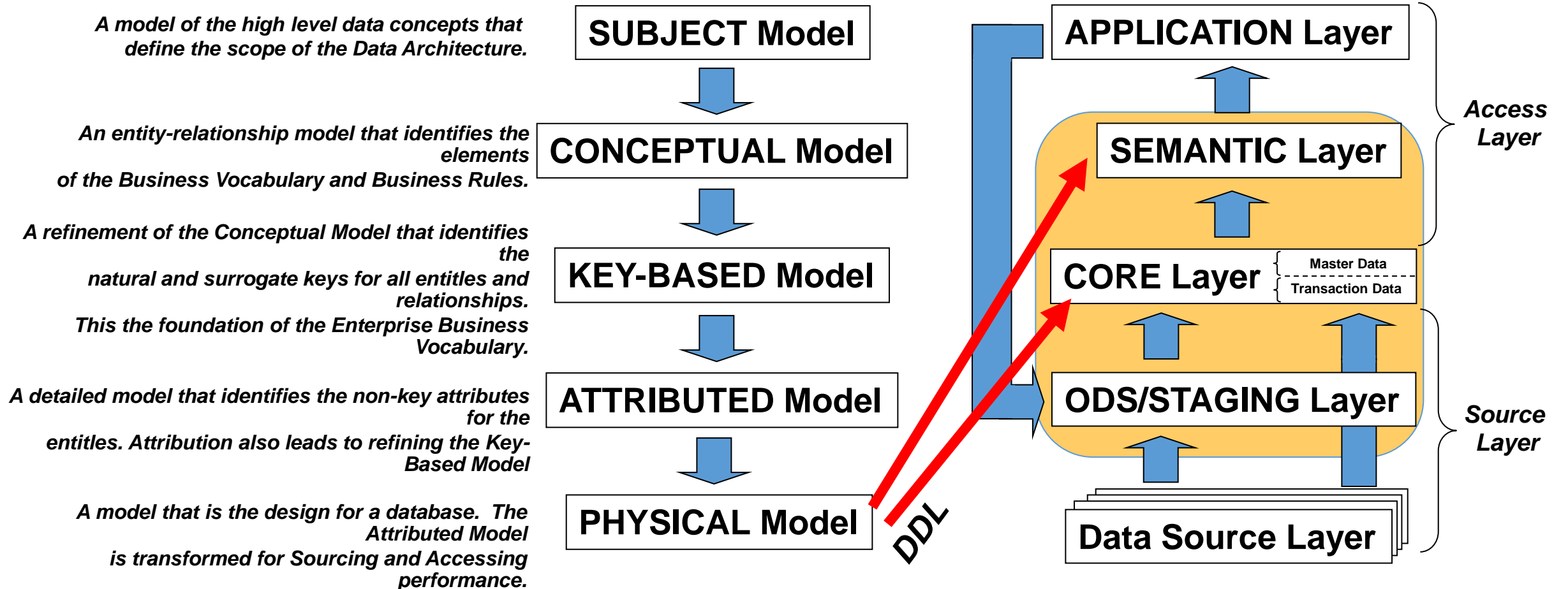
| Logical      | Physical         |
|--------------|------------------|
| Entity       | Table            |
| Attribute    | Column           |
| Relationship | PK/FK Constraint |
| Key Group    | Index            |



# Data Modeling Process

## Data Modeling

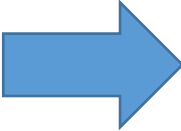
## Data Warehouse implementation







# Why we need DWH modeling?

- We need Data Modeling for DWH to speed up development and minimize costs and risks
  - Data model elements – entity, attribute, relationship, key
  - Physical Data Model is real-life implementation of logical data model
  - DWH Data modeling and implementation process should be streamlined and packaged industry data model should be attributed and prepared as much as possible for physical modeling to speed up the implementation
- 
- Not only pure LDM layer, but also attributes within entities and aggregation / data mart layer
  - All elements (subject areas, entities, attributes) documented in detail within the model
  - Clear visual representation and color-coding of model objects and Subject areas
  - Detailed model documentation including implementation guides



# Industry Standard Banking Data Warehouse Data Model

Poslovna Inteligencija





# Model benefits

- Easy to customize according to bank's needs
- Implementation can be done incrementally delivering fully functional phases
- Enables business users to more effectively control and reduce the time taken to scope their requirements, as well as subsequent customization and any extension of the data warehouse
- Provides a solid basis for regulatory reporting as well as decision support and executive information applications
- Minimizes development costs
- Reduces the risk of failure by facilitating an incremental approach to delivering integrated data warehouse solution





# Based on practical experience

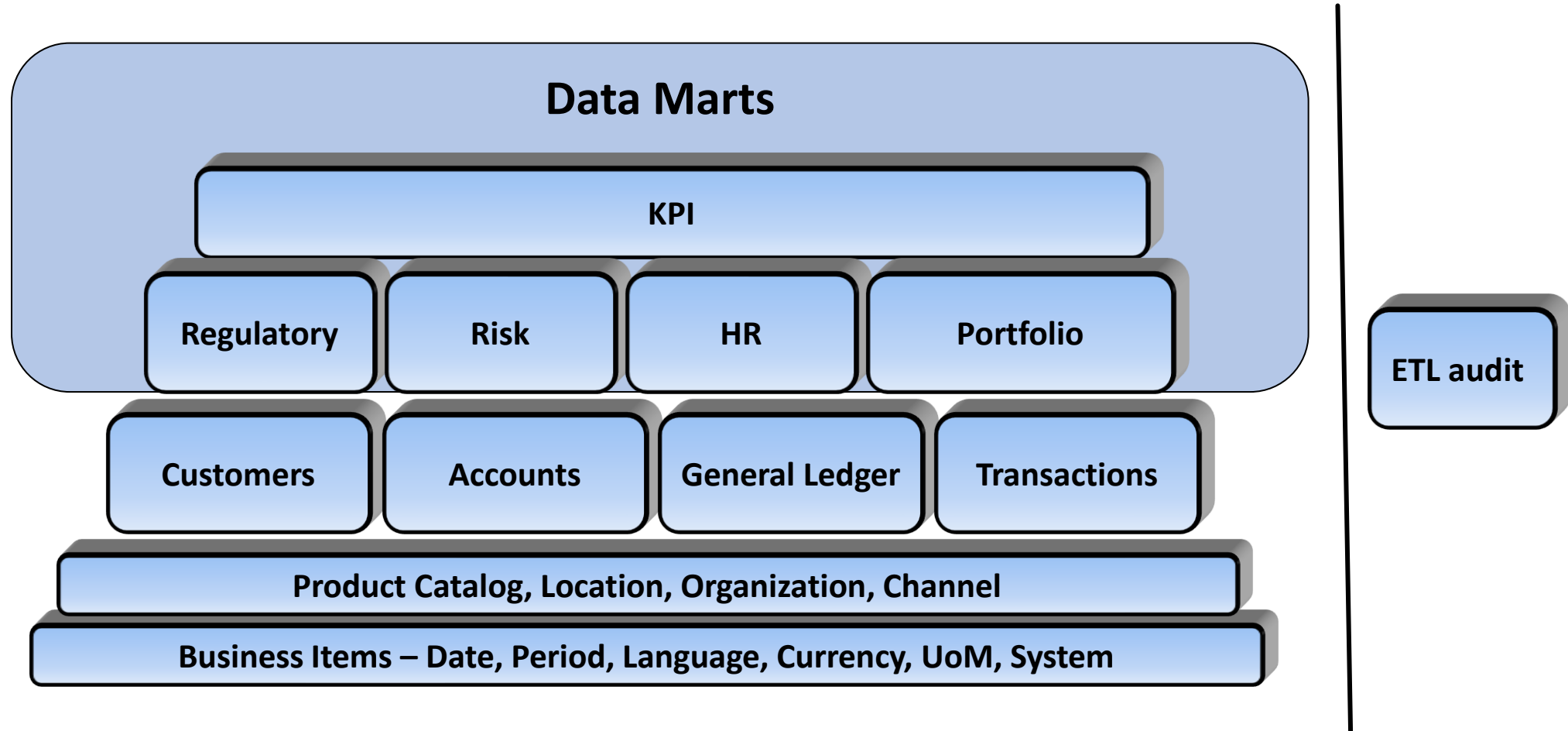
- developed since 2010, on the basis of the experience in implementation in different financial institutions
- Open to customizations
- Possibility to integrate data between bank and leasing company or bank and insurance company in one common data warehouse model
- Work on the model is a continuous endeavor, in respect to:
  - Banking regulations
  - New Analytical requirements
  - New markets
  - Data modelling standards







# Business Areas – logical view







# Different Data Marts ...

- Reporting and regulatory compliance
- Portfolio:
  - Deposit Portfolio
  - Credit Portfolio
  - NPL
- Customer – different perspectives
- Profitability
- ...

|   |  |
|---|--|
| <b>REPORTING AND REGULATORY COMPLIANCE</b><br><br>General Ledger Balance<br>General Ledger Average Balance<br>General Ledger Summary<br>Kpi Summarized Level<br>Kpi Detailed Level<br>Report Position Values                                    | <b>PORTFOLIO</b><br><br>Credit Portfolio<br>Deposit Portfolio<br>Customer Portfolio<br>Credit Risk Portfolio |
| <b>PROFITABILITY</b><br><br>Customer Profitability<br>Product Profitability<br>Organization Unit / Cost Centre Profitability<br>Location Profitability<br>Channel Profitability<br>Profitability Analysis<br>Transaction Profitability Analysis | <b>360 DEGREE VIEW OF CUSTOMER</b><br><br>Customer Summary<br>Customer Behavior Analysis<br>Customer Loyalty |





# KPI's, reports, dashboard tech perspective

- 100+ Standard Banking KPI's and calculations are defined from model tables
- Standard analytical models and set of 40+ reports and 5+ dashboards
- More than 380 entities (tables) in core DWH model
- Possibility to create number of datamarts based on the currently existing entities
- Created in Erwin model. Can be exported to any format.
- Descriptions of attributes / entities, with examples (possible values), where applicable
- Database independent - possibility to create DB schema on any of the standard RDBMS



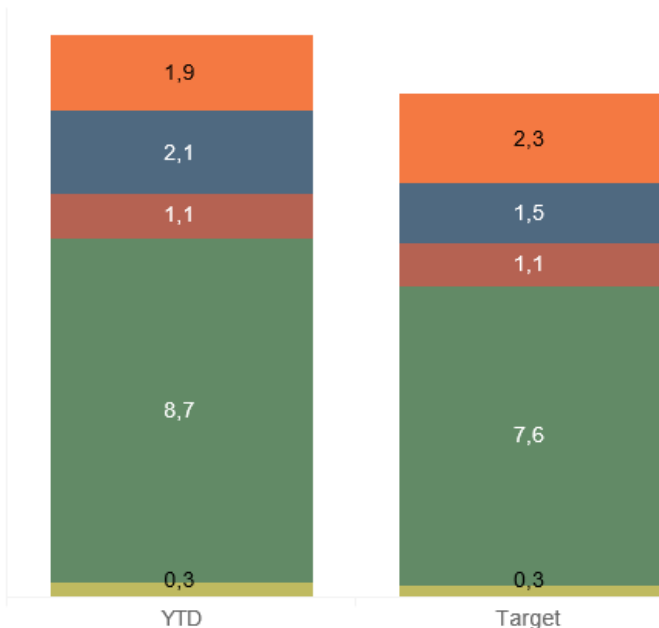
# IBS Dashboard samples - finance

## KEY FINANCIALS

poslovna  
inteligencija  
Poslovna means Business

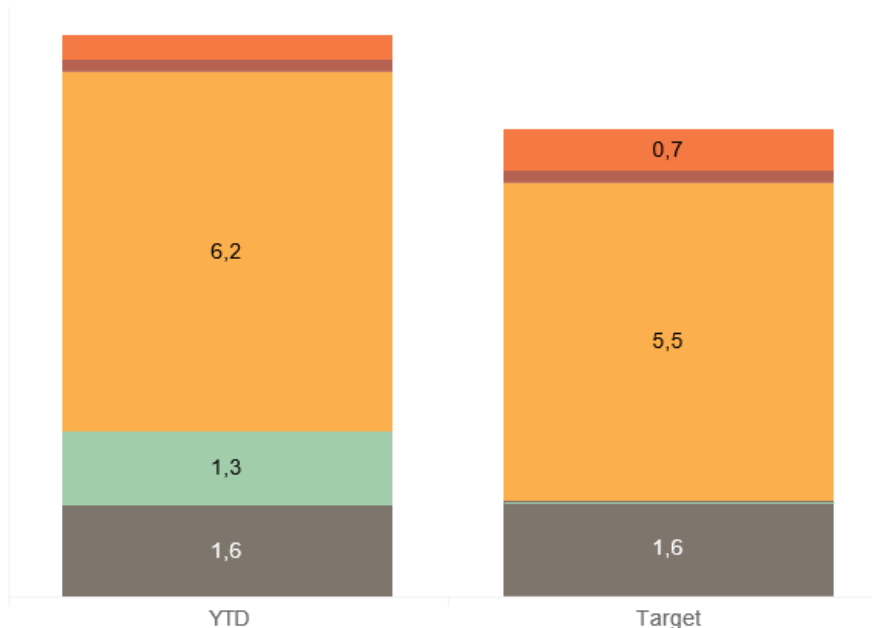
### Balance sheet

Assets (incl. Risk Provisions)



Financial Assets  
Liquidity  
NIBL  
Living Loans  
Other assets

Liabilities



Interbanking  
Other liabilities  
Primary funds  
Refi-lines  
Total equity

|                            |   |        |
|----------------------------|---|--------|
| Interest rate living loans | ↓ | -0,05% |
| Interest rate deposits     | ↓ | -0,01% |
| Net interest margin        | ↑ | 0,08%  |
| Cost / income ratio        | ↑ | 3,89%  |
| NPL Coverageratio          | ↑ | 0,39%  |
| Return on equity           | ↓ | -2,23% |

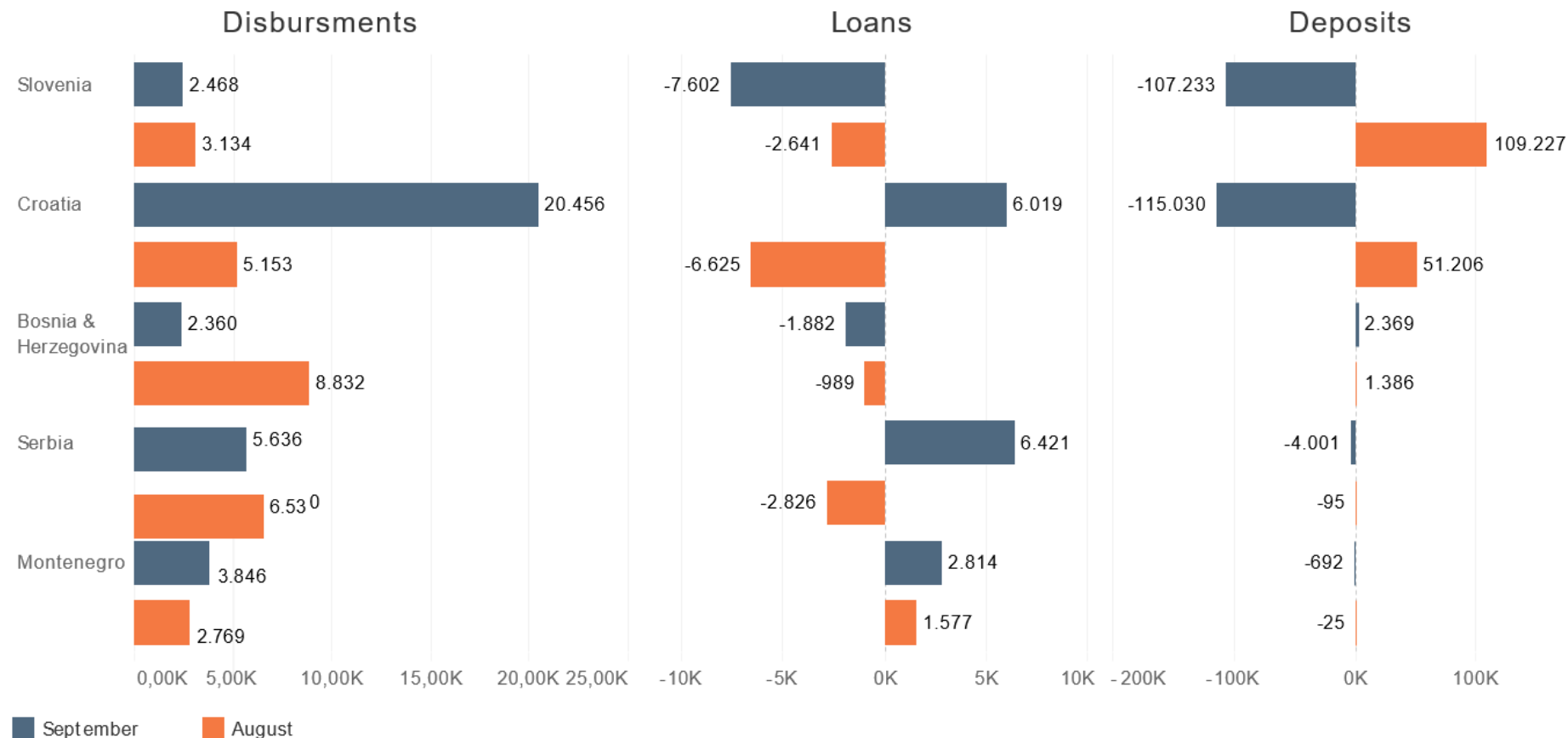


# IBS Dashboard samples - sales

## SALES – Market development

poslovna  
inteligencija  
Poslovna means Business

Business Segment  
Large Business



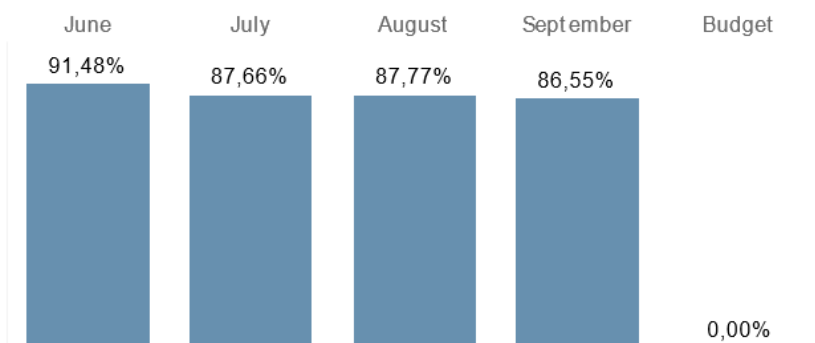


# IBS Dashboard samples - risk

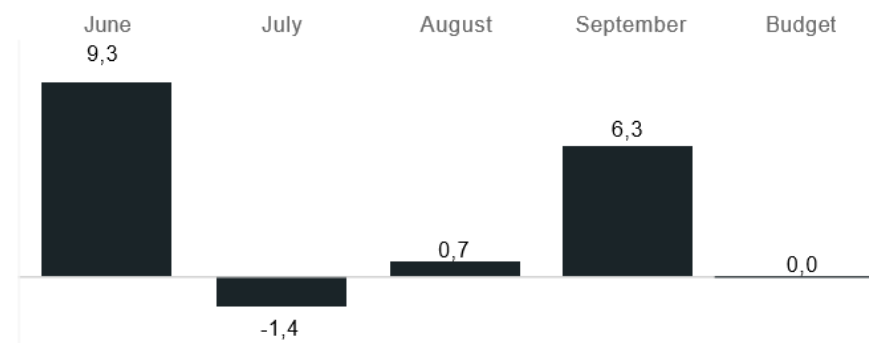
## RISK – Overview

poslovna  
**inteligencija**  
Poslovna means Business

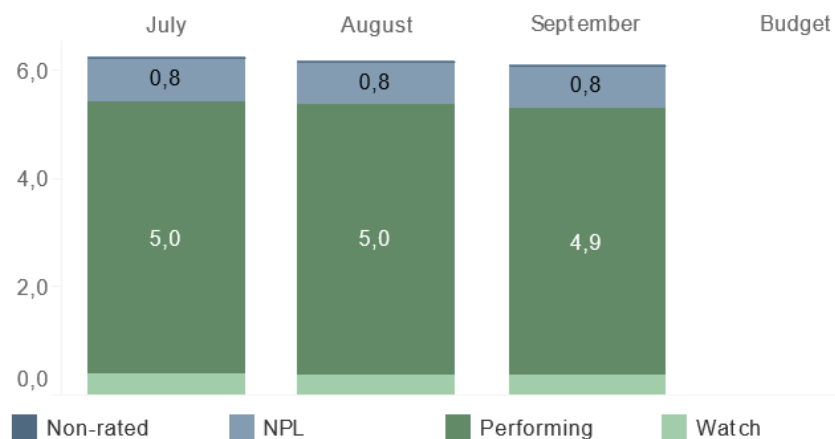
Risk bearing capacity utilization ( % )



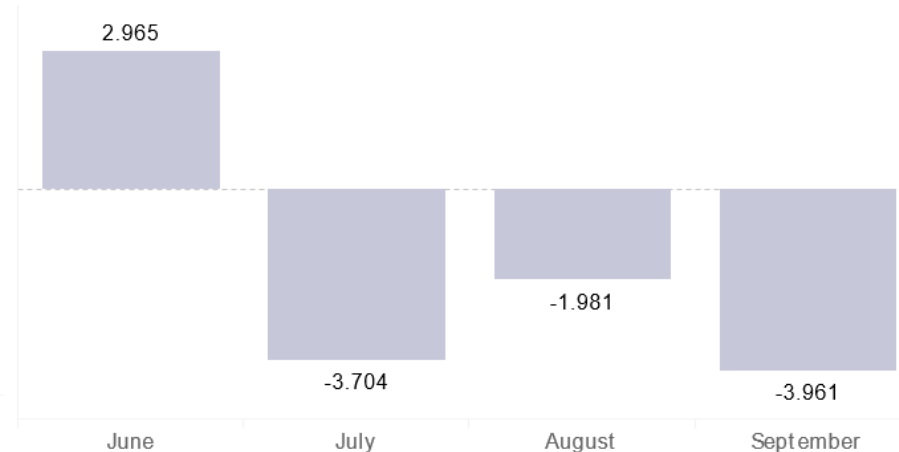
P/L effect of risk provisions monthly (EUR M)



Overview by asset quality



Credit risk shortfall







# Model Roadmap

- Major release (X.0) available every 2 years
  - New Subject Areas and related supported Analysis and KPI's
  - Everything included in Minor releases
  - Next major release planned in Q4 2017
- Minor release (2.X) available 3 times a year
  - New Tables in existing Subject Areas
  - New attributes in Existing tables
  - Other minor enhancements
  - Next minor release planned in Q1 2017





# Model packaging

- Model in ERWin format and exported to other required formats
- Detailed Subject Areas Documentation (~50 pages)
- Detailed ERWin Model Report Documentation (~250 pages)
- Business Glossary with 100+ definitions
- KPI list with 100+ standard banking KPI definitions
- Customization Guide Documentation
- Source to Target Mapping templates
- Detailed Model content and Methodology Powerpoint Presentation (300+ Slides)





# Other DWH Models

- Insurance

- Organizes data around a number of key business subjects – Products & KPI's, Finance&HR, Customer, Policy – Actuary, Claims, Sales.
- Supports stable decision-making process
- Includes a number of most commonly used pre-defined summarizations
- Enables consistent and consolidated reports
  - Delivers competitive advantage by enabling the consolidation of clean data across multiple systems
  - Supports rapid implementation of warehousing solutions with meaningful data

- Telcos

- standard industry data warehouse model
- applicable for fixed and mobile telecommunications operators
- traditional Business Intelligence and Big Data Analytics
- follows TM Forum's Information Framework Shared Information Data Model (SID)
- all standard Telecommunication reporting and analysis Data Marts





# erwin data modelling

An Agile Foundation for Building the Data-Driven Enterprise







# About erwin Data Modeling

erwin delivers a unique “Any<sup>2</sup>” approach to connecting enterprise stakeholders to any data, anywhere in support of the data driven enterprise

- The leading data modeling solution , with over 30% market share (IDC)
- Built on the vision and experience of data modelers worldwide
- De-facto standard in data model integration
- Preferred data model format for industry data model providers: ADRM, Teradata, IBM.....

erwin should be evaluated by any company that is about to embark on a major application development initiative or an application portfolio overhaul.”

- Ovum

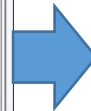
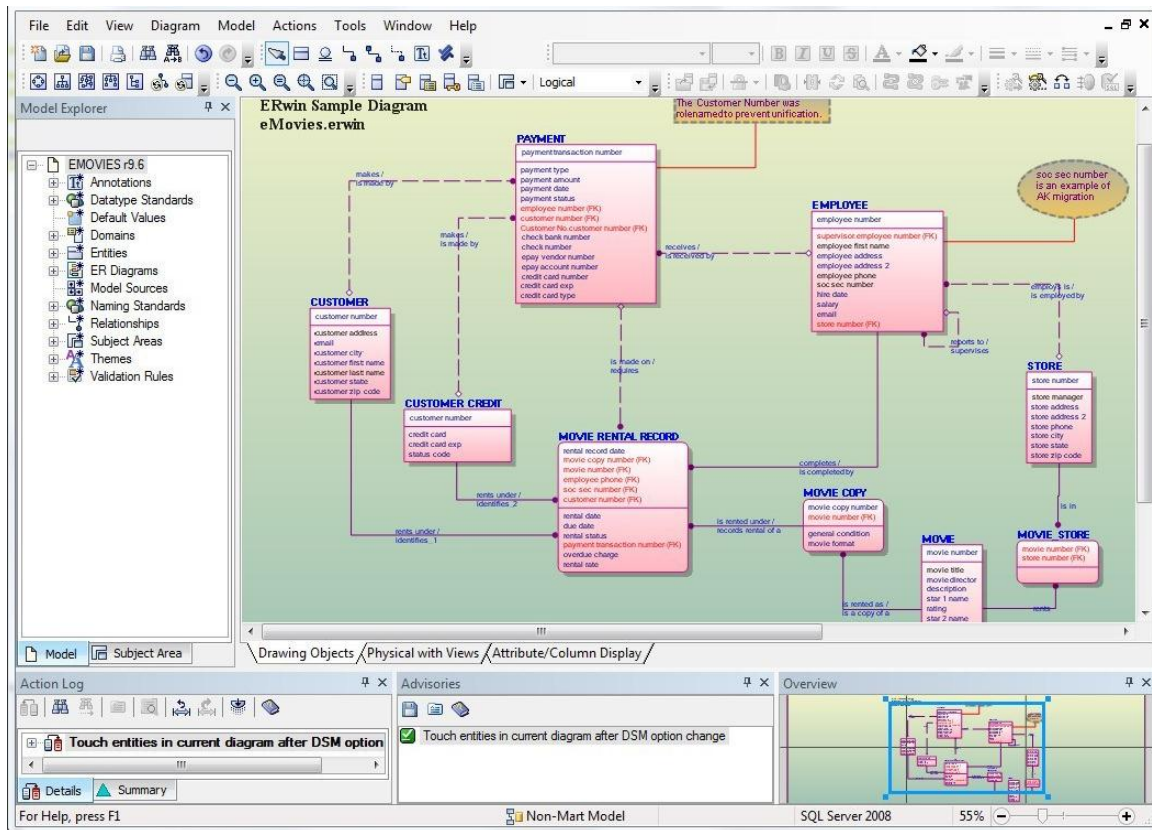








# Use Case 2: Design and Deploy New Data Sources



```

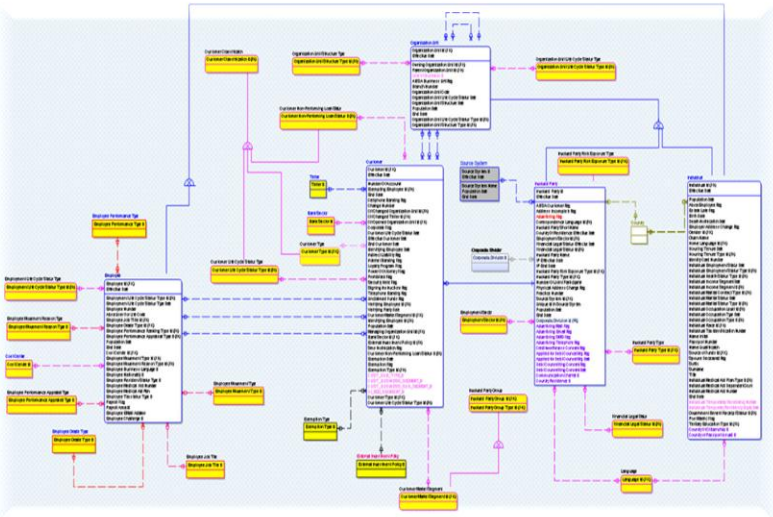
SQL Script

25 CREATE TABLE [ErrorLog] (
26     [ErrorLogID] INTEGER IDENTITY(1,1) NOT NULL,
27     [ErrorTime] DATETIME NOT NULL
28     CONSTRAINT [DF_ErrorLog_ErrorTime] DEFAULT (getdate()),
29     [UserName] sysname NOT NULL,
30     [ErrorNumber] INTEGER NOT NULL,
31     [ErrorSeverity] INTEGER NULL,
32     [ErrorState] INTEGER NULL,
33     [ErrorProcedure] NVARCHAR(126) NULL,
34     [ErrorLine] INTEGER NULL,
35     [ErrorMessage] NVARCHAR(4000) NOT NULL,
36     CONSTRAINT [PK_ErrorLog_ErrorLogID] PRIMARY KEY ([ErrorLogID])
37 )
38 GO
39
40 CREATE TABLE [Illustration] (
41     [IllustrationID] INTEGER IDENTITY(1,1) NOT NULL,
42     [Diagram] XML NULL,
43     [ModifiedDate] DATETIME NOT NULL
44     CONSTRAINT [DF_Illustration_ModifiedDate] DEFAULT (getdate()),
45     CONSTRAINT [PK_Illustration_IllustrationID] PRIMARY KEY ([IllustrationID])
46 )
47 GO
48
49 CREATE TABLE [ScrapReason] (
50     [ScrapReasonID] SMALLINT IDENTITY(1,1) NOT NULL,

```

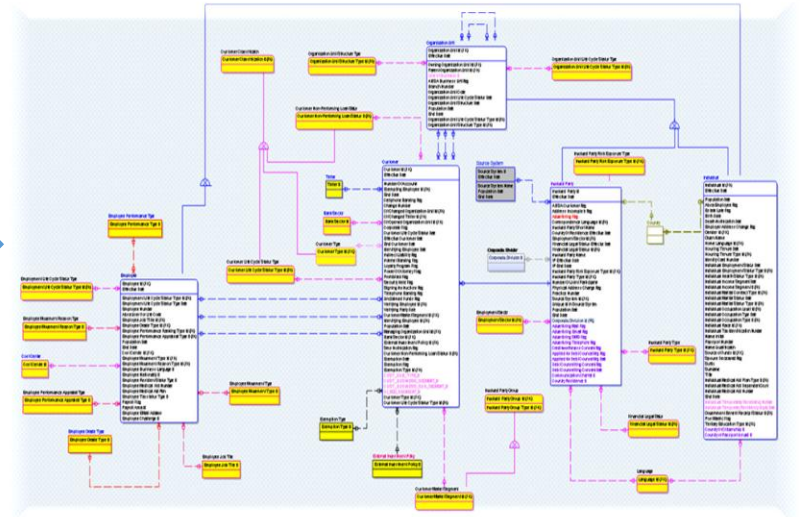


# Use Case 3: Visually compare, analyze and resolve business and technical differences

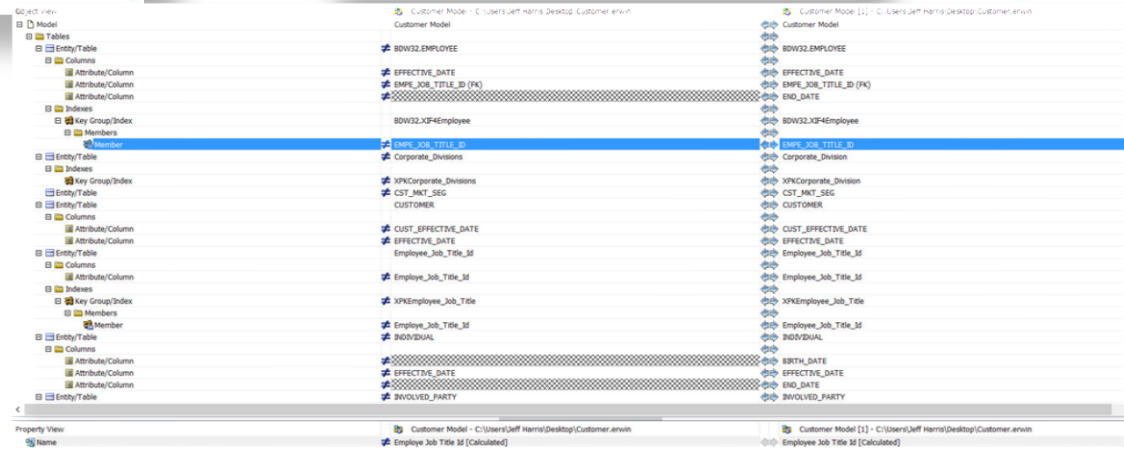


## Complete Compare

- Identify inconsistencies
- Analyze differences
- Synchronize metadata and structures
- Mark and document differences
- Report results of compare and sync processes



- Models
- Databases
- Scripts

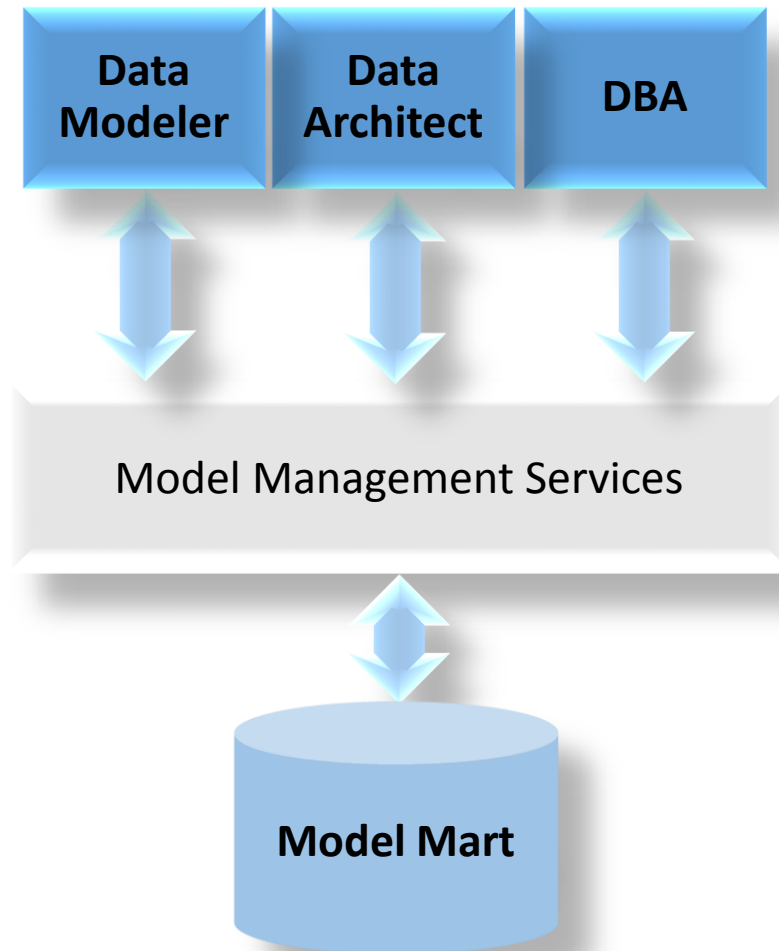


- Mark Differences
- Reuse Profiles
- Report Results





## Use Case 4: Govern Your Data Modeling and Definition



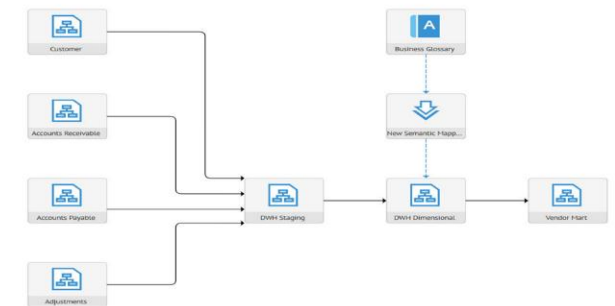
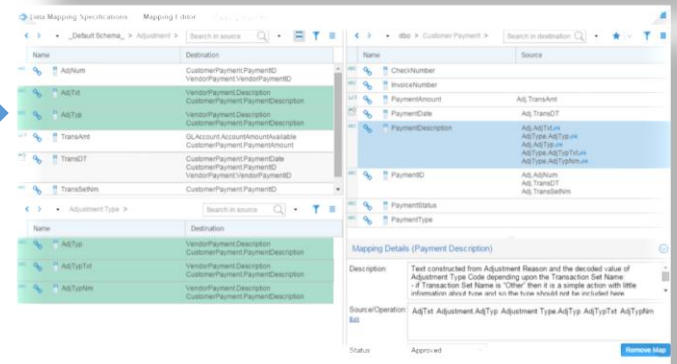
### Model Management Services

- User Access and Permissions
- Model Check In/Check Out
- Change Management
- Concurrent Modeling
- Conflict Resolution
- Version Management
- Cross Model Reporting
- Centralized Standards





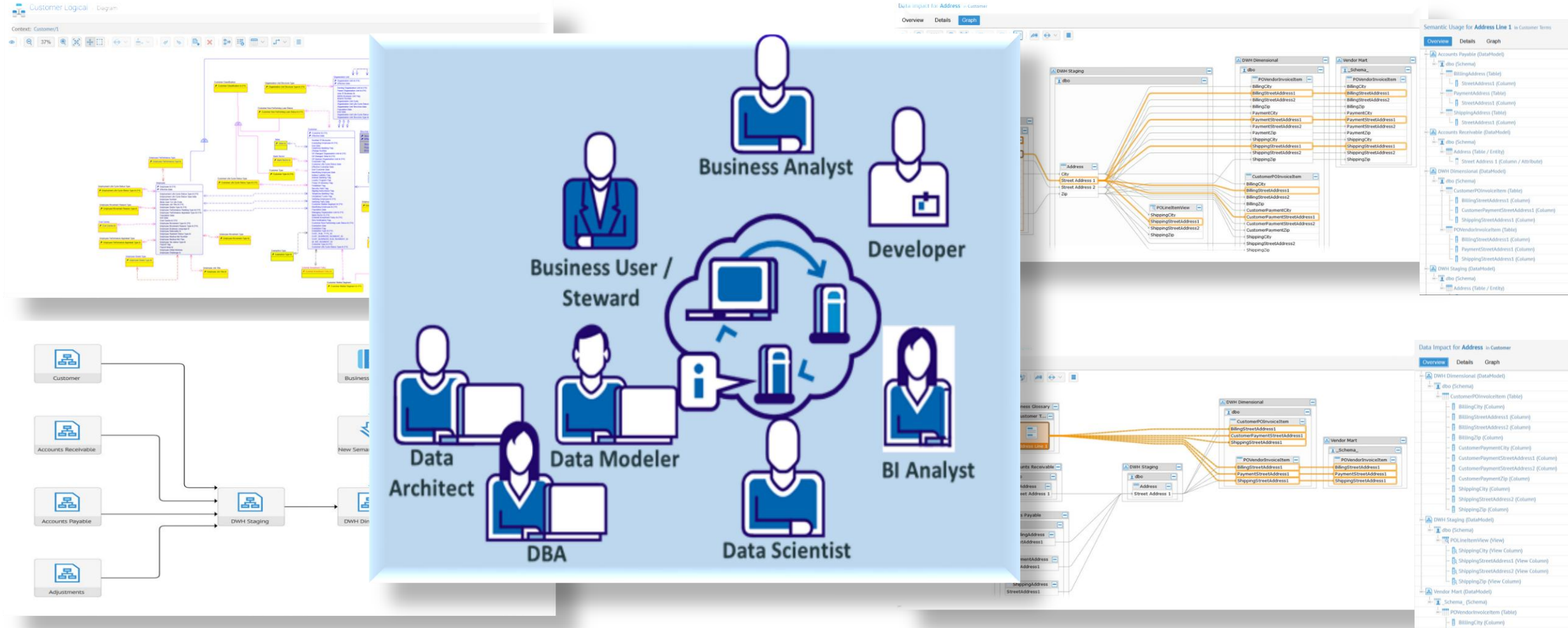
- Glossary Derivation and Authoring
- Semantic Mapping
- Dataflow Mapping
- Configuration Models
- Lineage and Impact Analysis
- Model Visualization
- Metadata Drill-Down
- Metadata Reporting
- Metadata Tags
- Metadata Authoring







# Use Case 6: Enable Stakeholder Exploration and Collaboration







# Key Characteristics of the Data-Driven Enterprise

Recognizes data as a strategic asset to drive growth and competitive differentiation

Integrates new data sources into existing business data architectures

Democratizes enterprise data through increased fluency, accountability and collaboration

Maintains an agile and effective data management approach that promotes increased strategic data usage while minimizing associated risks