Understanding Data Warehouse Needs Session #1568 Trends, Issues and Capabilities

Dr. Frank Capobianco
Advanced Analytics Consultant
Teradata Corporation

Tracy Spadola CPCU, CIDM, FIDM
Practice Lead - Insurance Industry Consulting
Teradata Corporation

Monday, March 9, 2009

ANTITRUST Notice

The Casualty Actuarial Society is committed to adhering strictly to the letter and spirit of the antitrust laws. Seminars conducted under the auspices of the CAS are designed solely to provide a forum for the expression of various points of view on topics described in the programs or agendas for such meetings.

Under no circumstances shall CAS seminars be used as a means for competing companies or firms to reach any understanding – expressed or implied – that restricts competition or in any way impairs the ability of members to exercise independent business judgment regarding matters affecting competition.

It is the responsibility of all seminar participants to be aware of antitrust regulations, to prevent any written or verbal discussions that appear to violate these laws, and to adhere in every respect to the CAS antitrust compliance policy.

The State of the Industry

Some Insurance Industry Drivers

- Competitive Landscape
  - > Protection & Growth of Market Share
  - > Need for Better Pricing and Segmentation Strategies
  - > Merger & acquisition vs. Organic growth
- Claims Costs
  - > Catastrophe losses
  - > Rising healthcare costs
  - > Need to contain overall Claims Expenses
- Current Financial Crisis
  - > Market and Credit Crisis
  - > Potential Loss of Consumer Confidence
  - > Investment Income Down

New Insights

- Using the data warehouse to “Close the Books”
  - > Many organizations have made the warehouse a key component of month-end close, giving them the ability to close the books faster and earlier in the month-end cycle
    - – Some can close the books at any time, and have results in minutes
    - – Statistical / Regulatory Reporting from the Data Warehouse
- Understanding the “Quote to Bind” process
  - > Many organizations are stepping up analysis to capture ALL quote iterations for better understanding of acquisition process and costs
- Moving to a Customer-Focused Organization
  - > Insurance companies has typically been organized by product.
  - > Insight into customer portfolios, cross sell / upsell, better risk segmentation and customer service

The Data Drivers

- Enhanced Customer Intelligence
  - > Single view of the Customer
  - > Move from Product to Customer View
- Increased / Improved Risk Analysis
  - – Improved Underwriting / Pricing
  - – Catastrophe Modeling
    - – Global warming issues
  - – Enterprise Risk Management
- Increased reliance on Predictive Modeling Capabilities
  - > Segmentation
  - > Loss characteristics
  - > Fraud & Abuse
- Protection / Growth of Market Share
  - > Analysis of distribution channels and pipeline activity

New Insights with New Data Types

- Telematics
  - > Capturing real time telematics for pricing and risk assessment for "Pay as You Drive" programs
- Unstructured Text
  - > Analysis of Claims / Adjuster notes for better Loss and Fraud analysis
  - > Analysis of Call Center records for better customer service / marketing "Voice of the Customer" analysis
- Geographic Information Systems (GIS)
  - > Leading US Insurers using geo-spatial data to support risk exposure analysis (concentration / location), reinsurance and regulatory compliance
- Dynamic Financial Analysis
  - > Risk (financial, market, credit) modeling for improved financial reporting, capital management and Enterprise risk management

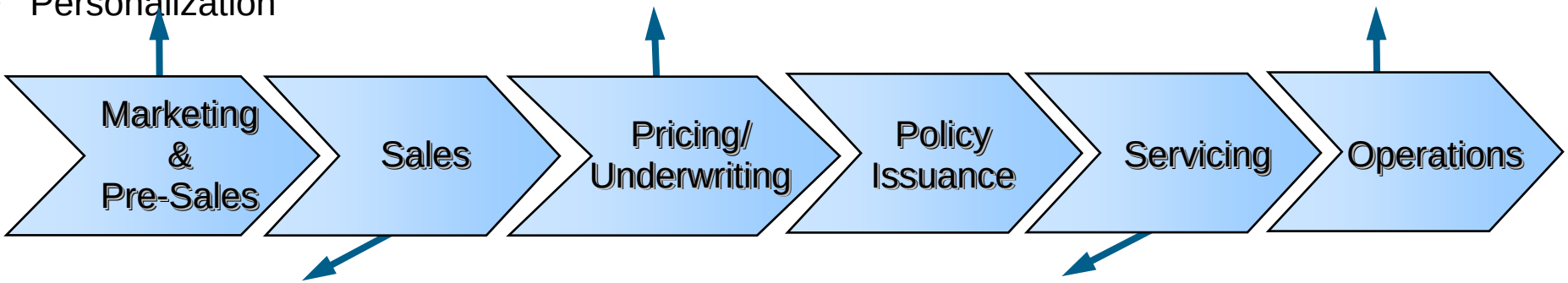
Data Warehousing Challenges and Capabilities

Why do Insurance Companies Use / Need Data Warehouses?

- Campaign development and customer segmentation
- Customer behavior assessment using predictive modeling
- Event-based marketing
- Channel analysis
- Personalization

- Using internal and third-party data to price products to risk
- Underwriting: pricing analysis, frequency/severity analysis and exposure analysis
- Risk profiling

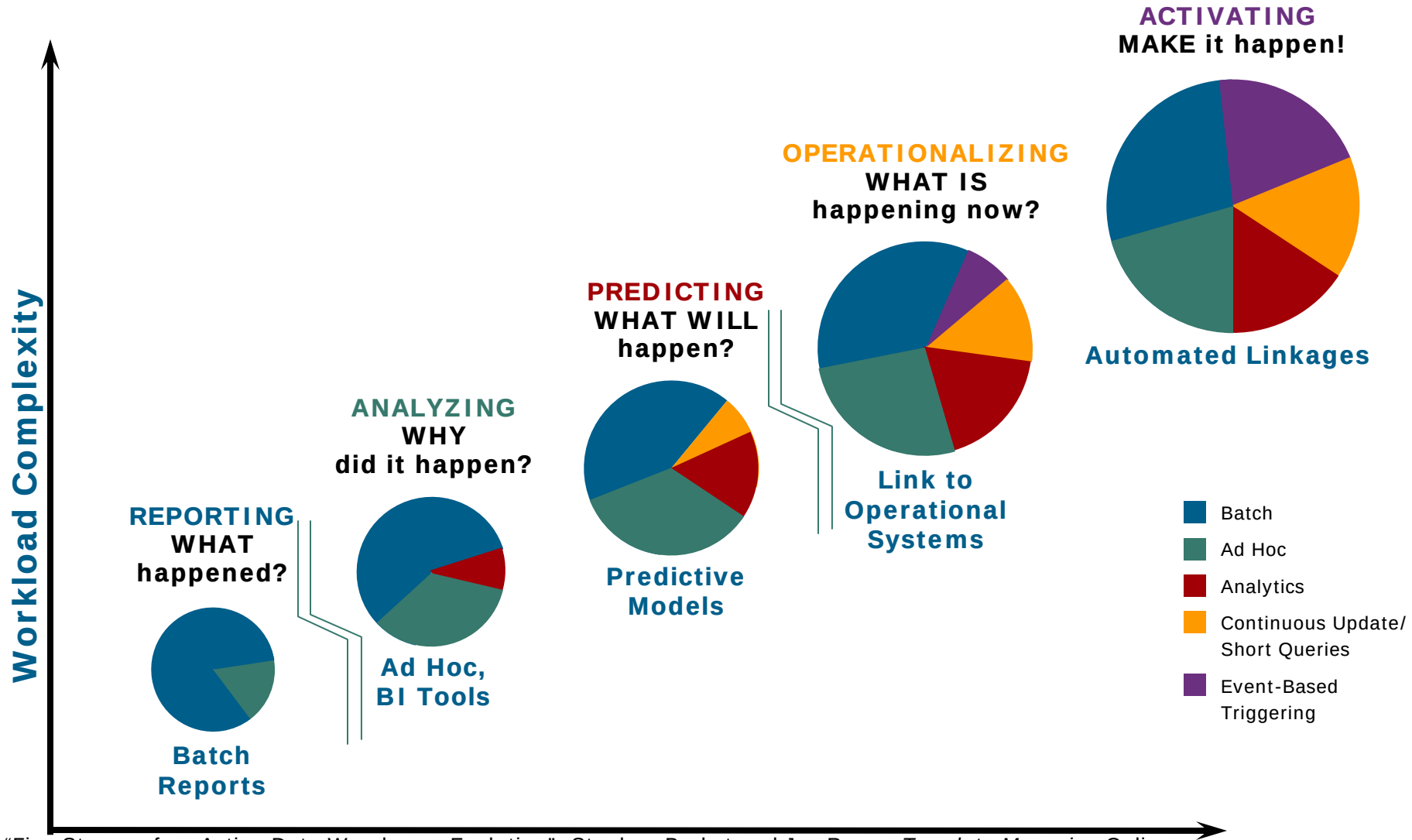
- Financial analysis
- Regulatory reporting
- Reinsurance analysis
- Operational performance analysis



- Identification of upselling and cross-selling opportunities
- Assessment of channel, region and agent performance
- Agent loyalty assessment

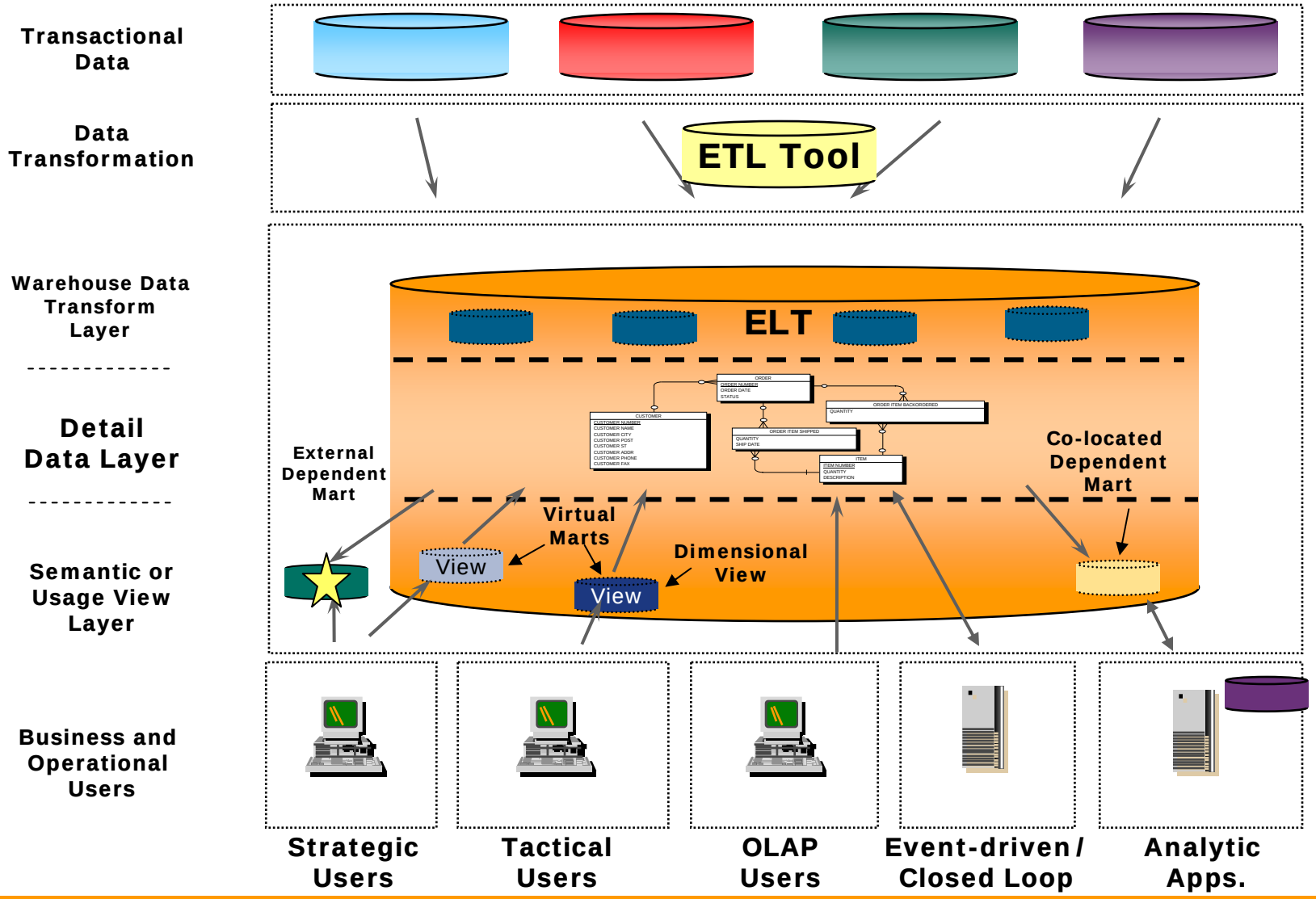
- Customer information analysis to determine service preferences
- Personalized and knowledgeable interaction via all channels as information is delivered to agents, brokers and other partners
- Claims performance and claims workflow assessment
- Fraud identification and detection

Life-Cycles of Data Warehousing



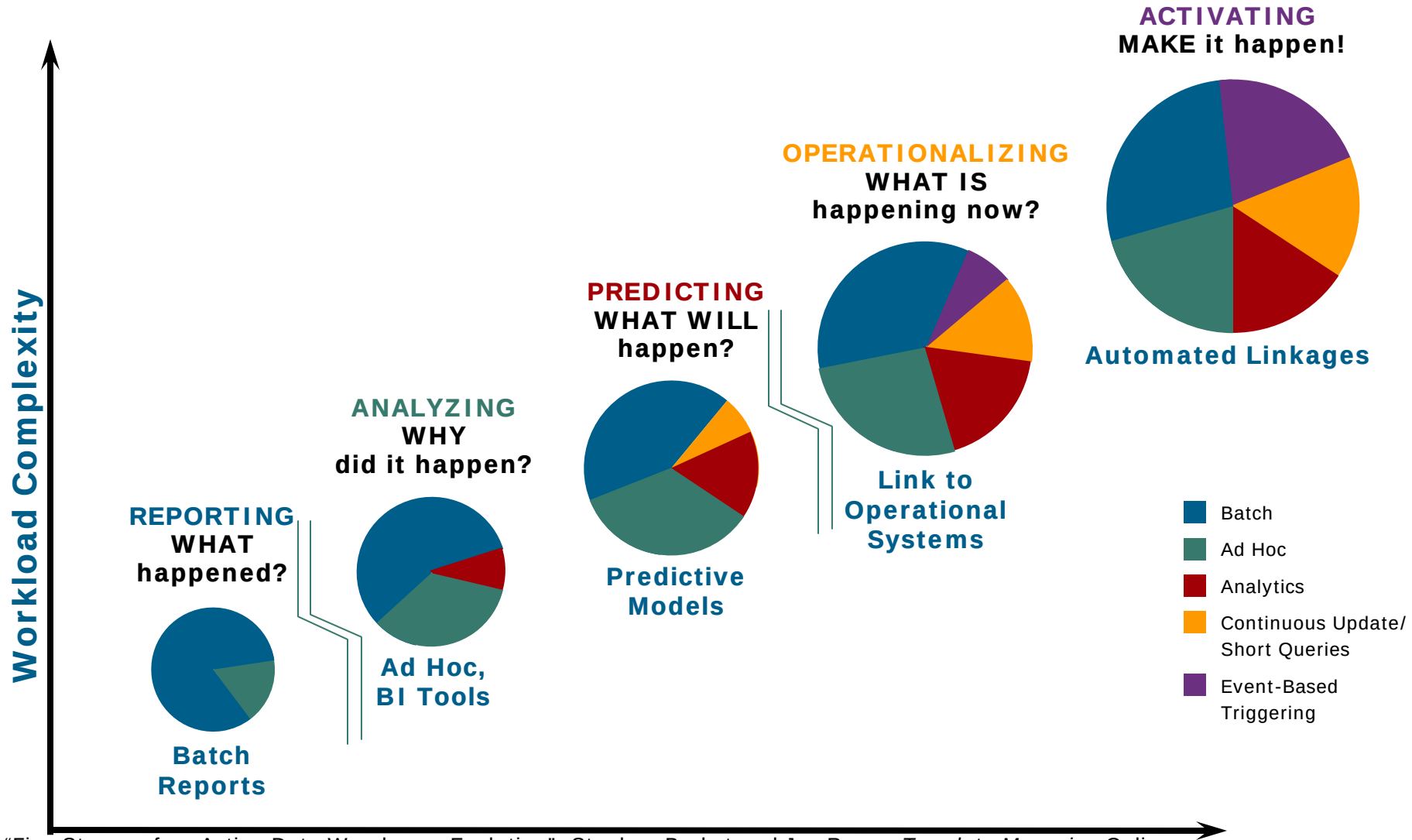
"Five Stages of an Active Data Warehouse Evolution", Stephen Brobst and Joe Rarey, *Teradata Magazine* Online

Warehouse Reference Architecture – The Target Topology



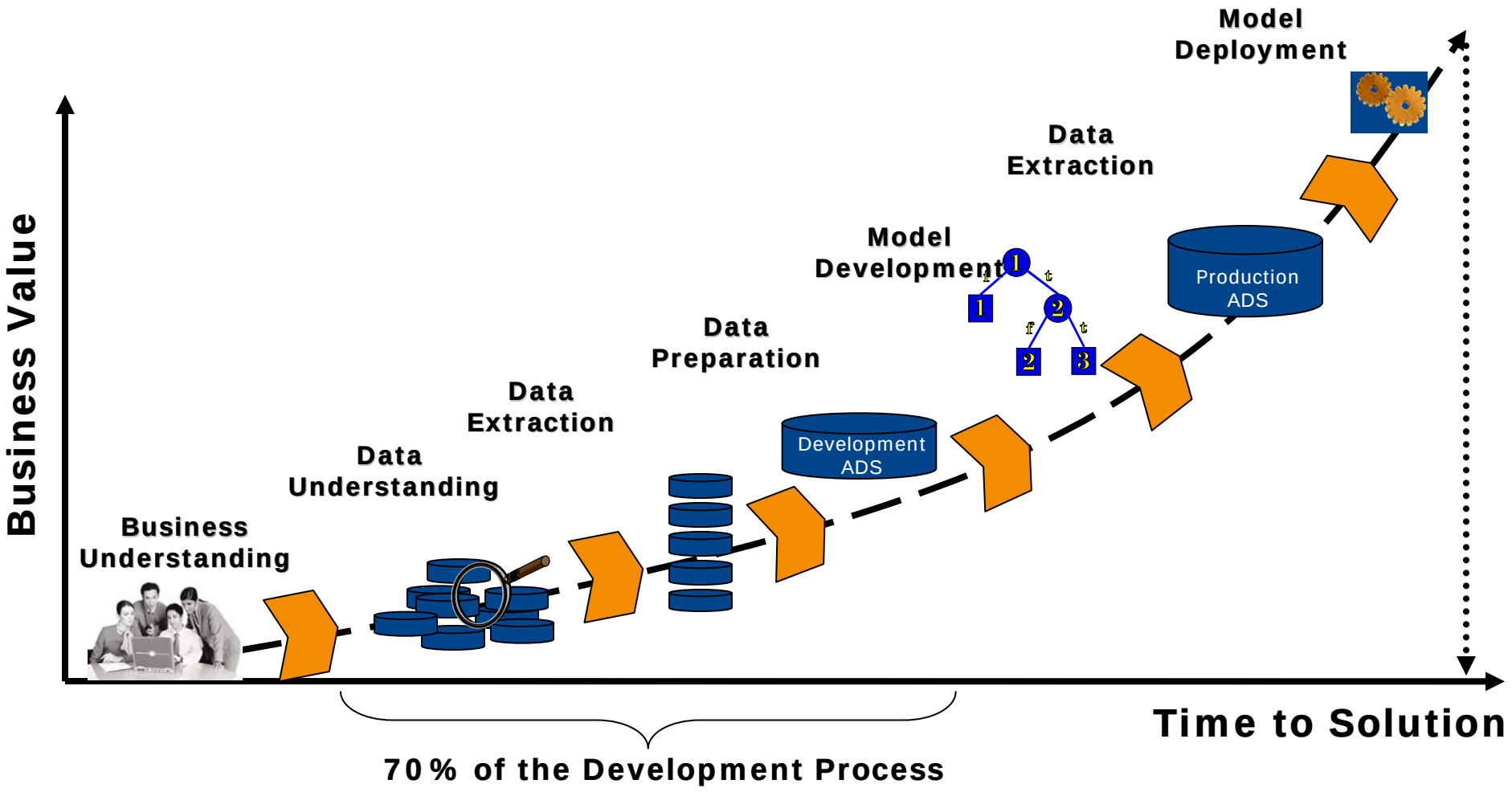
Supporting the Analytics Enviroment

Business is driven by need for actionable information



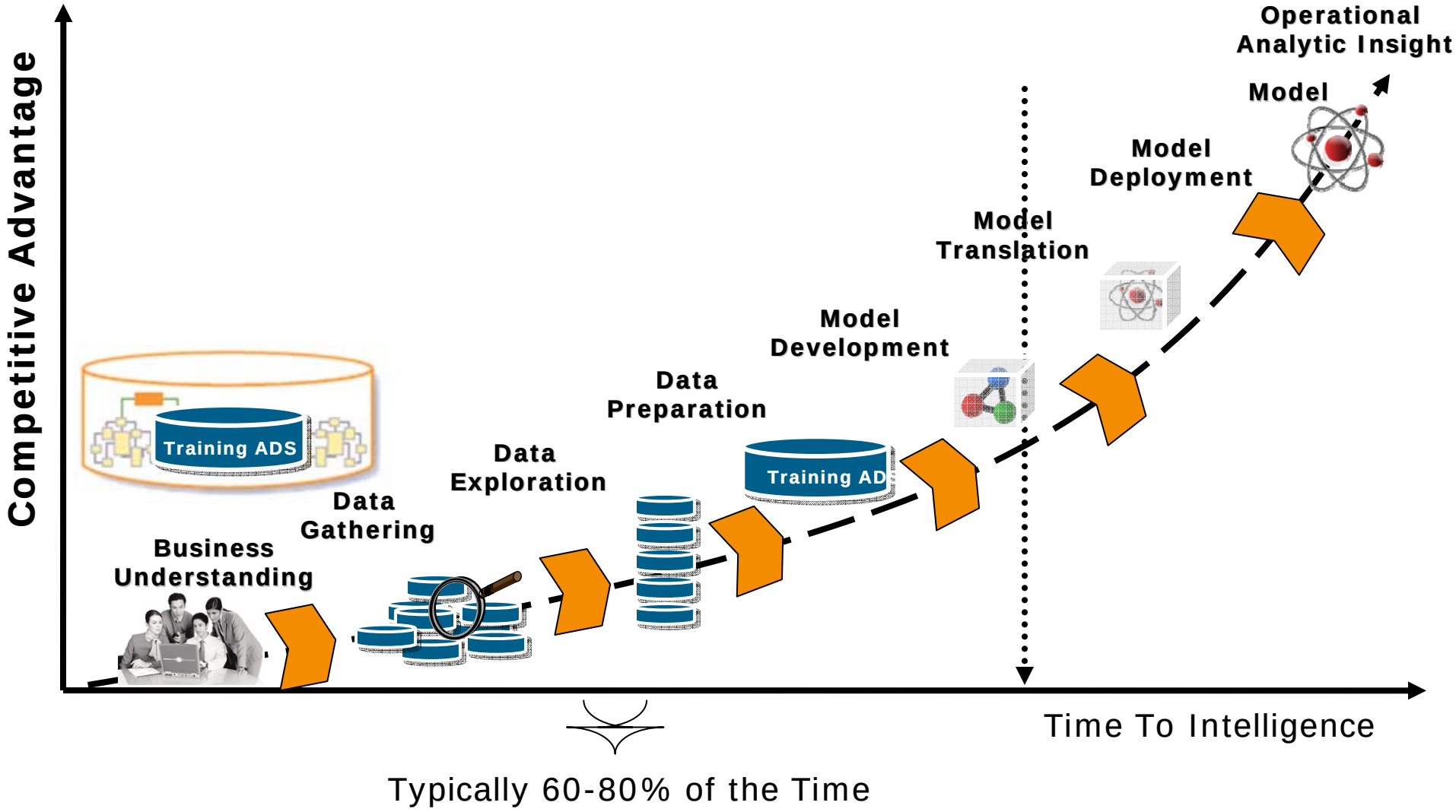
"Five Stages of an Active Data Warehouse Evolution", Stephen Brobst and Joe Rarey, *Teradata Magazine* Online

A Typical Analytic & Predictive Process

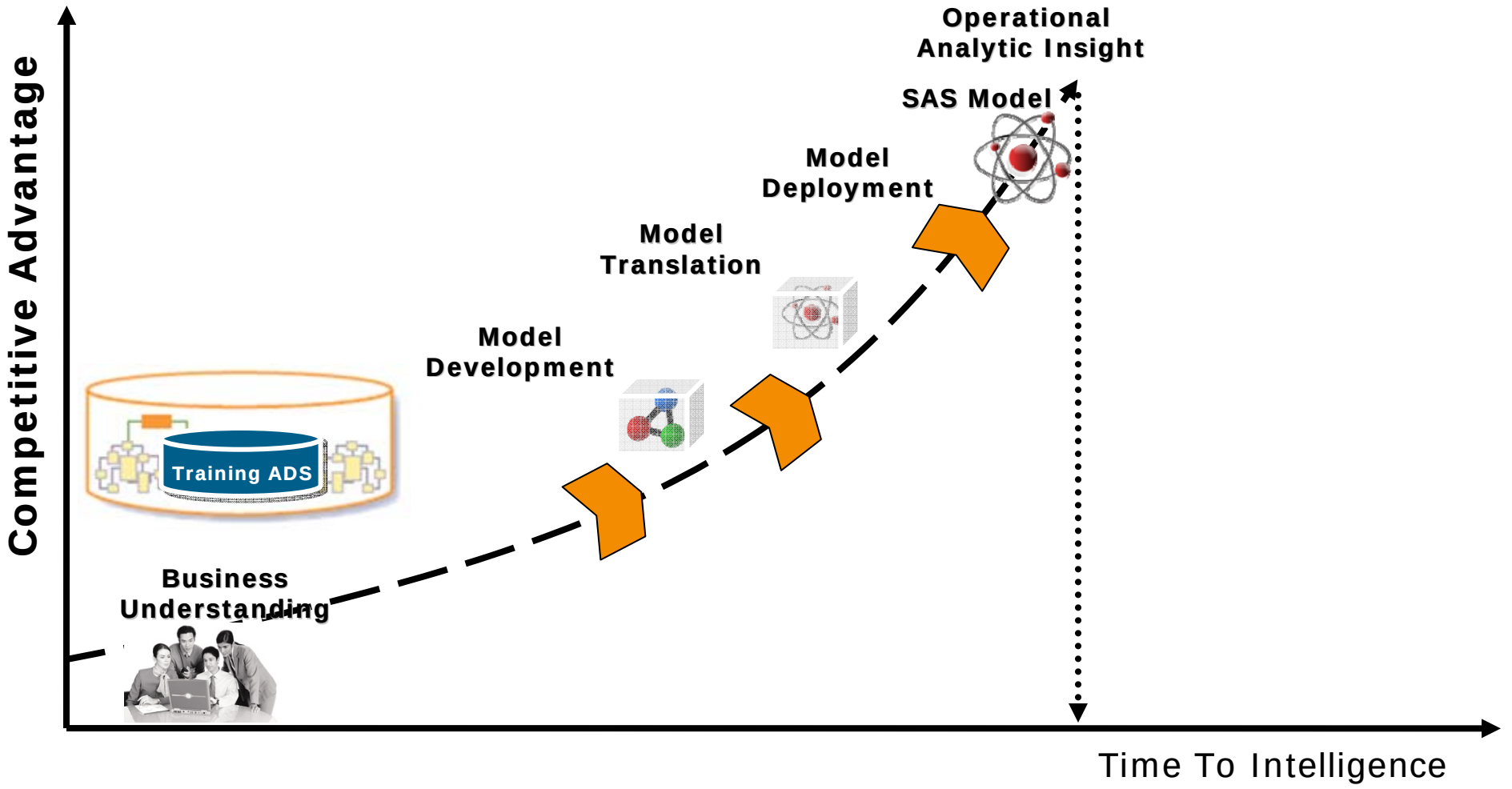


Time to Build & Deploy Models
Weeks to Months

A Typical Analytic & Predictive Process

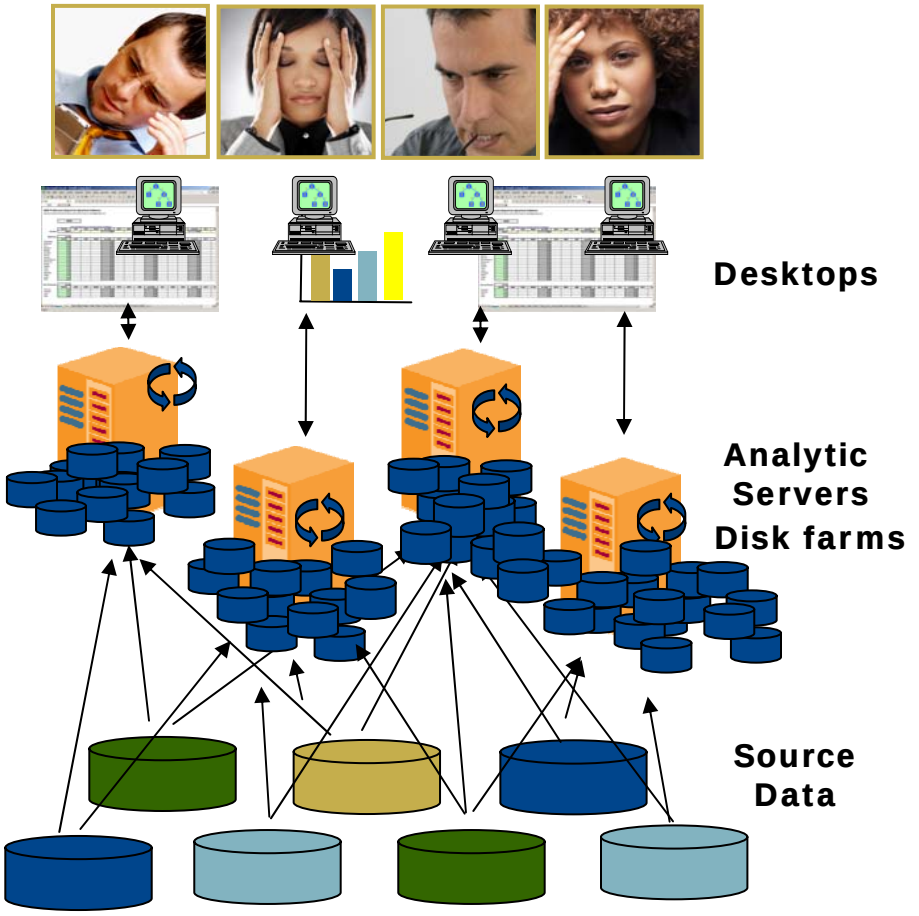


Gain Agility, Speed and Quality



Current Environment Challenges

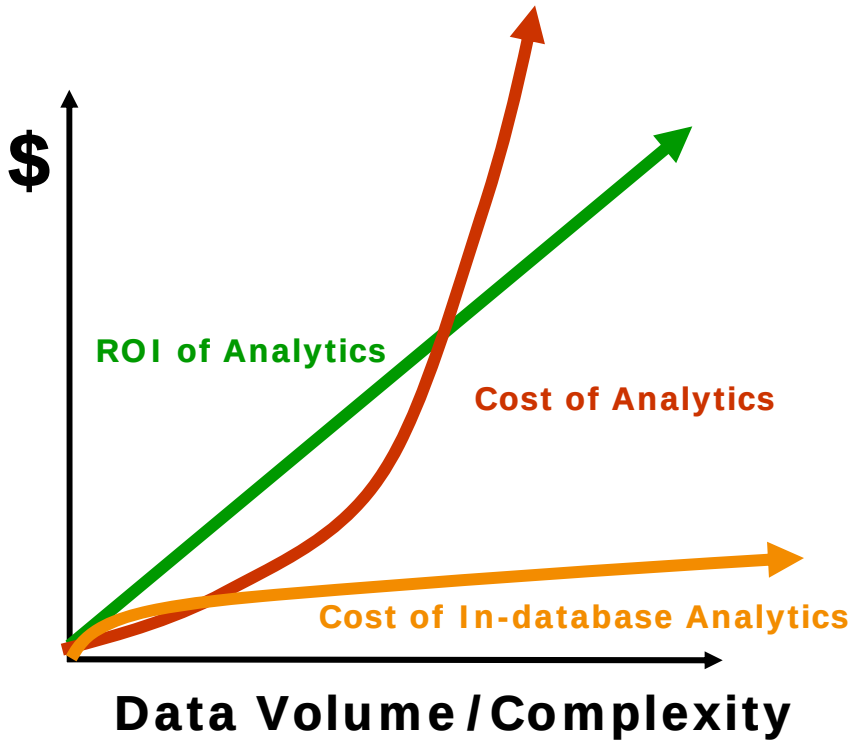
- Data volumes & complexity continues to grow exponentially
- Many tools replicate data on to an analytic server/data mart for processing
- Inefficient process: Slow Time-to-Market
- Inconsistency throughout your enterprise



The Cost of A Suboptimal Environment

As data complexity and volumes grow, so does the cost of building analytic models.

- How much time is your analyst spending on data movement and SQL programming vs. analysis?
- How much time is model development taking?
- How much time is your IT department spending extracting data for your analytic modelers?
- Is data movement saturating your network?
- Is your analytic server under capacity?
- How long does it take to get the results of your predictive model to your business users?



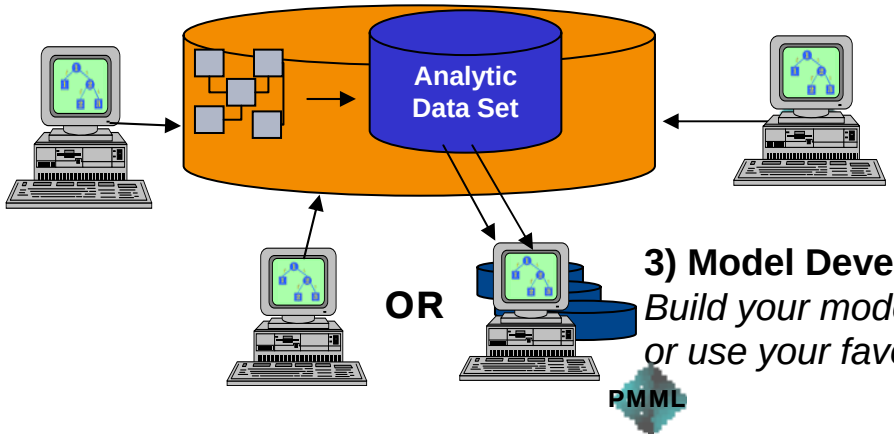
Optimal Analytic Environment
- exploits in-database processing

Provides enterprise analytic capabilities to efficiently develop and deploy advanced analytics

Delivers a flexible, scalable environment for enterprise analytics

- Performs appropriate tasks on appropriate platform
  - ✓ Data intensive tasks (exploration, preparation & scoring) in the database
- Leverages partner products integrated with the database
  - ✓ Partner products to leverage the database for data prep & scoring
  - ✓ Minimize data movement into an analytic mart or server
  - ✓ Execute model and applications directly in the database

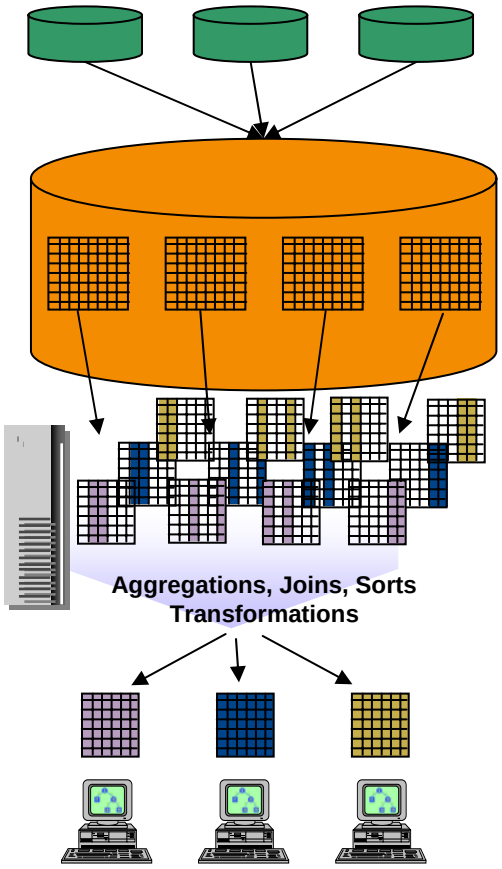
1) Data Exploration &
2) Data Preprocessing
Perform data Intensive steps in the database



4) Model Scoring
Export Code or PMML model from your favorite data mining tool

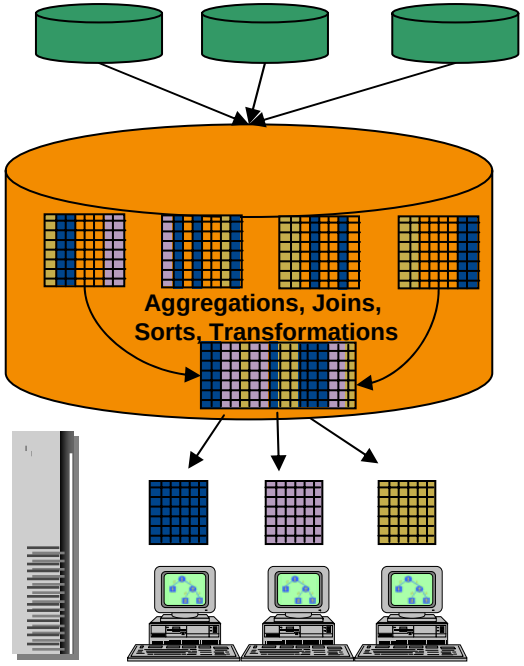
3) Model Development
Build your model in-database for efficiency or use your favorite data mining tool

Data Exploration and Data Preprocessing



Data Exploration and Preprocessing Outside the Database

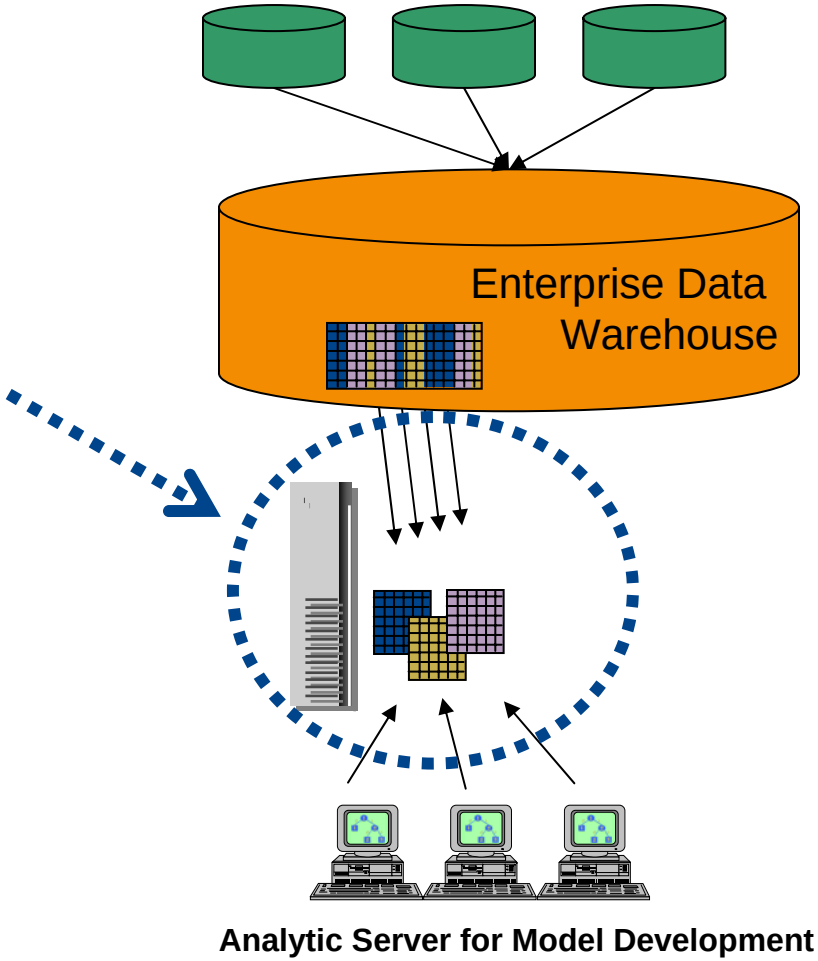
- Know your data prior to extracts
  - Explore and identify your data relevant to your application
  - Avoid pulling irrelevant tables and columns
- Prepare data prior to extracts
  - Perform derivations, aggregations and transformations then extract the results
- Leverage in-database tools to streamline process
  - Easy to use environment for Analyst and IT team
  - Automated SQL generation
  - Reusable metadata in-database



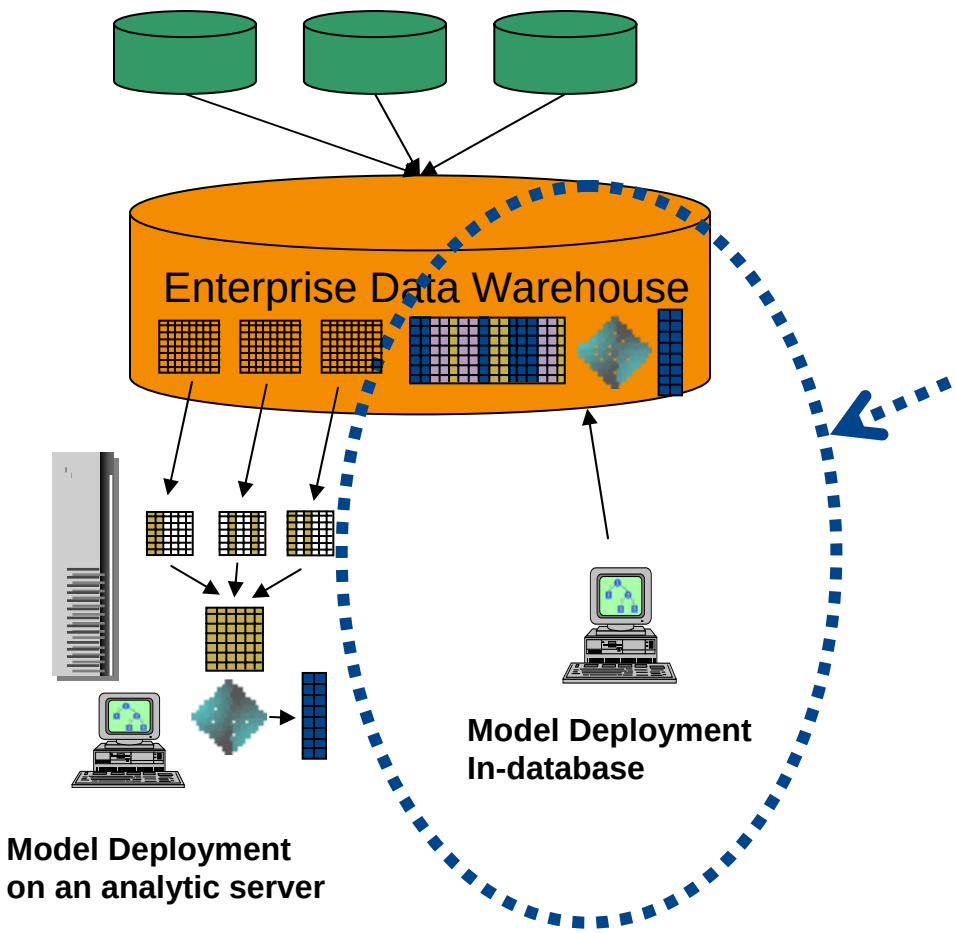
Optimized Environment

Analysis and Advanced Analytic Modeling

Leverage tools like SAS for its strength: Analytics



In-Database Model Deployment



- Suboptimal methods:
  - (1) Extract data and score outside the database
  - (2) Scoring code re-written in SQL and submitted to RDBMS
- Optimal method:
  - Use automatic score generation capabilities of data mining tools, and run them in-database against the Analytic Data Set

Example: SAS/Teradata Partnership

Data Exploration

Explore all the data directly in the database with Teradata Profiler

TERADATA. Profiler

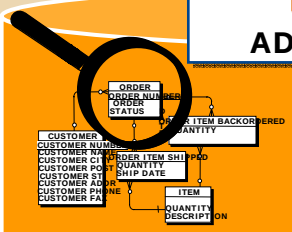
Analytic Modeler

- Know your data before extracting
- Eliminate redundancy and unused data
- Analyze large complex tables quickly

Data Preparation

Transform and aggregate data in the database with Teradata ADS Generator

TERADATA. ADS Generator



Modeling ADS

Model Development

Sample your ADS data and build your model on a server with SAS Enterprise Miner

sas Enterprise Miner

Analytical Model

sas Access To Teradata

Production ADS

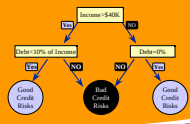
- Visualize your sample analytical data store
- Build analytical models more efficiently
- Produce consistent results that are "actionable"

Model Deployment

Deploy your SAS model as a In-DBS Function in the Teradata database

sas Scoring Accelerator

In-DBS Functions



Enterprise Consumer

- Export SAS models as Teradata functions for in-database scoring
- Eliminate labor-intensive model translation efforts
- Get to better answers faster

Success Story: Discover Financial Services

Key Points

Discover Financial Services operates the Discover Card, America's cash rewards pioneer

Discover Network with millions of merchants and cash access locations

PULSE ATM/debit network with more than 260,000 ATMs, and serving over 4,500 financial institutions

Business Challenge

- Each analyst pulled data from 1000s of source files for analytics
- Slow and inefficient model development process required 3+ weeks to complete
- Challenges with scoring extended runtimes to 7+ days
- Inconsistent data compromised the results

Outcome and Benefits

- Built a 1400 variable analytical data store in Teradata
- Reduced data preparation from 3+ weeks to 90 minutes
- Model development time was reduced from 14 to 2 weeks
- Accelerated analytic model runtime from 7+ days to 36 minutes
- Trained 450 analysts on better leveraging SAS and Teradata, with plans to increase analytic output by10-fold from 40 to 500 models

Case Study: Warner Home Video

Key Points

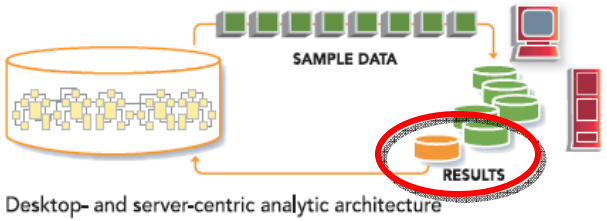
Forecasting New DVD Demand

The window of opportunity is short. With a major DVD release, 45 per cent of the total video sales happen in first week, and 75 per cent within the first month

Using the combined strength of SAS and Teradata, some smart forecasting helped Warner Home Video sell 68 million DVDs of the three Harry Potter movies, and lead the industry with sales of $4.75B

Business Challenge

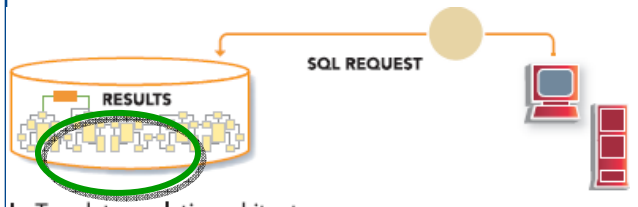
- **Accurately estimating the demand and shipping of new movie titles**
- **A small window of opportunity and constantly changing predictors presented significant challenges**
- **New release forecasts were requiring 36 hour runtimes for approximately 300 DVD titles**



Desktop- and server-centric analytic architecture

Outcome and Benefits

- **Higher sales with forecasting decisions based on analytic intelligence**
- **Using the combined strength of Teradata and SAS, the same new release forecast now takes an hour and 15 minutes**
- **Similar performance gains were realized across other scoring activities, ie from 2 hours to 1 minute**



In-Teradata analytic architecture

Contact Information

Dr. Frank Capobianco has many years experience as an advanced analytics consultant for Teradata. In that role, Dr. Capobianco has helped clients solve data mining problems, such as posed by forecasting, fraud detection or marketing campaign development. His clients span the financial, insurance, telecommunications, retail, travel, transportation, and manufacturing industries. Dr. Capobianco received his Ph.D. in mathematics from the University of Virginia.

Email Frank at ***frank.capobianco@teradata.com***

Tracy A. Spadola is the insurance practice lead in the Industry Consulting Group. She has worked across multiple industries to help develop Enterprise Data Management strategies and uncover Business Value for these projects. Ms. Spadola is the current president of the Insurance Data Management Association (IDMA).

Email Tracy at ***tracy.spadola@teradata.com***