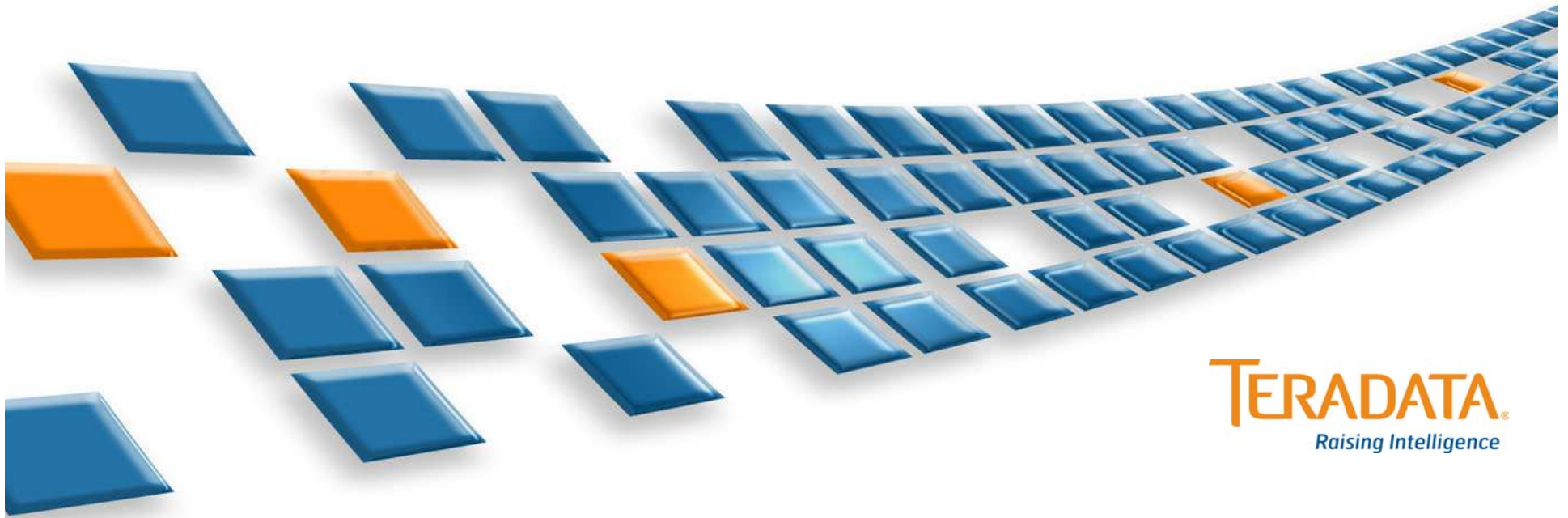


The Data Warehouse and Business Information

Alison Torres

Director Teradata Global Warehouse Consulting

Teradata Certified Master V2R3, V2R5



TERADATA
Raising Intelligence

Alison Torres

Teradata Evangelist

- Alison Torres, Director Teradata Warehouse Consulting, has been working with Teradata since 1986, and is recognized globally for her mastery in the areas of data warehouse design, implementation, and performance.
- Her work in the computer industry has spanned over 25 years and has been dedicated to driving business value from analytical processing. She has led teams that have delivered some of the most successful enterprise data warehouses in existence.
- Alison has been a Teradata Partners Conference Speaker for 20 years. Some of her past presentations include: Explain the EXPLAIN; A Guide to System Management; I've Got the EXPLAIN - Now What?; An Introduction to Teradata; and Embedded Views: Maze, Menace, or Magnificent?
- Alison is a TDWI Faculty member.



Alison Torres
Teradata Certified Master



Agenda

- Data Warehouse Definitions
- Enterprise model for data warehouses
- ETL/ELT processes
- Role of data warehouses in business intelligence
- Data mining / Data analytics
- Data visualization / Managing Data
- Industry practices in business intelligence



Data Warehouse Definitions

Data Warehouse Definition - From Wikipedia

Data warehouse is a repository of an organization's electronically stored data. Data warehouses are designed to facilitate reporting and analysis.

- This definition of the data warehouse focuses on data storage. However, the means to retrieve and analyze data, to extract, transform and load data, and to manage the data dictionary are also considered essential components of a data warehousing system. Many references to data warehousing use this broader context.
- Thus, an expanded definition for data warehousing includes business intelligence tools, tools to extract, transform, and load data into the repository, and tools to manage and retrieve metadata.
- In contrast to data warehouses are operational systems which perform day-to-day transaction processing.

Enterprise Model for Data Warehouses

Enterprise Data Warehousing

Data warehousing is the *process* of capturing, storing and analyzing data to gain insight. This process is built on an enterprise data warehouse which is a **single, centralized, application-neutral, integrated repository** of an organization's current and historical data.



Active Data Warehousing

The foundation of enterprise intelligence, active data warehousing **drives** valuable information into the company's operations and to hundreds of front-line decision-makers. It enables smarter, more competitive decisions through **near real-time** information access and analysis and **predictive analytics** on customer or business activity. Active data warehousing supports both long-running strategic and short-running tactical queries.

The active data warehouse provides **strategic** and **tactical intelligence** throughout the enterprise.



Active Enterprise Data Warehouse

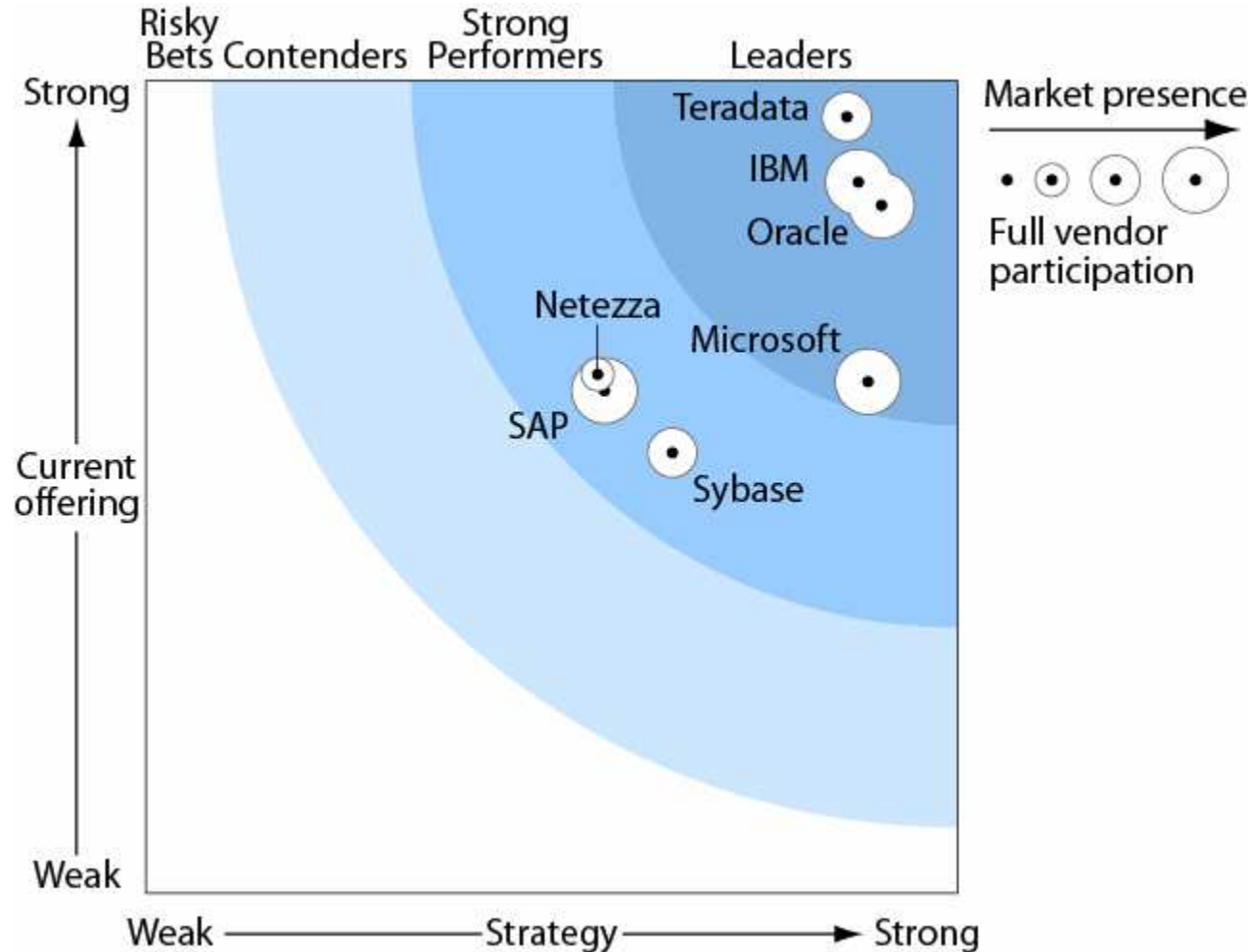
Active Enterprise Intelligence

- **Active enterprise intelligence** is what an active data warehouse delivers. Active enterprise intelligence is the **alignment of strategic and operational systems, people, technology and processes** to make better, faster decisions throughout all levels of the organization.
- This capability helps decision-makers and front-line workers achieve new productivity levels, speed and precision in decision-making, manage business complexities, and **respond to business events while they are occurring.**



The Forrester Wave™: Enterprise Data Warehousing Platforms, Q1 '09

FORRESTER®



February 2009 "The Forrester Wave™: Enterprise Data Warehousing Platforms, Q1 2009"

TERADATA
Raising Intelligence

ETL/ELT Processes

Data Integration Techniques

Three main approaches:

ETL Approach:

Extract from the source systems

Transform inside the ETL engine on integration engine servers

Load into target tables in the data warehouse

ELT Approach:

Extract from the source systems

Load into staging tables inside the data warehouse RDBMS servers

Transform inside the RDBMS engine using generated SQL with a final insert into the target tables in the data warehouse

Hybrid ETLT Approach:

Extract from the source systems

Transform inside the ETL engine on integration engine servers

Load into staging tables in the data warehouse

Apply further **Transformations** inside the RDBMS engine using generated SQL with a final insert into the target tables in the data warehouse

ETL versus ELT

When does **ETL** win?

- Ordered transformations not well suited to set processing.
- Integration of third party software tools best managed by data integration engine outside of the RDBMS (e.g., name and address standardization utilities).
- Maximize in-memory execution for multiple step **transformations that do not require access to large volumes of historical or lookup data** (note: caching plays a role).
- Streaming data loads using message-based feeds with “real-time” data acquisition.

ELT versus ETL

When does **ELT** win?

- Leverage of high performance DW platform for execution reduces capacity requirements on ETL servers – this is especially useful when peak requirements for data integration are in a different window than peak requirements for data warehouse analytics.
- Significantly reduce data retrieval overhead for **transformations that require access to historical data** or large cardinality lookup data already in the data warehouse.
- Batch or mini-batch loads with reasonably large data sets.
- Optimize performance for large scale operations that are well suited for parallel execution using set operations such as complex joins and large cardinality aggregations.

Which Approach is Best?

- The Answer: Hybrid **ETLT**
- Best of both worlds!
- Use the appropriate combination of **ETL** and **ELT** based on transformation characteristics.
- Data integration product creates and defines metadata to drive transformations independent of the execution engine.
- **ETL** versus **ELT** execution is selectable within data integration package and can be switched back-and-forth without the redefinition of transformation rules.

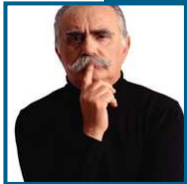
Role of Data Warehouses in Business Intelligence

Role of Data Warehouses in Business Intelligence

- A new kind of BI: **Active Enterprise Intelligence** to drive better, faster Operational Decisions
- Active Enterprise Intelligence has high value
 - > Align Strategic and Operational decisions
 - > Accelerate Operational decisions
- Many leading-edge consumers are using Active Enterprise Intelligence to make a difference
 - > Drive profitability, customer service, and operational excellence
 - > Across all industries
 - > Across multiple business functions

What Is Active Enterprise Intelligence?

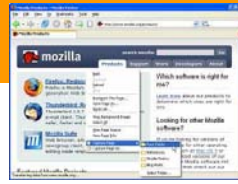
Helps You Make Better Decisions, Faster



Strategic Intelligence:
Great Insights about the Business



Align and Accelerate



Operational Intelligence:
Operations People and Systems
Become Smarter and Faster

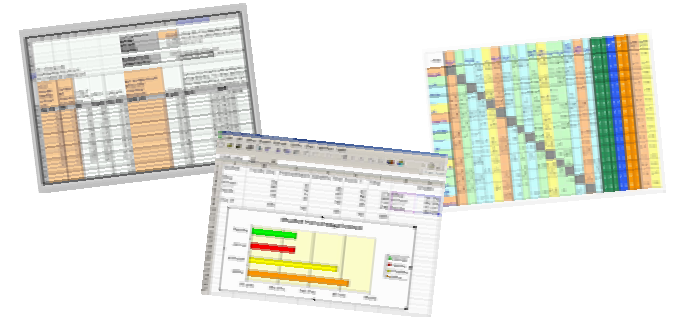
What Is Active Enterprise Intelligence?

- **Active**
 - > Better, faster decisions that drive actions
 - > Responsive and agile, make changes when needed
- **Enterprise**
 - > Consistency – single view of the business
 - > Scope – across appropriate business functions
 - > Reach – new operational users, processes, and applications
- **Intelligence**
 - > “Strategic” *Intelligence* – aligned to drive
 - > “Operational” *Intelligence* – front line



Strategic Intelligence

- **“Back Office” Company Users**
 - > Tens to hundreds of users
 - > Knowledge workers in Strategic Planning, Marketing, Finance, Quality Assurance
- **“Intelligence” for Strategic Analysis**
 - > Reporting, Forecasting, Inventory Analysis, Product Profitability Analysis, Financial Management, Customer Segmentation, Profitability, Compliance, etc.
 - > BI-Tool Centric



Operational Intelligence

- **“Front Line”** users and customers
 - > Thousands to tens of thousands of users and customers
 - > Agents, Cashiers, Workers, Customer Service Agents, Customers, and Suppliers
 - > Self-Service Systems: POC, Web
- **“Intelligence”** for **Operational Execution**
 - > Customer look up, individualized customer offers, transaction exceptions, visibility, event detection and notifications
 - > Application-Centric



Accelerating Operational Decisions

- While the service agent is still interacting with the customer
- While the customer is still at the counter
- While the customer service agent is still talking/listening
- While the web page is still in front of the customer
- While 'the crook' is still defrauding your state
- While the vehicle is still being loaded
- While the supplier still has an on-sale or in-stock situation

Point-of-Contact



Call Center



Self-Service Kiosk



Self-Service Web



Loading Dock



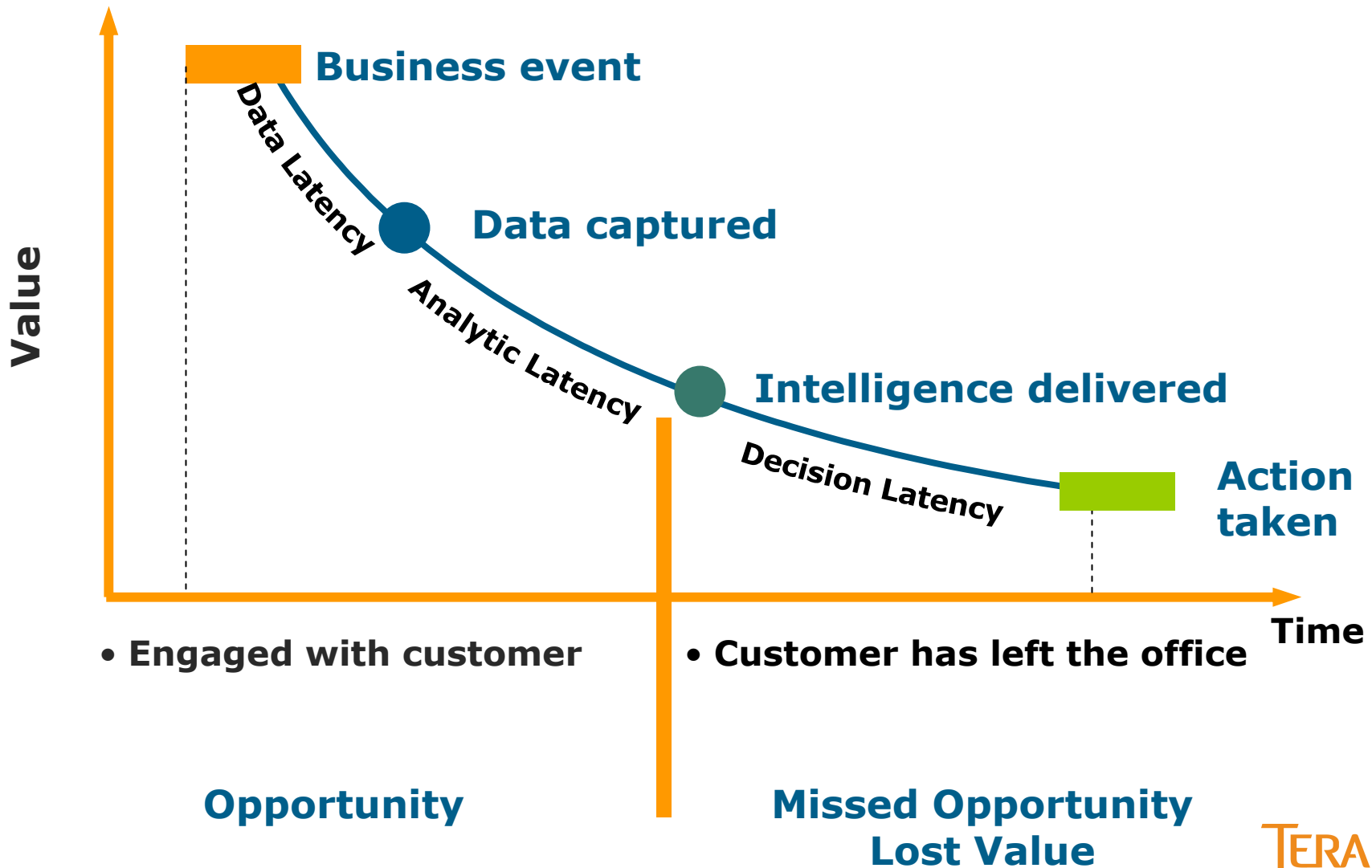
Scheduling



Supplier



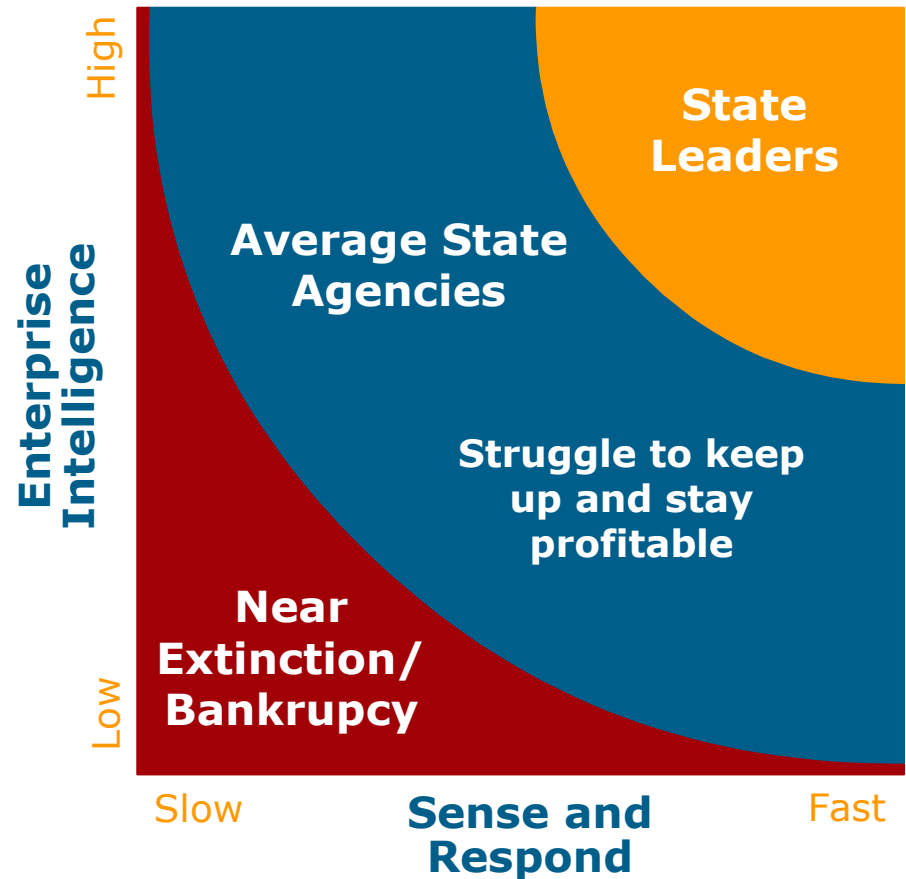
Accelerating Operational Decisions



Why Align and Accelerate?

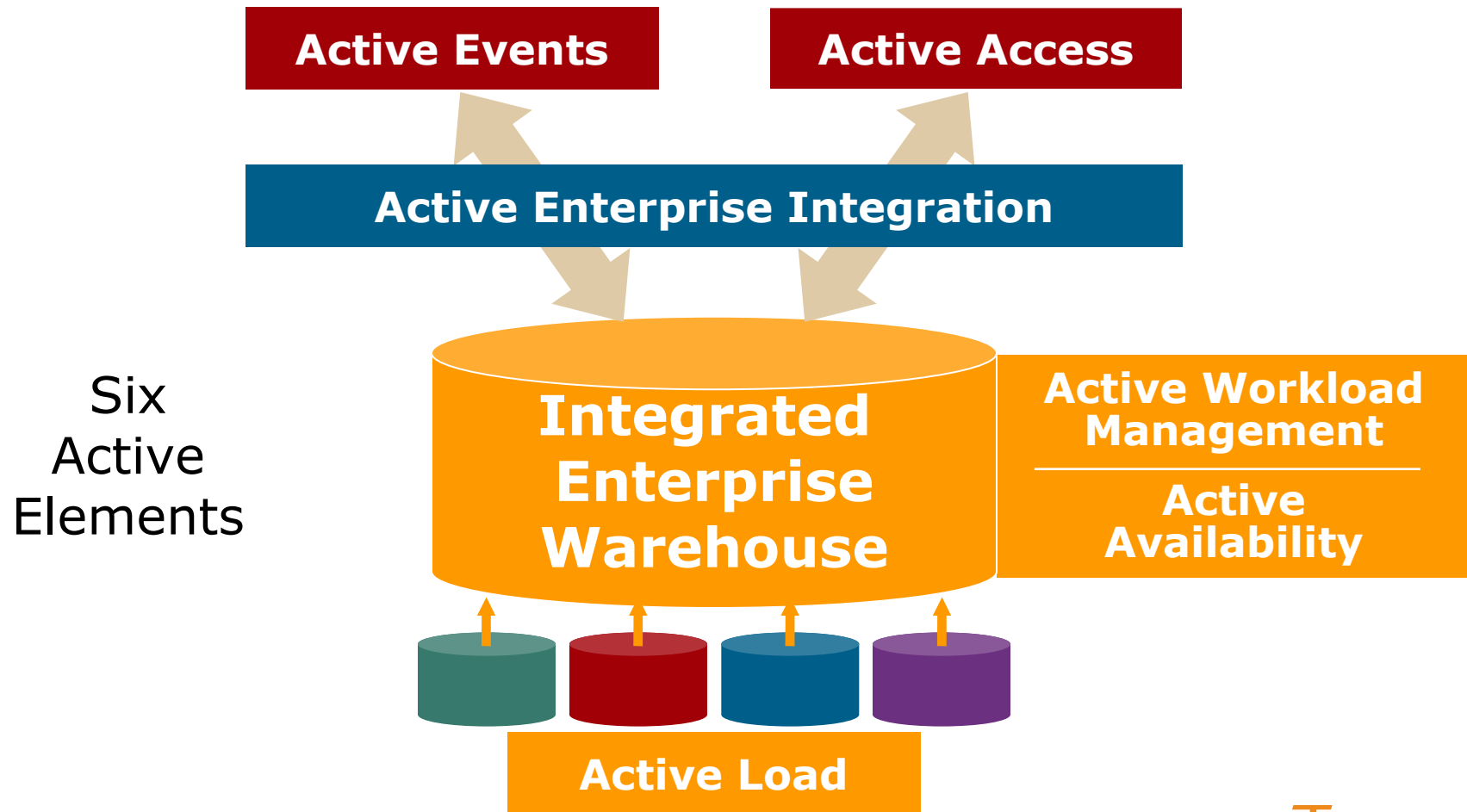
- Opportunities and threats at
 - > Strategic Level
 - > Operational Level
- Operational excellence
 - > 1000s of small daily decisions add up = reputation and profitability
 - > Aligned Insights → Right decision, just in time
 - > Acceleration → Efficiency
- Strategic excellence
 - > Foresight + planning = broad scope optimizations
 - > Insights → Right plan
 - > Speed → Time to re-strategize

Active Enterprise Intelligence



Active Enterprise Intelligence Technologies

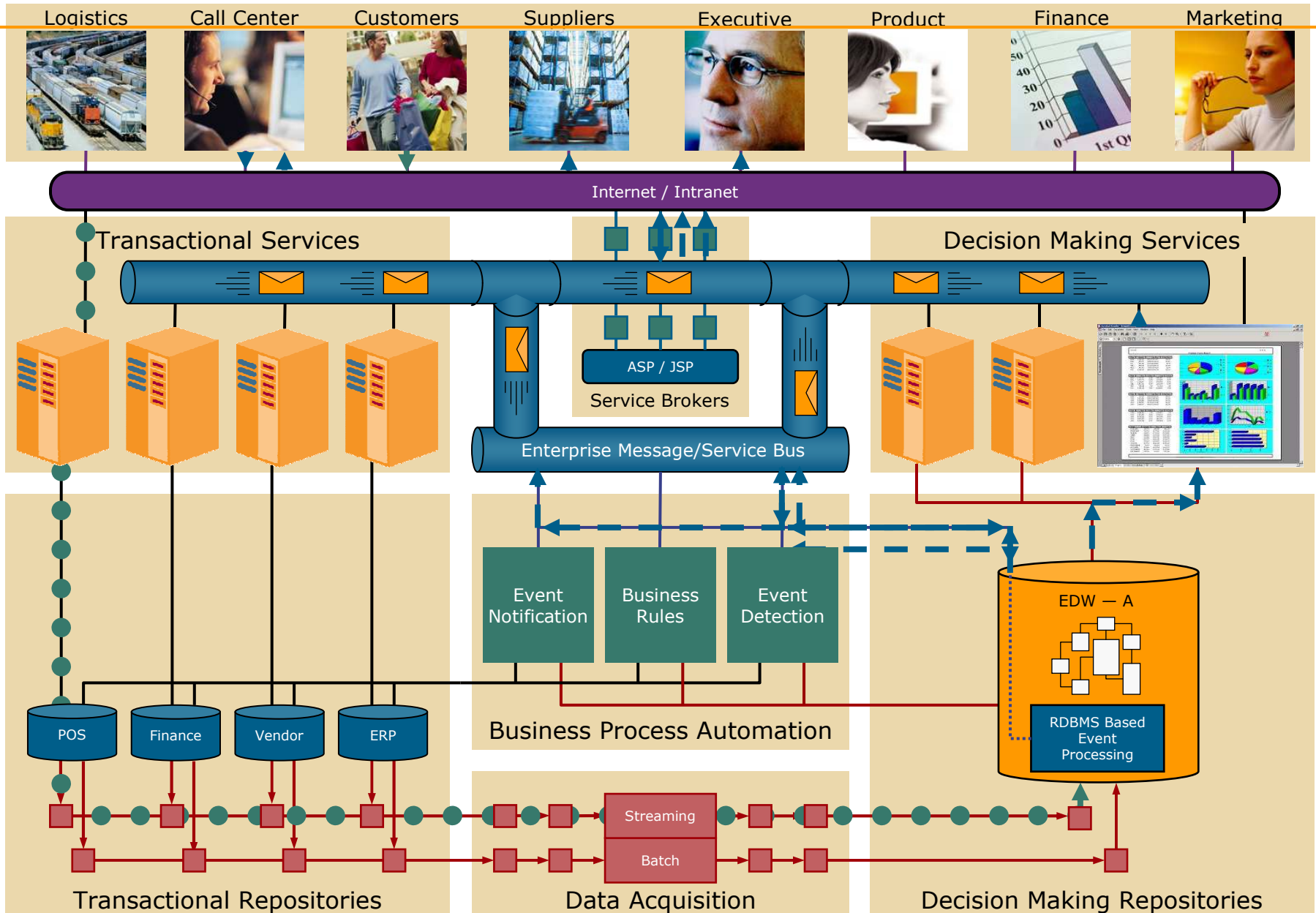
Powered by an Active Data Warehouse from Teradata



ADW – Brings It All Together

Front Line "Operational" Users

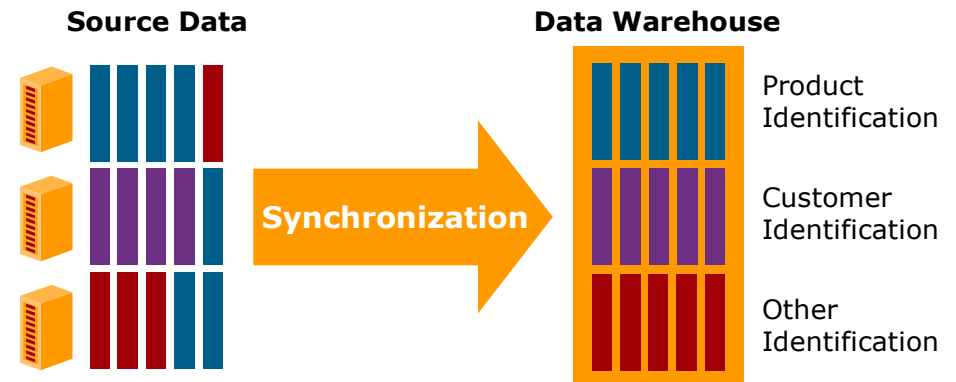
Back Office "Strategic" Users



Data Mining / Data Analytics

Data - Quest for the 360° View of Your State

- Consistent organization of business data in the data warehouse is *essential*
- This is an *ongoing process* which effective tools can both enforce and enhance



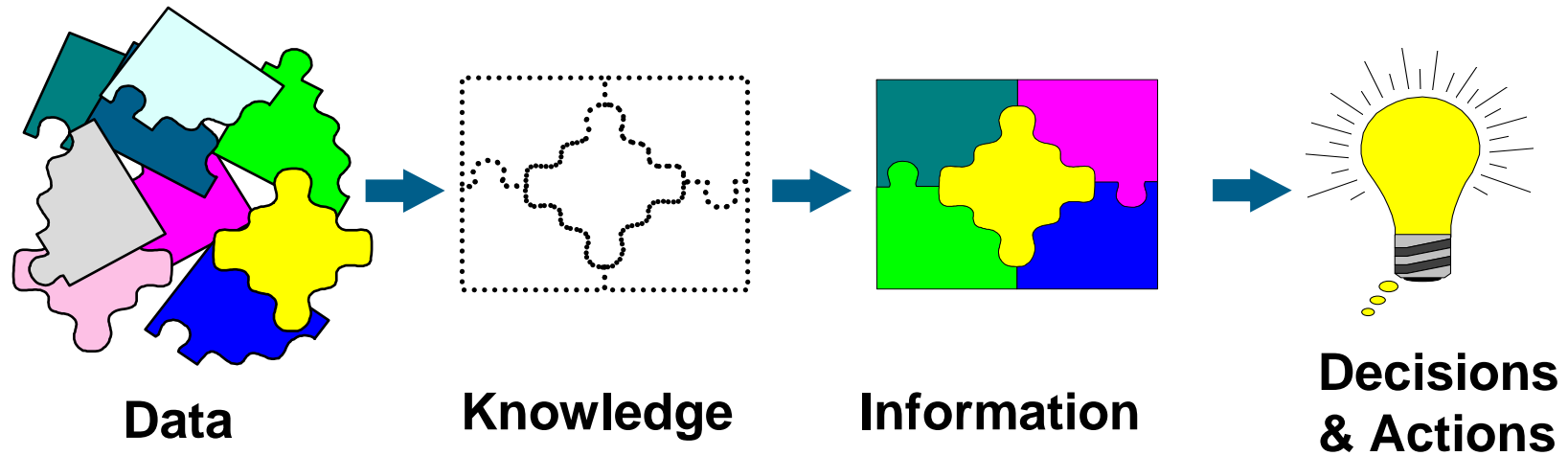
Leads to...	
Customers: Incomplete view of customer relationships	<ul style="list-style-type: none"> • Poor customer service • Ineffective cross/up-sell planning and execution
Products: Fragmented view of state of affairs	<ul style="list-style-type: none"> • Inaccurate view of campaign results • Reduced leverage in collections
All Areas: No complete view of critical areas	<ul style="list-style-type: none"> • Constant checking and correcting of data • Low confidence in strategic data » Low use/ROI

Can the IT organization respond as quickly as you want them to?

Data Mining

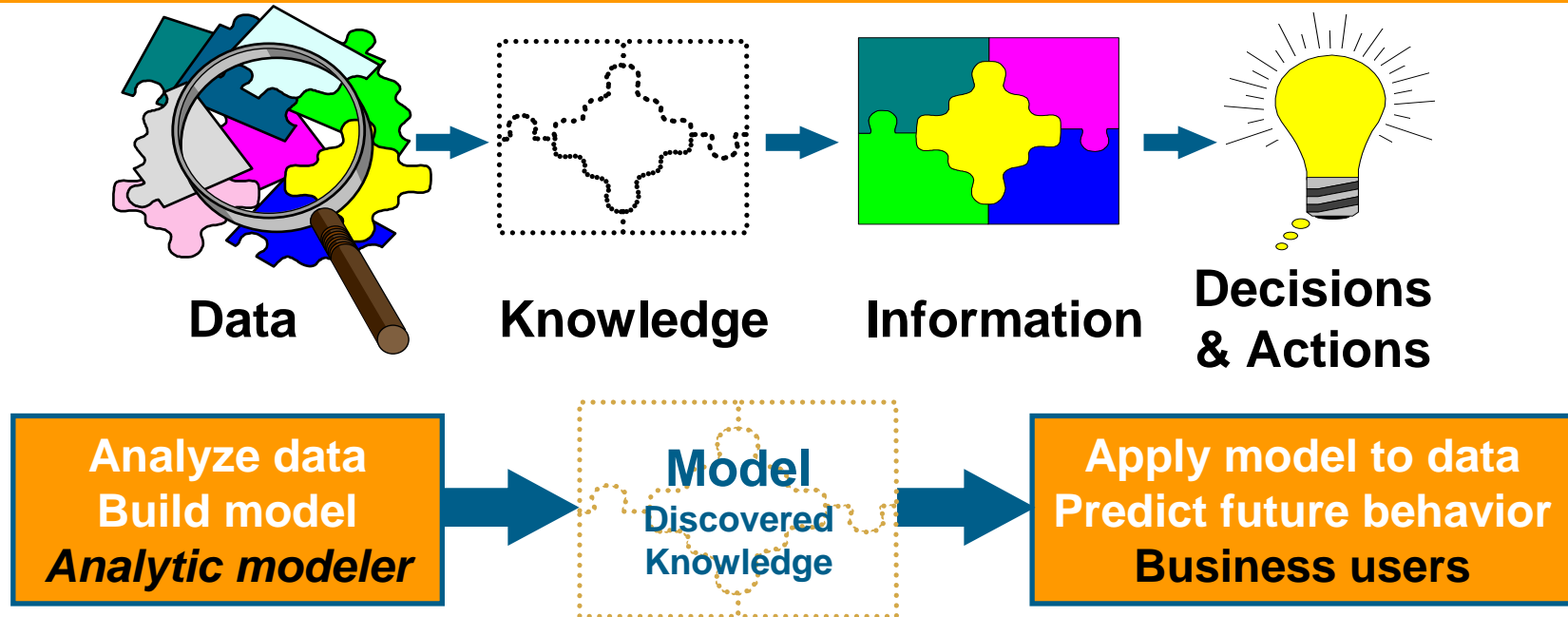
- Data mining is the process of sorting through large amounts of data and picking out relevant information. It is usually used by business intelligence organizations, and financial analysts, but is increasingly being used in the sciences to extract information from the enormous data sets generated by modern experimental and observational methods.
- It has been described as "the nontrivial extraction of implicit, previously unknown, and potentially useful information from data" and "the science of extracting useful information from large data sets or databases." Data mining in relation to enterprise resource planning is the statistical and logical analysis of large sets of transaction data, looking for patterns that can aid decision making.
- A class of analytical applications that search for patterns in a data base. Data mining is the process of sifting through large amounts of data to produce data content relationships.

What is Data Mining?



Data Mining is a process of discovering and interpreting patterns in data to solve problems

Data Mining Discovery and Deployment



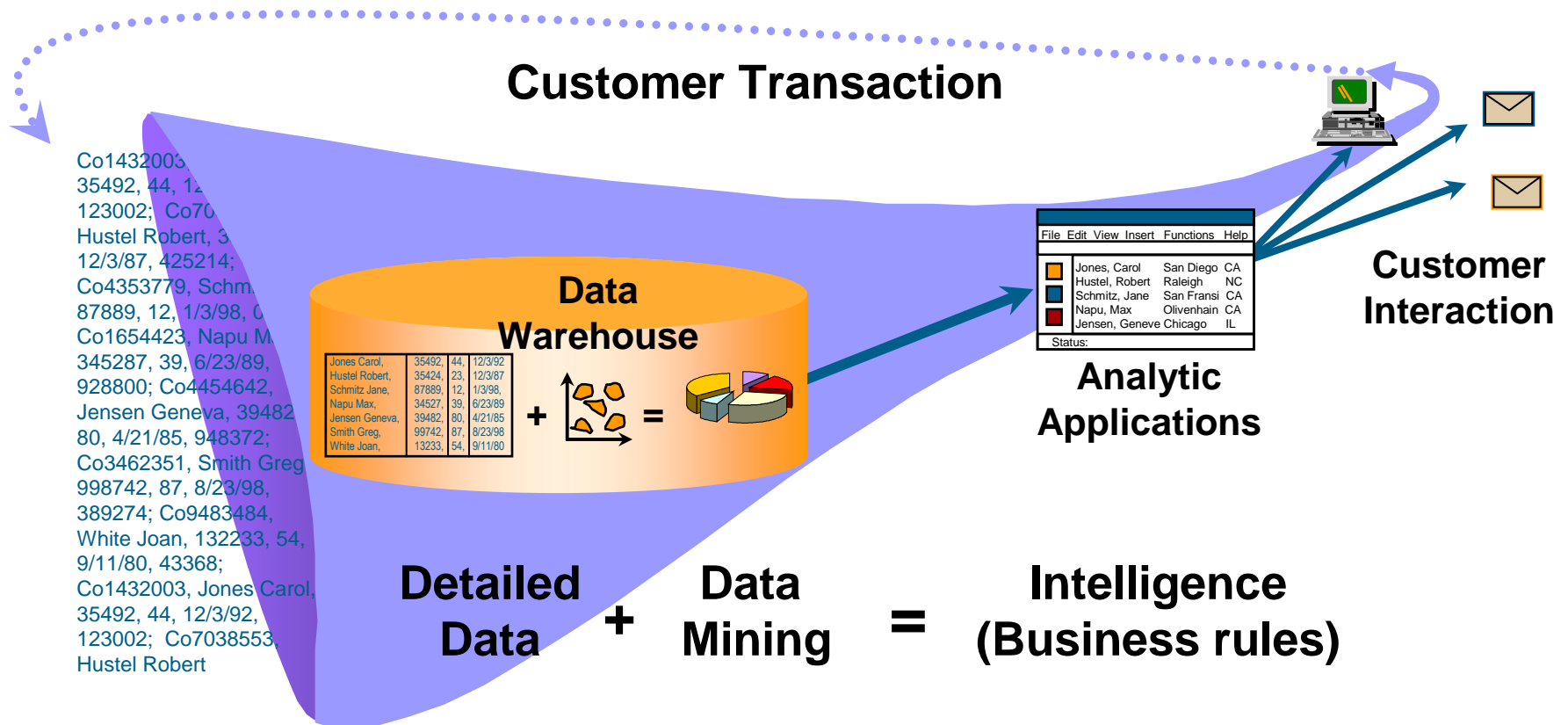
Discovery:

Finding the hidden patterns that transform data into information, typically done by statisticians or analytic modelers

Deployment:

Applying discovered knowledge for a useful purpose - e.g. prediction done by Business Users or IT organization

Data Mining Drives Intelligence into Your Analytic Applications



Analytic Applications

Typical Applications

Customer Segmentation

Propensity to Pay

Profitability Modeling & Profiling

Customer Attrition

Channel Optimization

Fraud Detection

What are my segments
and who are my customers
by segment ?



**Personalize customer relationships.
Higher satisfaction = Higher retention**

Better understand your customers' behavior, build better relationships, and retain customers

Analytic Applications

Typical Applications

Customer Segmentation

Propensity to Pay

Profitability Modeling & Profiling

Customer Attrition

Channel Optimization

Fraud Detection

Which customers are good candidates for our new long term payment play ?



**Targeting customers based on their needs.
More product sales = Greater loyalty**

Better understand your customers' behavior, build better relationships, and retain customers

Analytic Applications

Typical Applications

Customer Segmentation

Propensity to Pay

Profitability Modeling & Profiling

Customer Attrition

Channel Optimization

Fraud Detection

What is the life time
profitability of my customers ?



**Increase high value customers based
on current & future profitability.**

Better understand your customers' behavior, build
better relationships, and retain customers

Analytic Applications

Typical Applications

Customer Segmentation

Propensity to Pay

Profitability Modeling & Profiling

Customer Attrition

Channel Optimization

Fraud Detection

Which of my most valuable customers are at risk of leaving ?



**Prevent loss of high value customers
and let go of lower value customers.**

Better understand your customers' behavior, build better relationships, and retain customers

Analytic Applications

Typical Applications

Customer Segmentation

Propensity to Pay

Profitability Modeling & Profiling

Customer Attrition

Channel Optimization

Fraud Detection

What is the best channel to reach my customers in each market segment?



Interact with customers based on their preference.

Better understand your customers' behavior, build better relationships, and retain customers

Analytic Applications

Typical Applications

Customer Segmentation

Propensity to Pay

Profitability Modeling & Profiling

Channel Optimization

Channel Optimization

Fraud Detection

How can I tell which transactions are likely to be fraudulent ?



Detect and prevent fraud to minimize loss.

Better understand your customers' behavior, build better relationships, and retain customers

Application of Data Mining for a Financial Institution: Business Problem

- Business Problem:
 - > \$10 Million in Tax Fees are not being collected
- Goal:
 - > Identify customers who do not pay



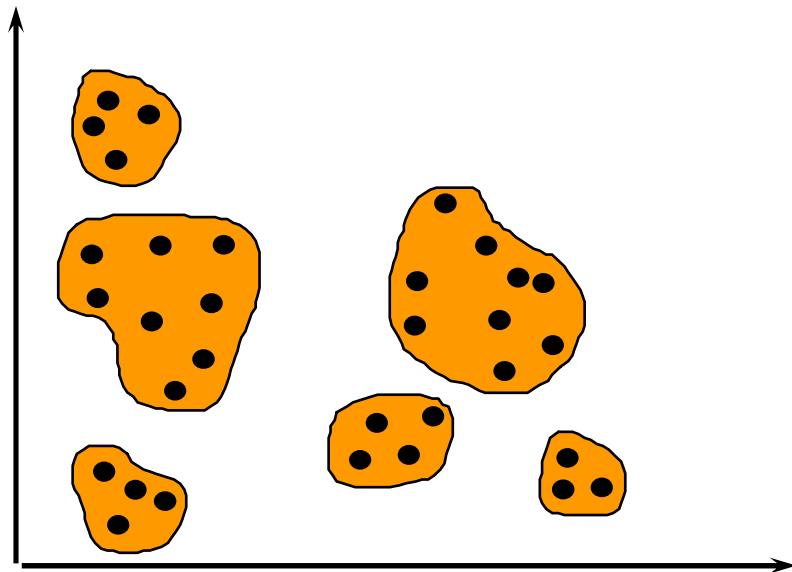
Application of Data Mining for a Financial Institution: Model Development

- Data Preparation

- > Gather detailed Data
- > Data Exploration
- > Create Analytic Data (assessed tax, fee/customer of a given period, % paid relative to billed amount, etc.)

- Model Development

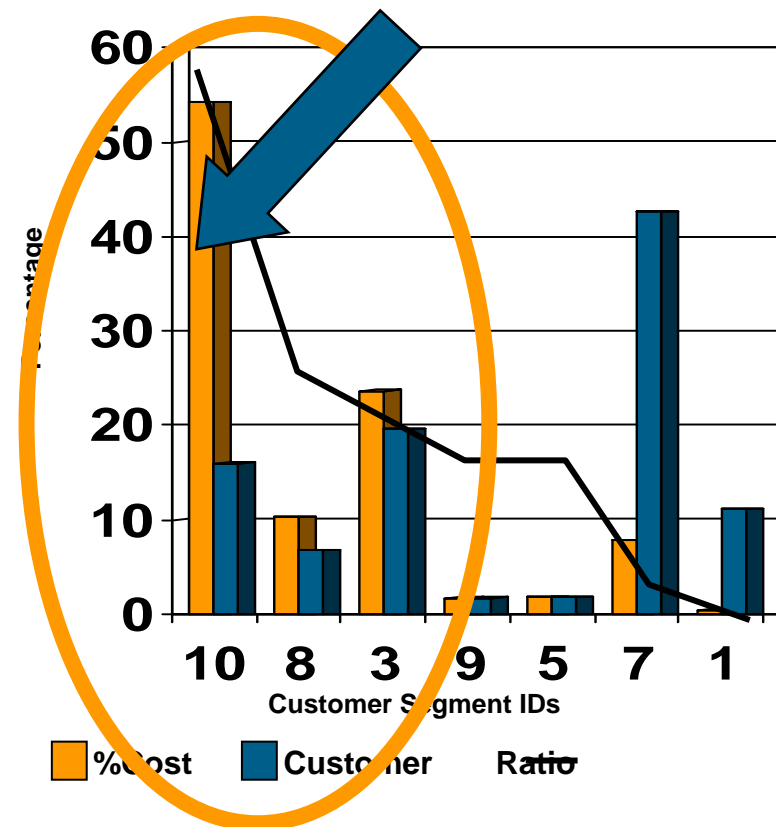
- > Factor Analysis
(reduce variables)
- > Clustering
- > Analysis



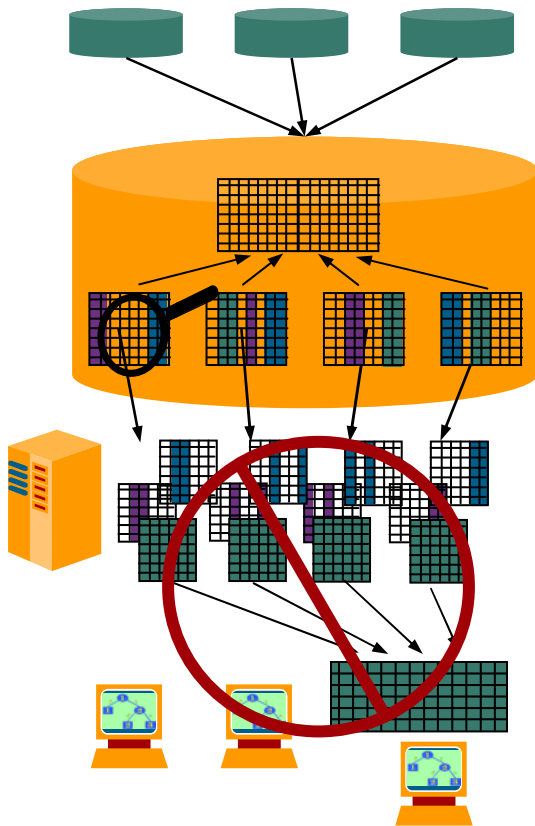
Application of Data Mining for a Financial Institution: Results

Collect Millions \$

- Collect taxes
- Collect fees
- Collect fines
- New Intelligence
 - > Tax fraud
 - > Unpaid child support/alimony
 - > Unpaid DMV
 - > Healthcare fraud
- Garnish wages
 - > Become paying citizens

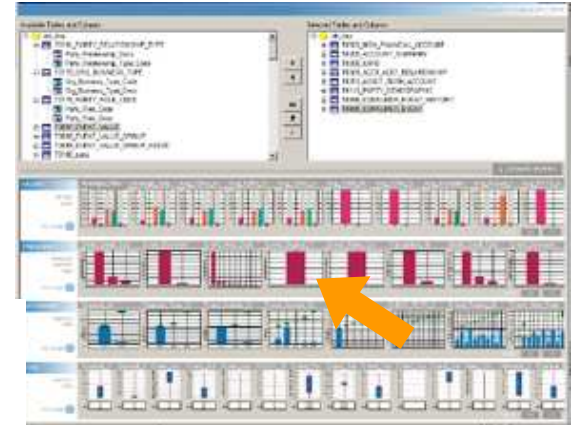


Understand Your Data



Data Exploration with Server-based Tools

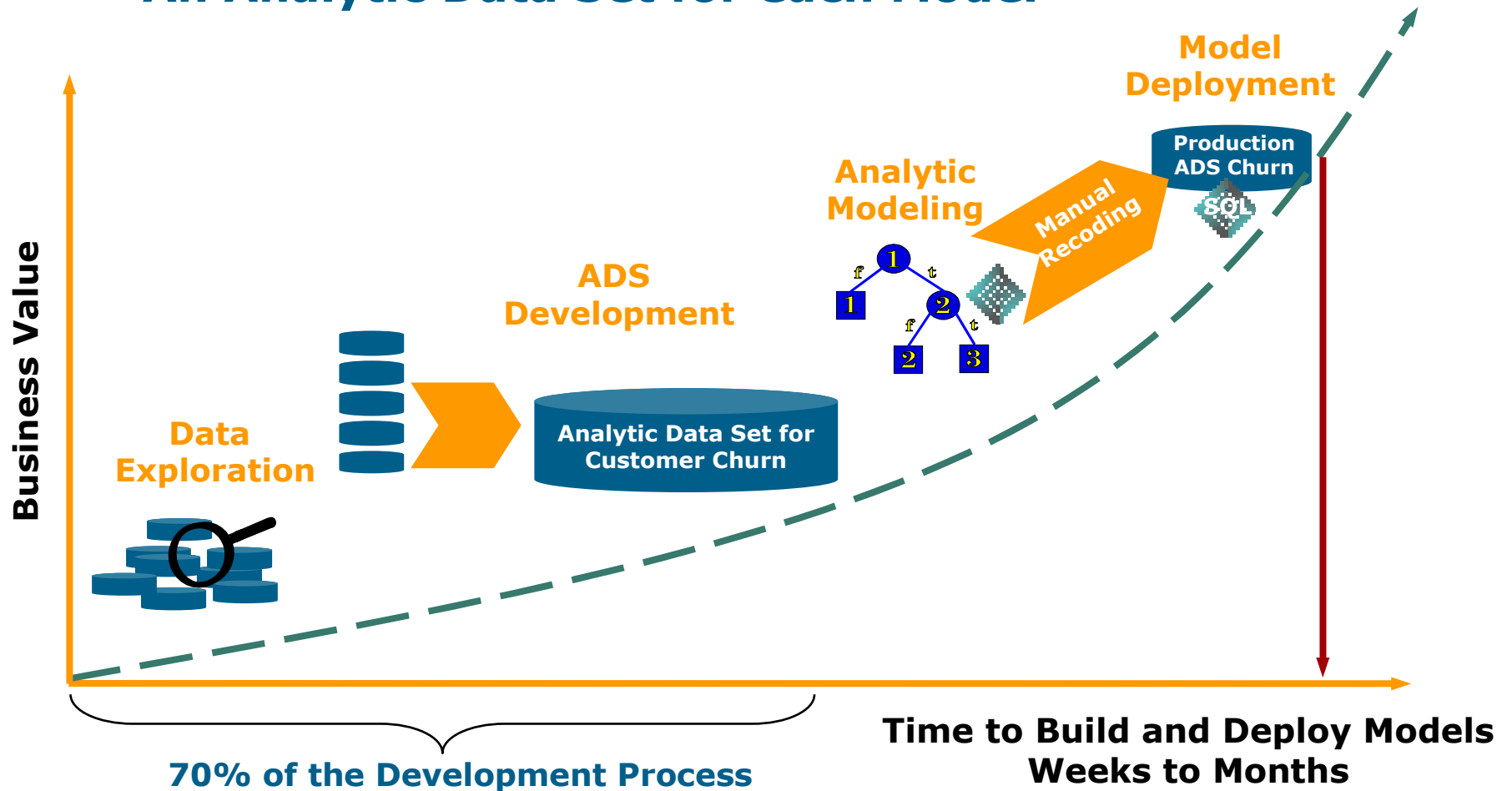
- Know your data prior to extracts
 - > Explore and identify data relevant to your application
 - > Avoid pulling irrelevant tables and columns
- Analyze large and complex tables
 - > Avoid pulling out all data and joining on the server
 - > Join then analyze data directly in the database
- Drill down to identify questionable data records
 - > Click on graphs to drill down on specific data records
 - > Take actions on those records



list	drilldown								
cust_id	income	age	years_with_us	nbr_children	gender	marital_status	name_prefix	first_name	
1362480	50890	33	3	2	M	2	Mr.	Hans	
1362481	20955	36	6	2	F	2		Darby	
1362484	10053	42	2	0	F	1	Dr.	Lana	
1362485	22690	25	4	0	F	1	Dr.	Elvira	
1362486	10701	76	6	0	F	3	Ms.	Kyle	
1362487	6605	71	1	0	M	2		ChengDye	
1362488	7063	77	7	0	F	2	Ms.	Stephane	
1362489	55888	35	5	2	F	3		Inge	
1362492	40252	40	0	5	F	3		Maggie	
1362496	0	13	2	0	M	1		Douglas	
1362497	0	14	6	0	M	1		Thomas	
1362498	9849	60	0	0	F	2		Rhonda	
1362499	37180	65	5	0	M	2		Lawrence	
1362500	56708	40	0	0	M	1		Dan	
1362501	65347	58	8	2	F	2		Rose	
1362502	19530	57	7	2	M	2		Gilbert	
1362503	5787	18	3	0	M	1		Cary	
1362504	0	15	0	0	F	1		Tammy	

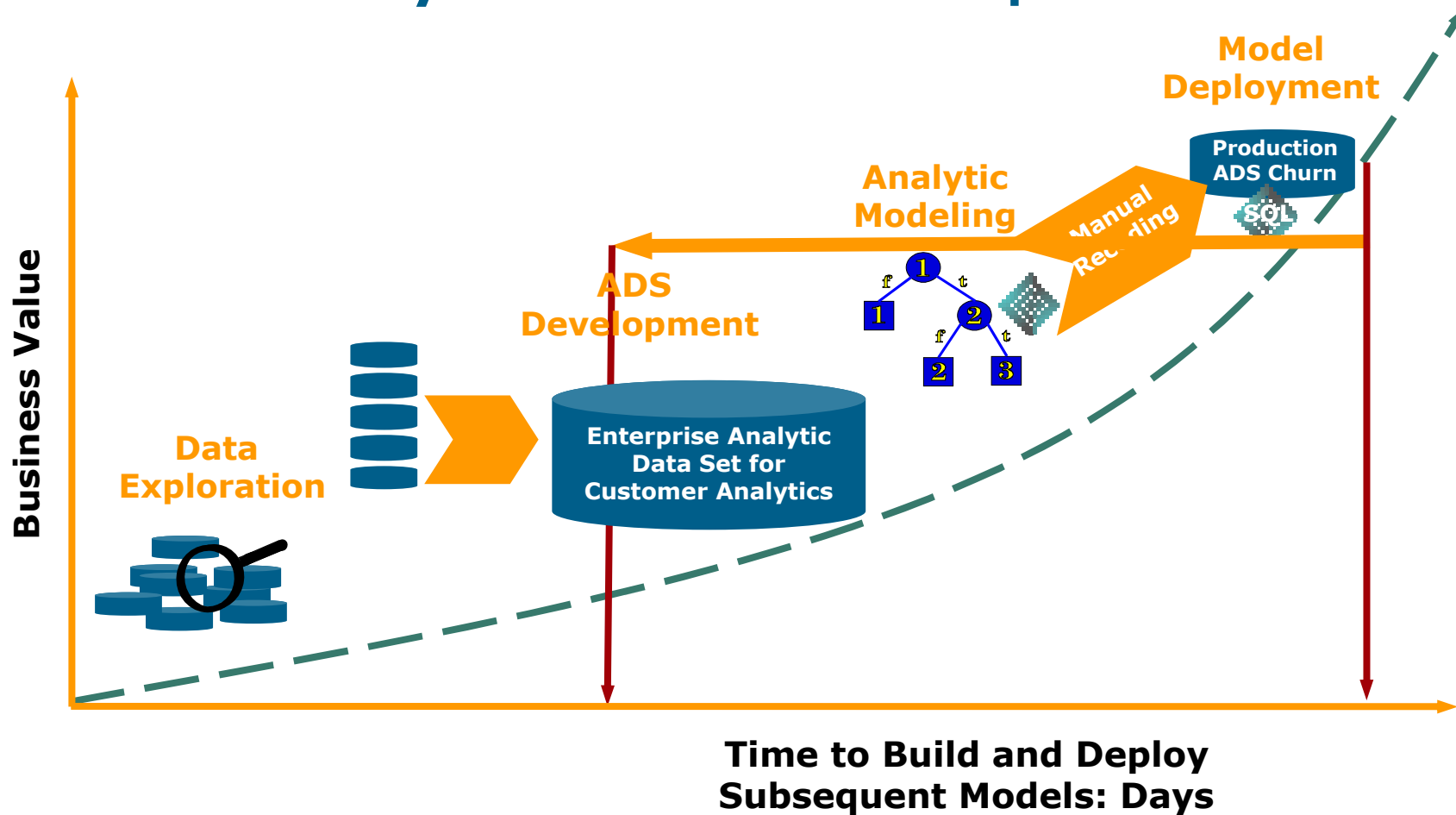
Typical Modeling Process

An Analytic Data Set for each Model



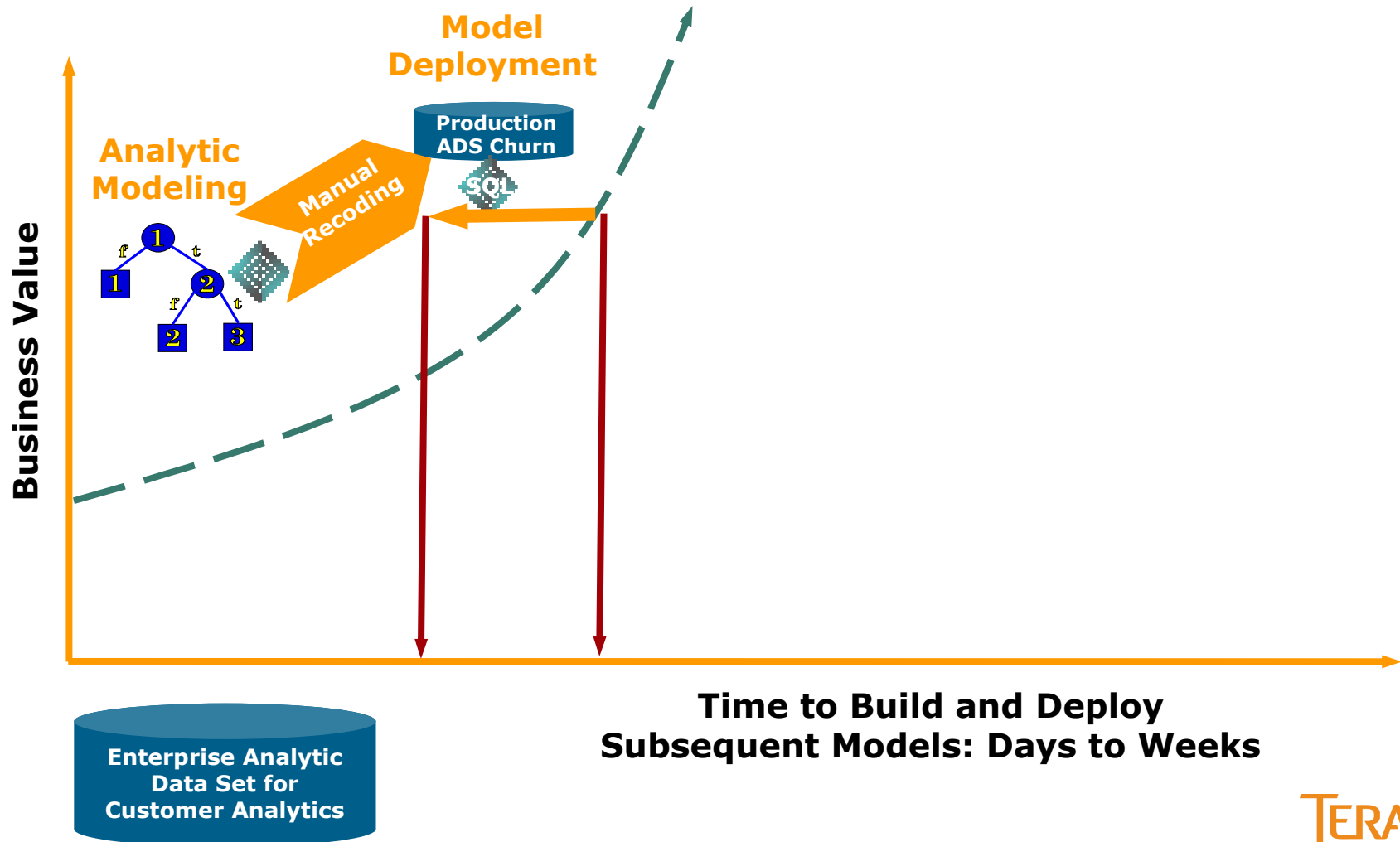
Increase Model Production

Reusable Analytic Data Set for Multiple Models



Shorten Time to Results

Automate Deployment



Data Visualization and Managing Data

Data Visualization - From Wikipedia

- The main goal of data visualization is to **communicate information clearly and effectively** through graphical means. It doesn't mean that data visualization needs to look boring to be functional or extremely sophisticated to look beautiful. To convey ideas effectively, both aesthetic form and functionality need to go hand in hand, providing insights into a rather sparse and complex data set by communicating its key-aspects in a more intuitive way. **Yet designers often tend to discard the balance between design and function, creating gorgeous data visualizations which fail to serve its main purpose — communicate information.**
- Data visualization is closely related to Information graphics, Information visualization, Scientific visualization and Statistical graphics. According to Frits Post (2003) data visualization is currently a very active and vital area of research, teaching and development. The term unites the established field of scientific visualization and the more recent field of information visualization

Delivering a 360° View of the Business

- Requires consistent *ongoing* identification and organization of master data. Traditional approaches are slow to address changing business needs, with serious impacts:
 - > Customer service
 - > Marketing campaign execution
 - > Time spent correcting data
 - > Cross/up-sell planning and execution
 - > Leverage with suppliers
 - > Satisfaction with business intelligence ROI
- Most master data values and hierarchies are ultimately defined by the business. Many required changes cannot be anticipated far in advance.

“Management of master data is the most frequent source of inconsistency and organizational conflict in data warehousing.”

– Forrester Research, J. Paul Kirby, Jan. 2007



Decisions Drive the Need for Analytics

- Decisions are the lifeblood of an organization
- Every person at every level in an organization makes multiple decisions every day
- Companies recognize the need to provide employees with information to aid in making those decisions
- Dedicated analytical environments have been around for more than 25 years



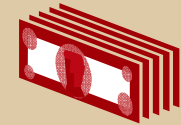
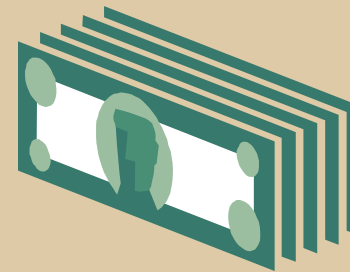
Options for Enterprise Analytics

Integrated, Enterprise Data Warehouse

Customer Name	Customer #	Address	List Price
Country	Region	Item #	Transaction #
Revenue	Margin	Vendor	Cost

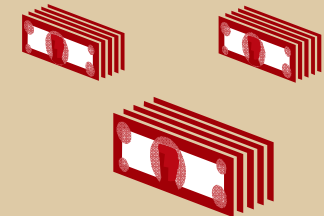
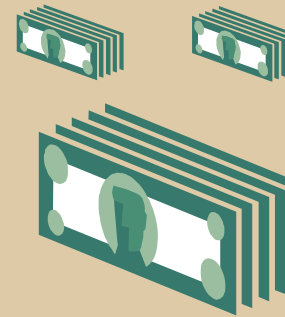
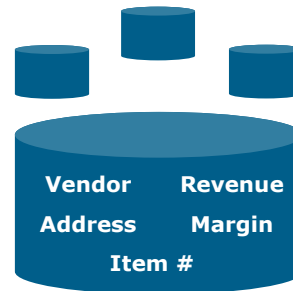
Business value

Cost to deliver

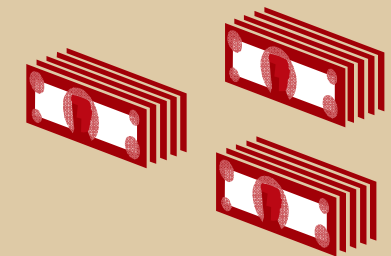
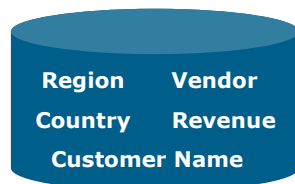


Some Data Integration

Customer Name	Customer #	Address
Country	Region	Item #
Revenue	Margin	



No Data Integration (Data Mart Strategy)

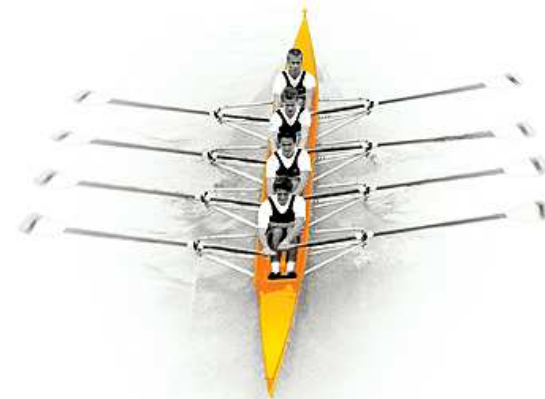


Data Integration Continuum

Attribute	No Data Integration (Data Mart Approach)	Some Data Integration (Some Data Marts, Some Integration)	Integrated Foundation – EDW, Logically Modeled
Data Duplication / Redundancy	High, across data marts	Medium, lessens as new projects use the integrated foundation	Low, source data once, use many times
Architecture Complexity and TCO	High, no data reuse	Medium, some data reuse	Low, data reuse
Development Process	Modeling, high Data sourcing, high BI development, high	Modeling, Med Data Sourcing, Med BI Development, High	Modeling, Low Data Sourcing, Low BI Development, Med
Time to Value: Incremental Applications	High, full development cycle for each application	Medium, some reuse occurs	Low, a shared utility, little modeling or ETL required
Ability to Innovate, Competitive Advantage	Low, minimal new insight, point solution	Medium, some new insight opportunities	High, game changing opportunities, enterprise leverage

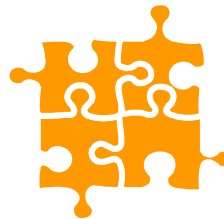
Benefits of Integrated Data

- **Cost Savings**
 - > Disk space
 - > Underutilized servers
 - > ETL effort (people, time, network)
 - > Modeling
 - > Maintenance (passwords, access rights, BAR)
 - > Data reconciliation
- **Efficiency/Optimization**
 - > Compliance
 - > Business Continuity Planning
 - > Security and Privacy
 - > Master Data Management
 - > Power, cooling, floor space
 - > Performance and Capacity planning
- **Business Opportunity**
 - > New business insights
 - > Manage business as whole
 - > Ask any question at any level
 - > Fast time to solution
 - > Fast response to market changes and competition
 - > Rapid time to decision



Make Informed Decisions

- Decisions are made based upon available data



All the right data at the right time

Industry Practices in Business Intelligence

Business Intelligence

Business Intelligence

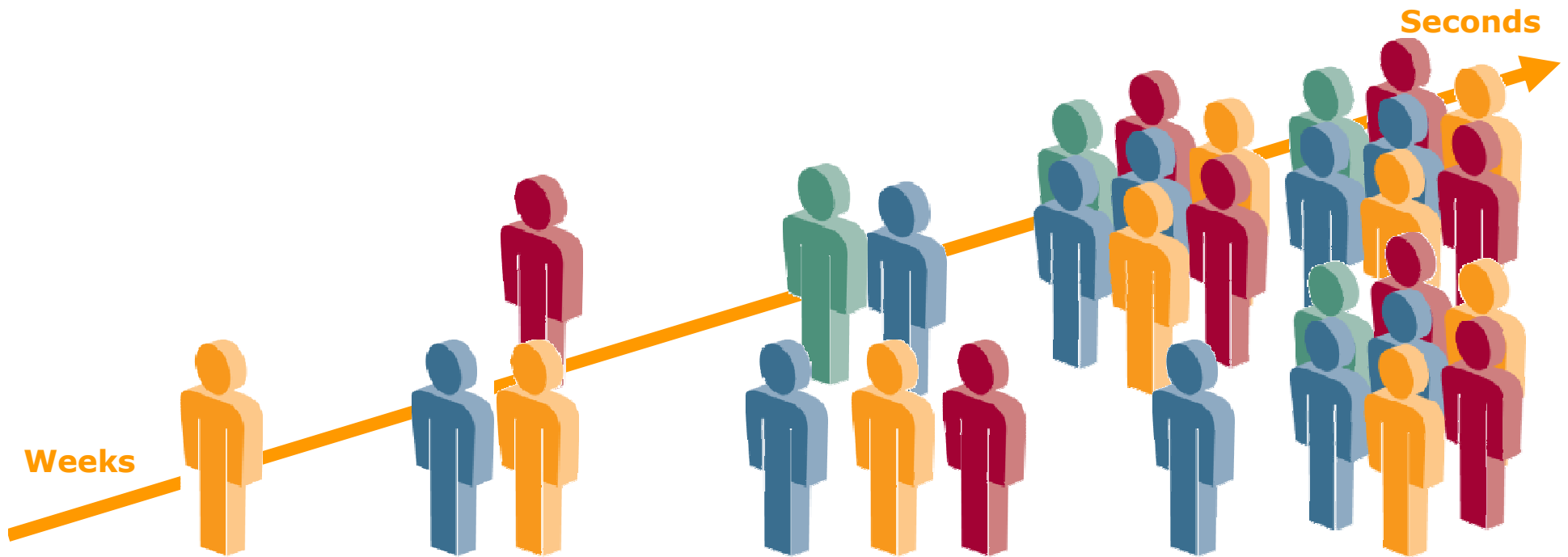
- Skills, knowledge, technologies, applications, quality, risks, security issues and practices used to help a business to acquire a better understanding of market behavior and commercial context.
- It undertakes the collection, integration, analysis, interpretation and presentation of business information.
- “Business intelligence” may refer to the collected information itself or the explicit knowledge developed from the information.
- BI applications provide historical, current, and predictive views of business operations, most often using data already gathered into a data warehouse or a data mart and occasionally working from operational data.

Business Intelligence

- Software elements support the use of this information by assisting in the extraction, analysis, and reporting of information.
- Common functionality of business intelligence applications includes reporting, OLAP, analytics, dashboards, scorecards, data mining, corporate performance management (CPM), and predictive analysis.
- Business intelligence — the term dates at least to 1958 — aims to support better business decision-making.
- One can also characterize a BI system as a decision support system (DSS): BI is sometimes used interchangeably with briefing books, report and query tools and executive information systems. In general, business intelligence systems are **data-driven** DSS.

Active Enterprise Intelligence

An Obvious Trend: More Speed, More Users



Strategic Intelligence

Enterprise Data Warehouse
 BI Tools & reports
 Analysis & visualization
 Predictive Analytics

Operational Intelligence

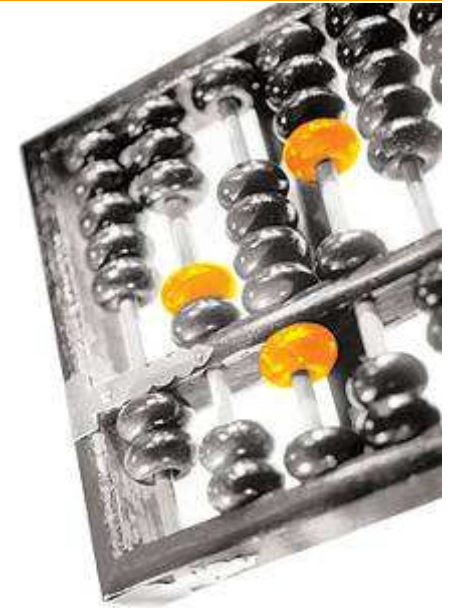
EDW Enterprise Integration
 Mixed workload management
 SOA, BPMS, IDEs
 Portals/composite applications

“Little Decisions Add Up”

“We judge leaders by how well they make **big, strategic decisions**. But corporate success also depends on how well rank-and-file employees make **thousands of small decisions**.

- > Do I give this client a special offer?
- > How do I handle this customer’s complaint?
- > Should I offer a long term payment to this customer?

By themselves, such daily calls – increasingly made with the help of enterprise decision-management technology – have little impact on business performance. **Taken together, they influence everything from profitability to reputation.**”



Active Enterprise Intelligence

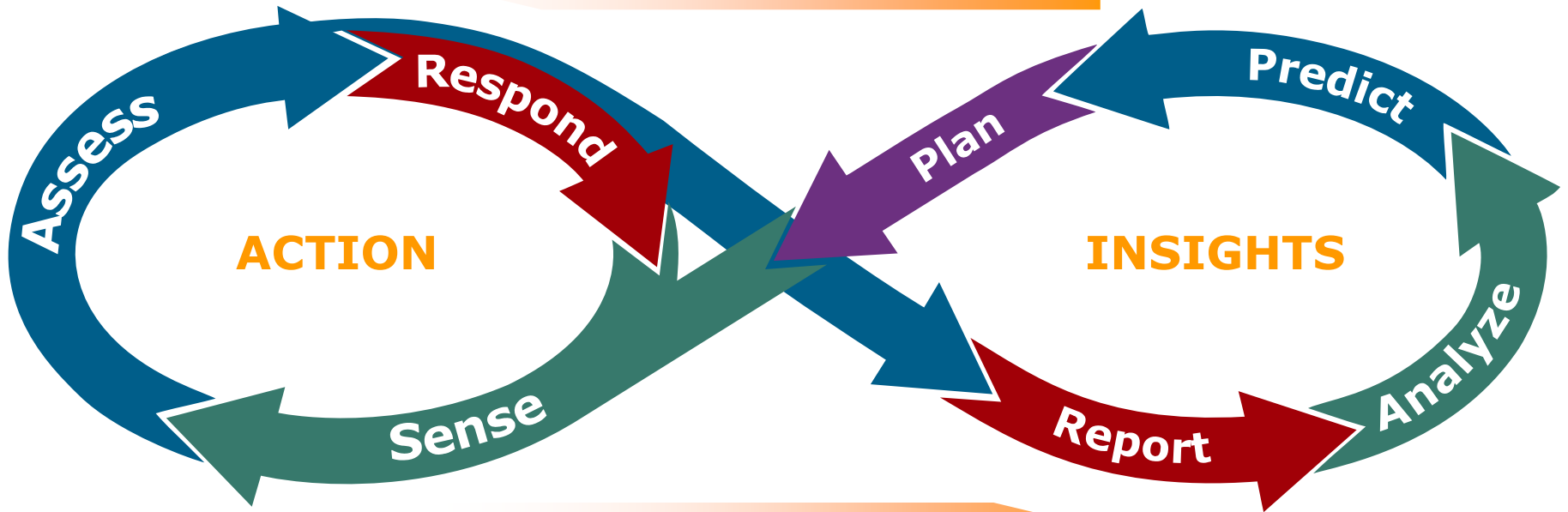
OPERATIONAL INTELLIGENCE



STRATEGIC INTELLIGENCE



Intelligence





TERADATA

Raising Intelligence