



Data Warehousing & Business Intelligence Seminar Series

Michael Haisten

VP – Business Intelligence

Norman Comstock

Director – OLAP Solutions

“A new factor, that of rapid change, has come into the world. We have not yet learned how to adjust ourselves to its economic and social consequences.”

Harvard Business Review
October 1932



What Is Business Intelligence?

A broad category of ***applications*** and ***technologies*** for ***gathering, storing, accessing, and analyzing*** data to help ***enterprise users*** make better informed ***business decisions***.

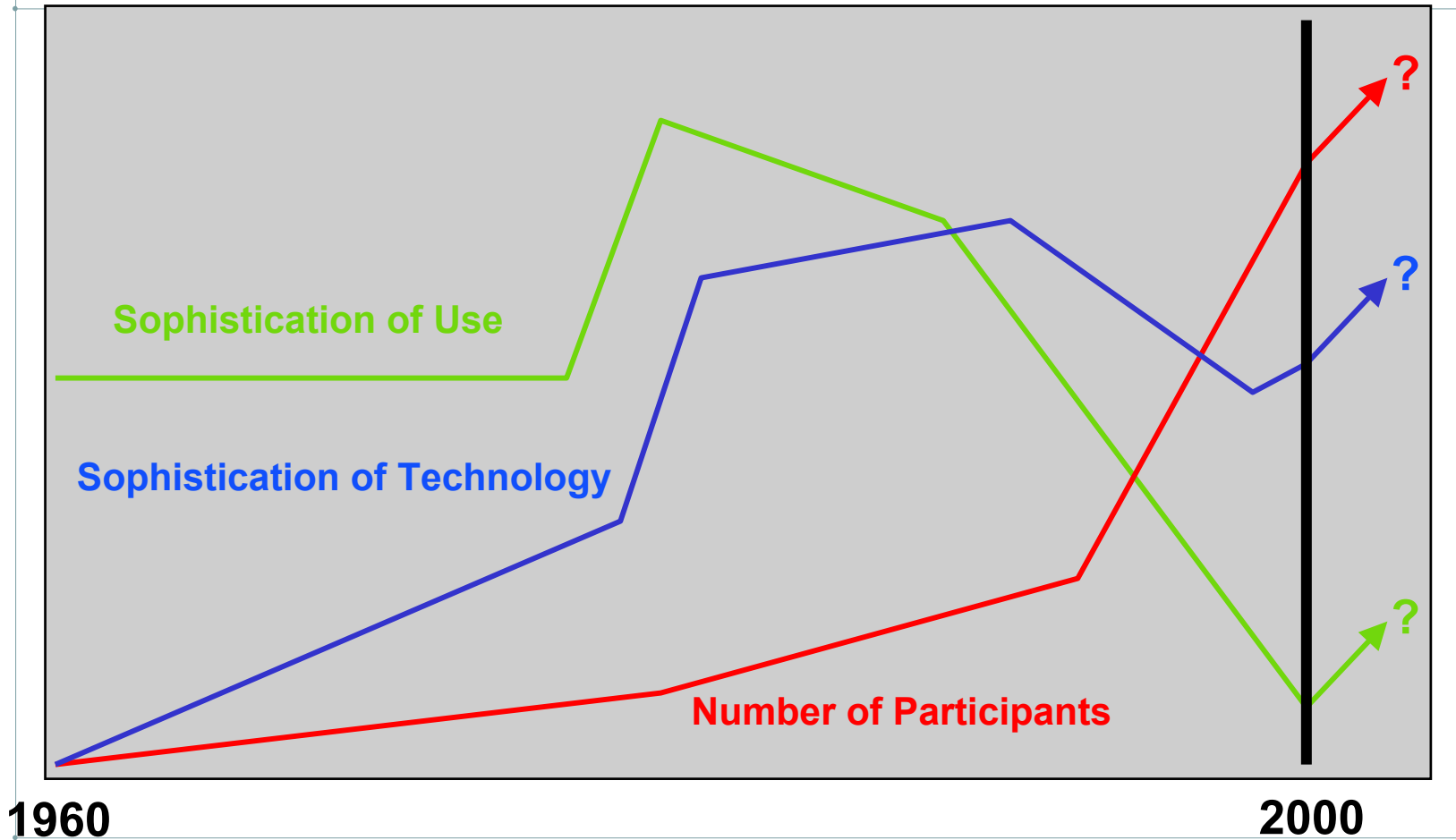


Business Intelligence Capabilities

- ❖ **Query and Reporting**
- ❖ **Decision Support Systems (DSS)**
- ❖ **Executive Information Systems (EIS)**
- ❖ **Enterprise Information Portals (EIP)**
- ❖ **Statistical Analysis**
- ❖ **Trending, Projections, and Forecasting**
- ❖ **On-Line Analytical Processing (OLAP)**
- ❖ **Pattern Identification & Conjecture (Data Mining)**



Business Intelligence Trends



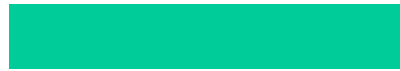
1960

2000



Sophistication of Use

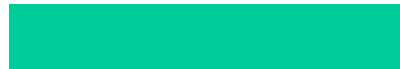
Control



Operational



Management



Analysis



?



Timeline



Analytical Application Trends

- **Underlying OLAP engines continue to become more robust in order to meet scalability demands.**
- **The cost of OLAP and complementary presentation software continues to slide as the technology improves.**
- **Web-based presentation is making analytical applications more accessible with a lower interpretation hurdle.**
- **Data mining is becoming more mainstream.**
- **Ubiquity of analytics is forcing integration of BI capabilities**
- **Closed Loop Analytics is now an essential capability.**



Data Warehousing – What's Next?

Michael Haisten

VP – Business Intelligence



Data Warehouse: What's up?

Data Warehousing is Mature.

- Everyone already has one.
- Everyone knows how to build one.

The Data Warehouse is Dead.

- Data warehousing failed.
- Not the concept but the execution.
- The successes are spectacular.
- The concept is as sound as ever.

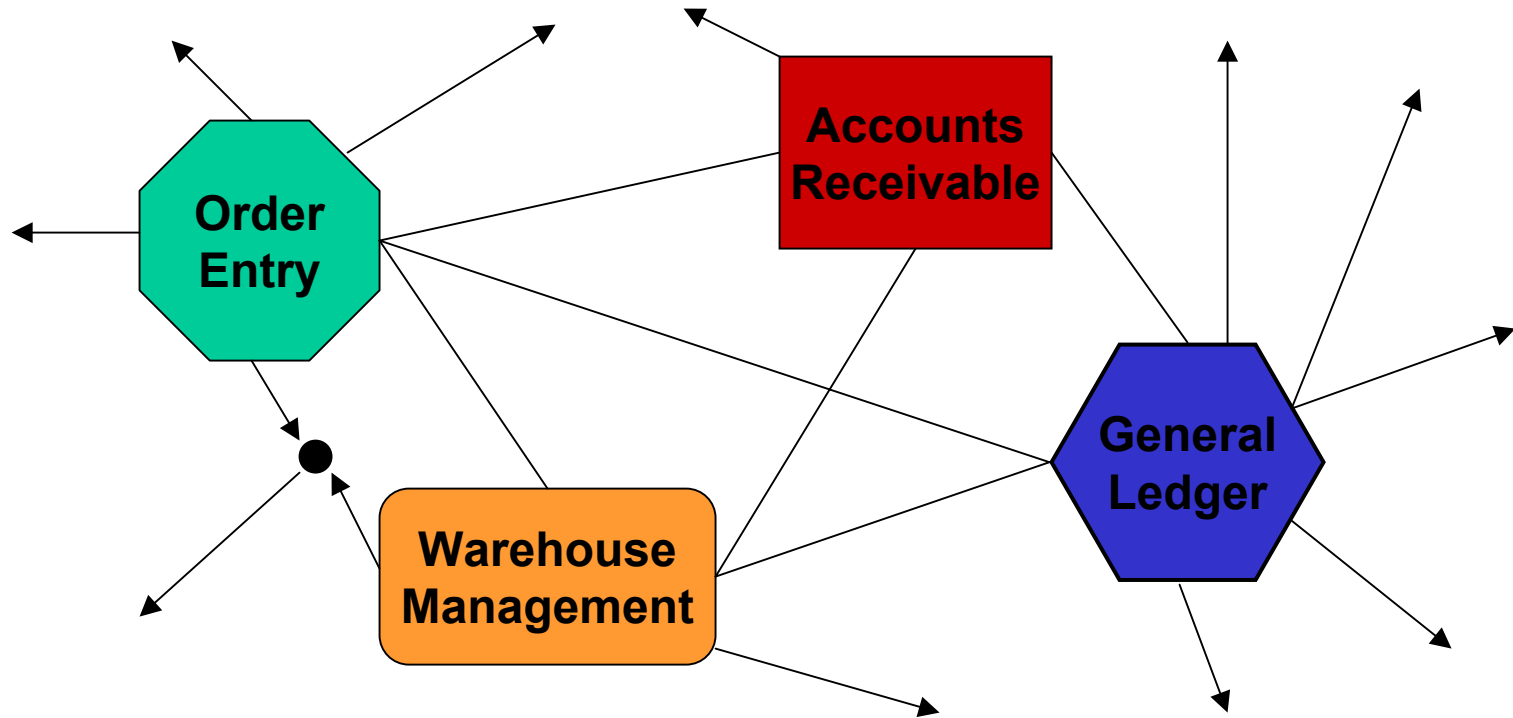


Data Warehousing: What's Next?

- ✓ **Integrate the “Data Warehouses”**
- ✓ **Expand Use of the Existing Resource**
- ✓ **Compressing the Information Supply Chain**
- ✓ **Increase Architectural Flexibility**
- ✓ **Rev-up Availability - RTDW (plus lower the cost & complexity)**
- ✓ **Close the Loop (directed operational feedback)**
- ✓ **Advance the Analytic Power**

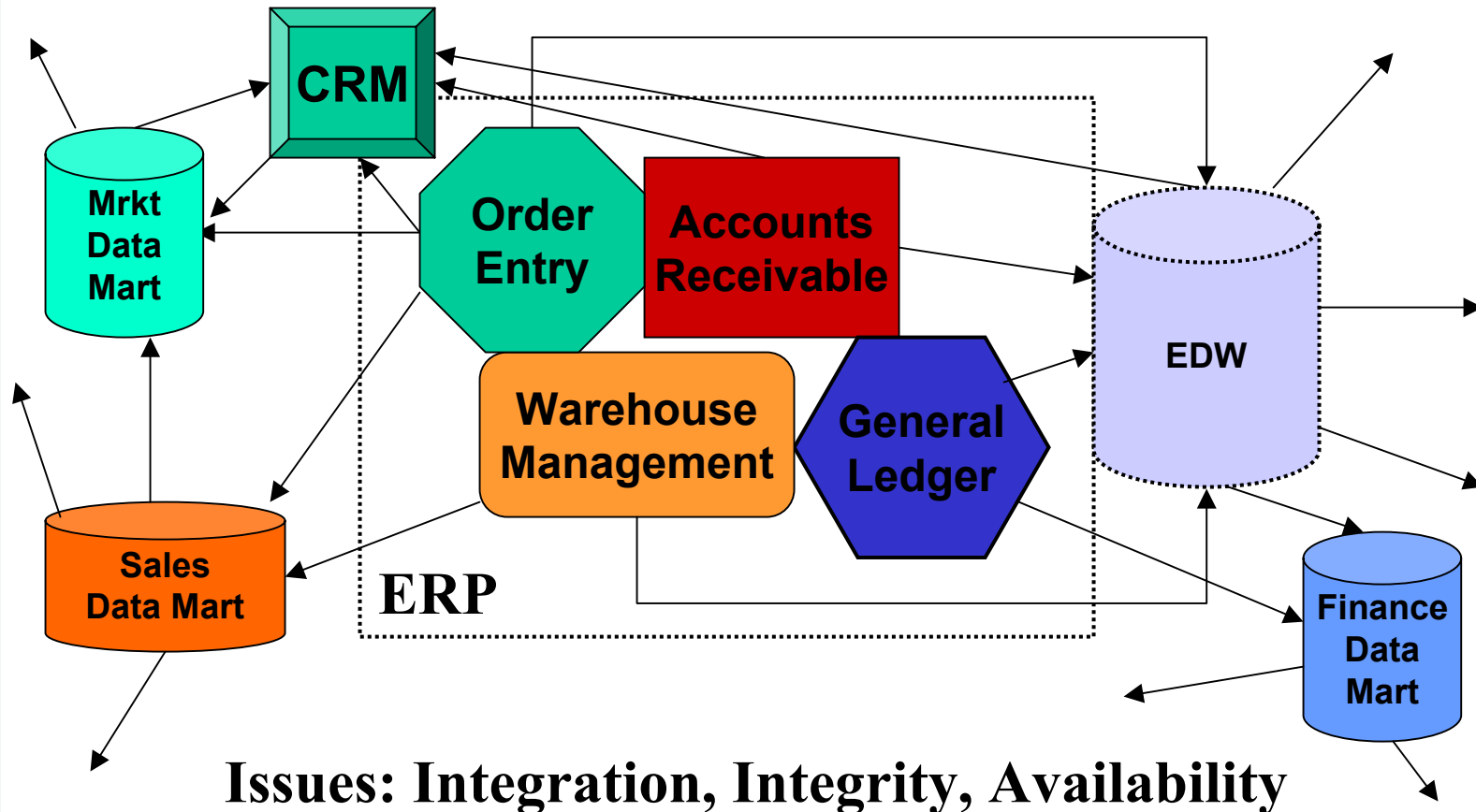


Islands of Information - 1991



Issues: Integration, Integrity, Availability

Islands of Information - 2001



Integration & Integrity – 2001 Forward

Abandon Backend Integration (DB to DB, APP to APP)

Create “Enterprise Back Plane” [Ralph Kimball]

- Concentrate on common dimensions (reference data)
- Collect transaction stream data (event detail)

Integrate in the Middle

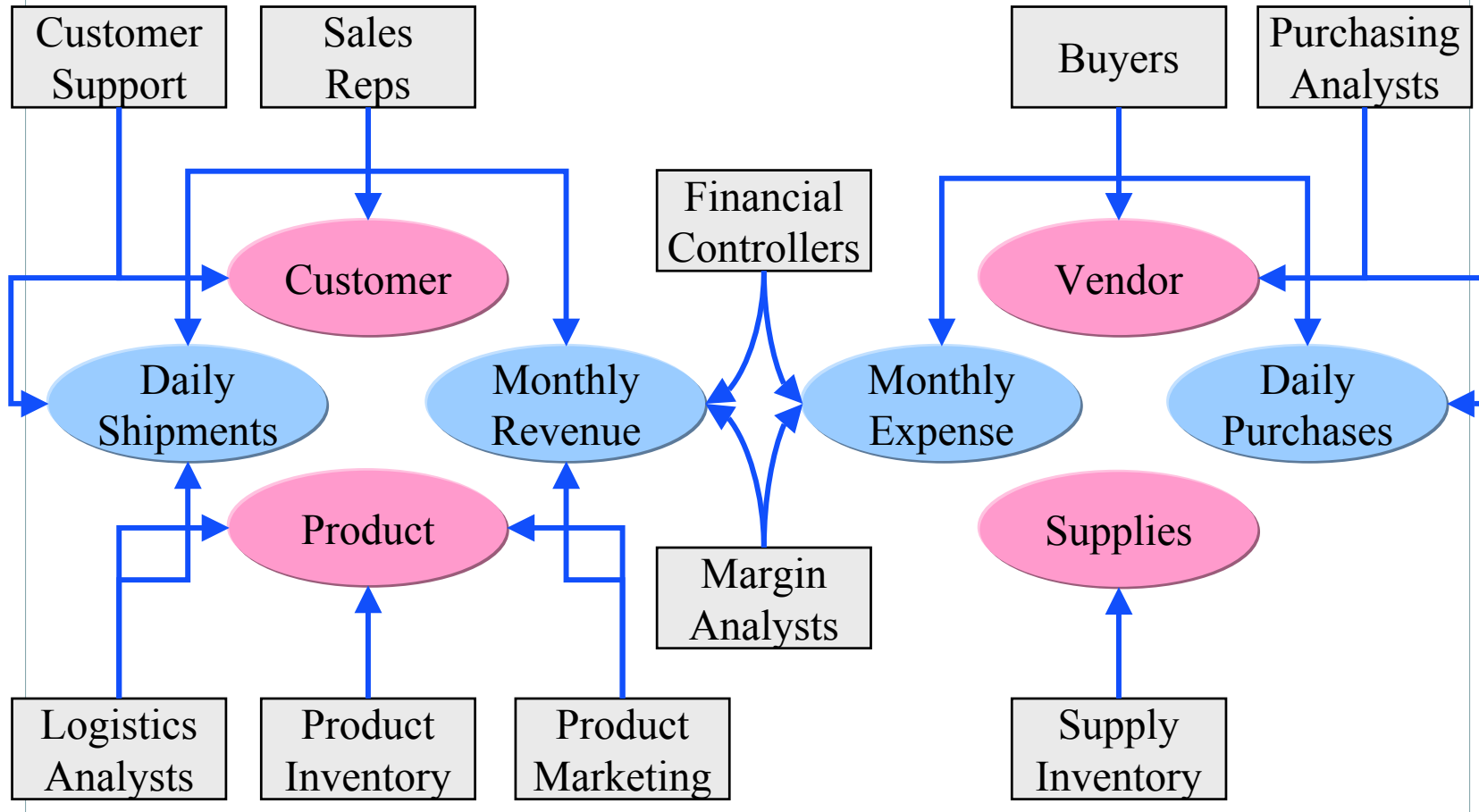
- Real – Message-based Transport, Mapping, and Transformation
- Virtual – Run-time translation into transient or persistent stores

Beware Frontend Integration



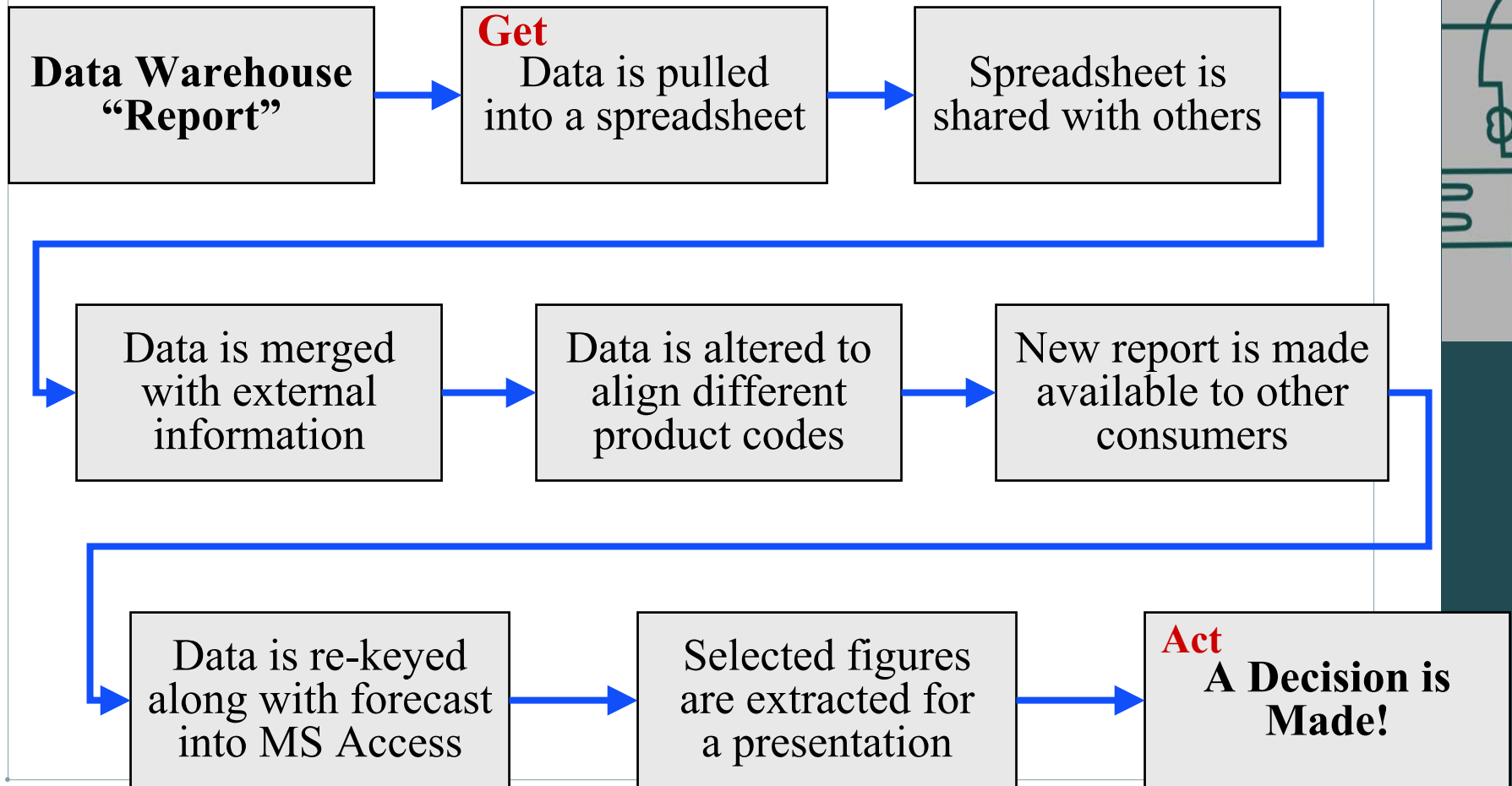
Expand the Use

Leverage Existing Investment

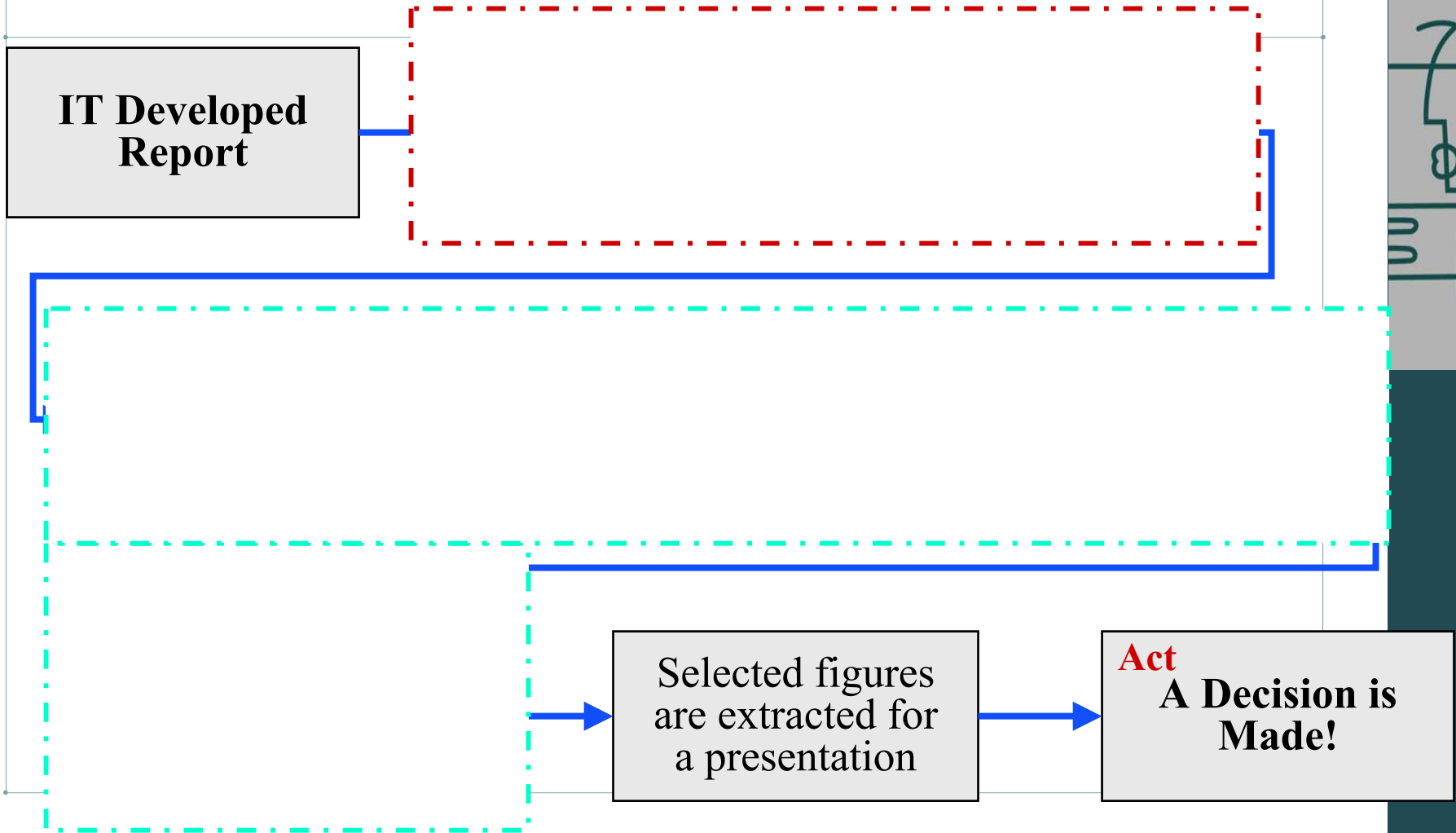


The Information Supply Chain

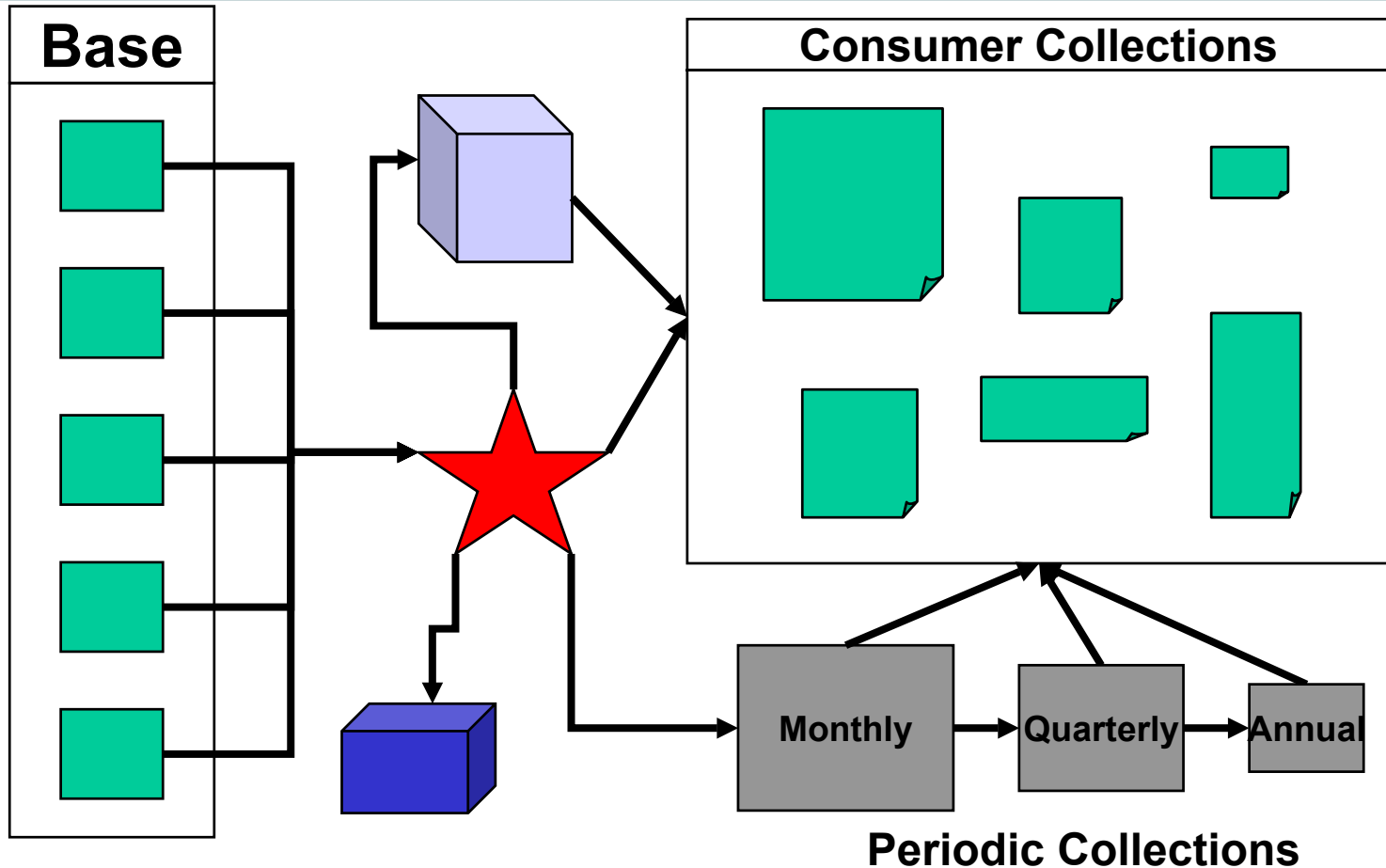
Steps Between “Get” and “Act”



Compressing the Info Supply Chain



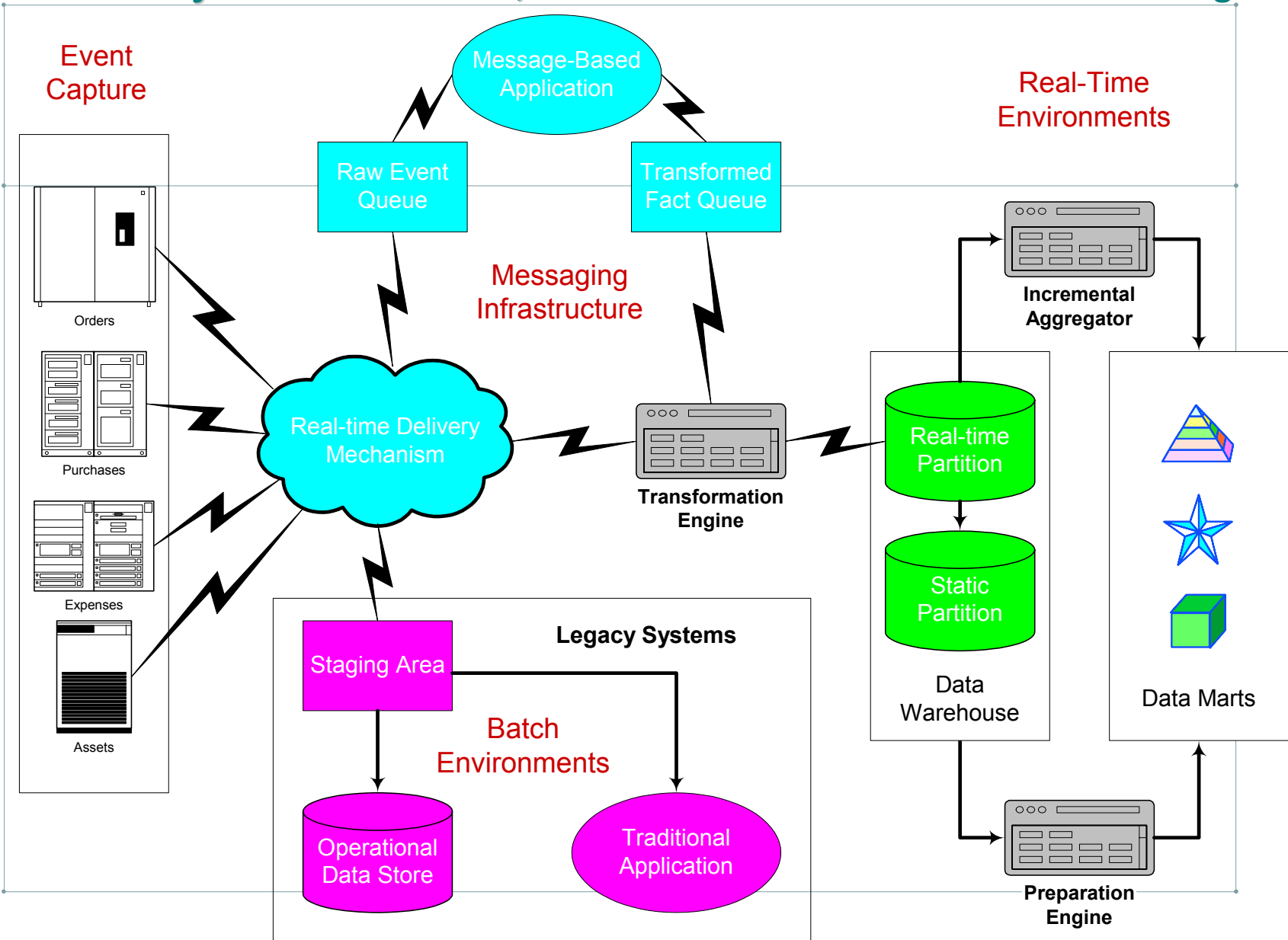
An Architecture for Flexibility



Components of a RTDW Environment

- **Real-Time Capture**
- **Real-Time Delivery**
- **Message-Ready Targets**
- **Transformation Engine**
- **Real-Time DW Partitioned Data Store**
- **Incremental Aggregator**
- **Preparation Engine**





Why Real-Time?

- Time Compression
- Infinite Snapshots
- Transitory States
- Quick Reconfiguration
- Perfect Capture

Real-Time DW provides much more than Real-Time Data!



The Essence of RTDW

- ✓ **Continuous capture of information**
- ✓ **Infinite snapshots with any time frame**
- ✓ **Reconfiguration on the fly**
- ✓ **Support for Active Closed Loop Systems**
- ✓ **Significant reduction in cost**
- ✓ **Monumental increase in value**



More On Real-Time Data Warehousing

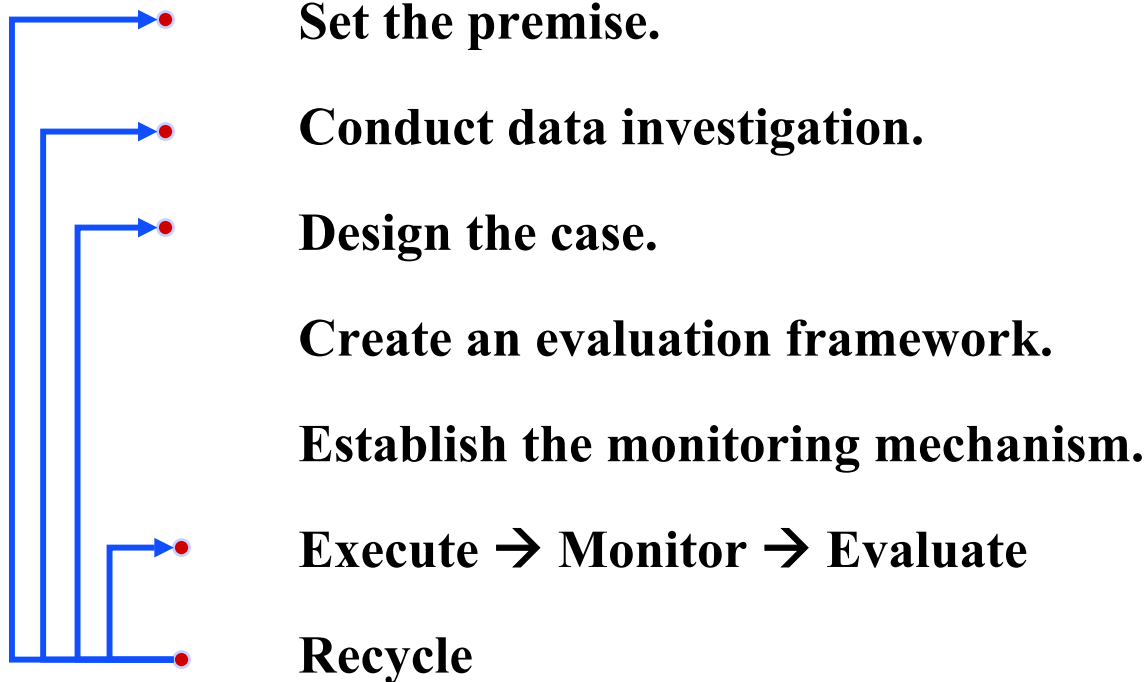
Data Management Review Website

Online Column by Michael Haisten

www.dmreview.com/master.cfm?NavID=152&AuthorID=608



BI Process for Closed Loop Analysis

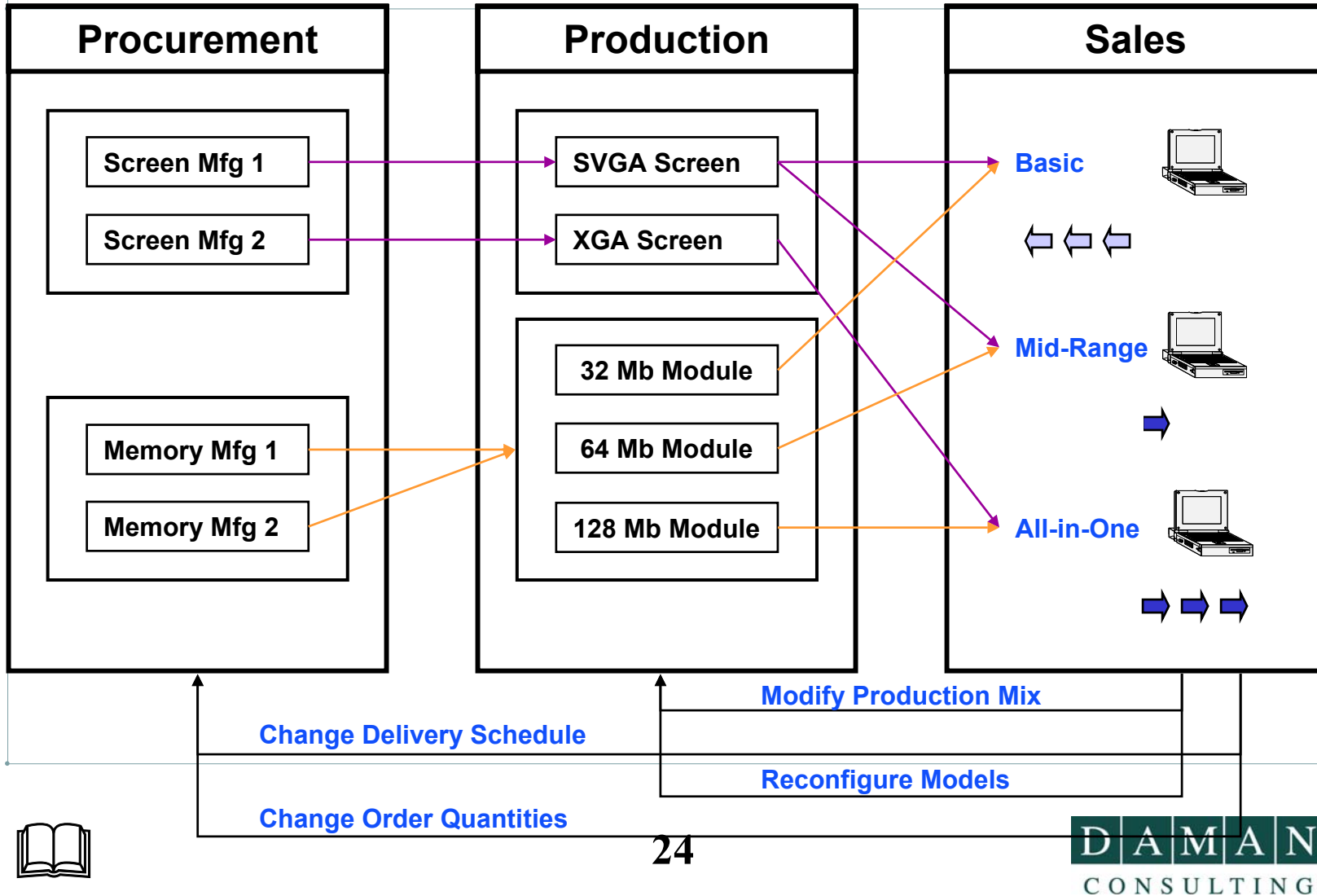


In eBusiness, closed-loop analysis is the business.



Closed Loop Operational Example

(Build-to-Order Laptop Manufacturer)





The BI Lifecycle

Michael Haisten

VP – Business Intelligence

Norman Comstock

Director – OLAP Solutions

BI Architectural Goals

- **Provide platform to deliver a great user-analyst experience**
 - With data that is consistent, centralized and easily accessible
 - Without getting in the way of OLTP systems
- **Ability to incorporate data from internal or external sources - regardless of format or platform**
- **Agility to adapt to changes in the business**

BI User Requirements

Support decision making – Strategy and Tactics

- Who/what/when/where/why/how of a business
- Facilitate queries without hindering operational systems performance or having to change the design
- Provide centralized repository of consistent data
- Answer complex queries quickly
- Enable data mining to discover new relationships in data

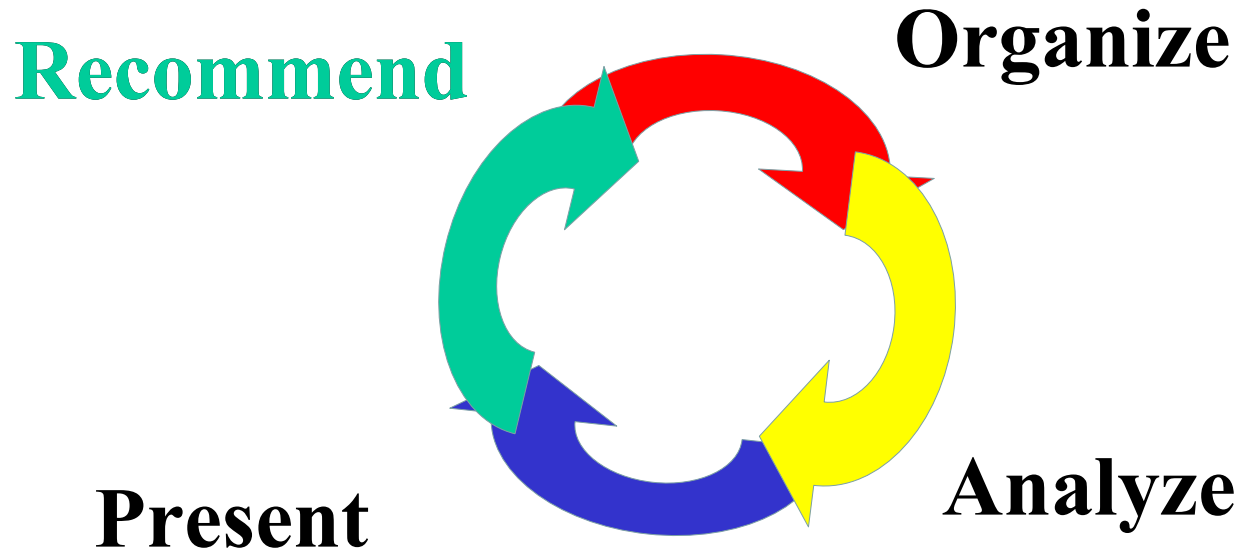
Provide different levels of analysis

- View data from many perspectives
- Easily navigate from summary to detail

Consumer acceptance is the true measure of success



The Business Intelligence Lifecycle



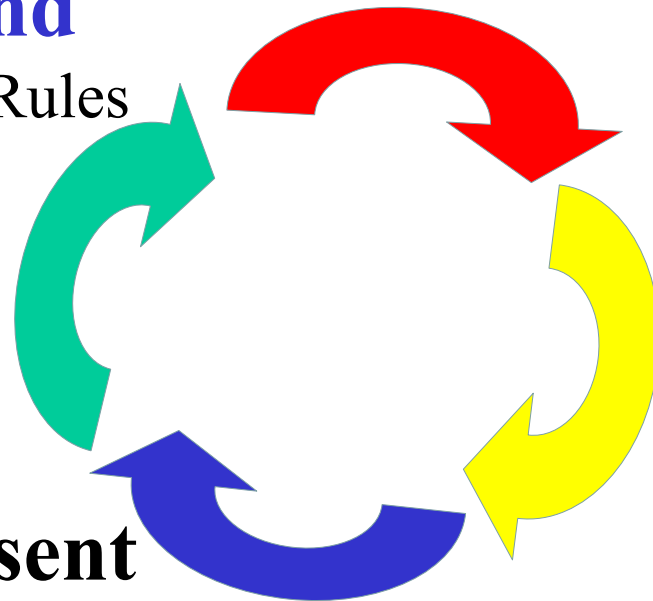
The BI Lifecycle - Organize

Recommend

- Business Rules
- Alerts
- Modify

Present

- Format
- Annotate
- Deliver



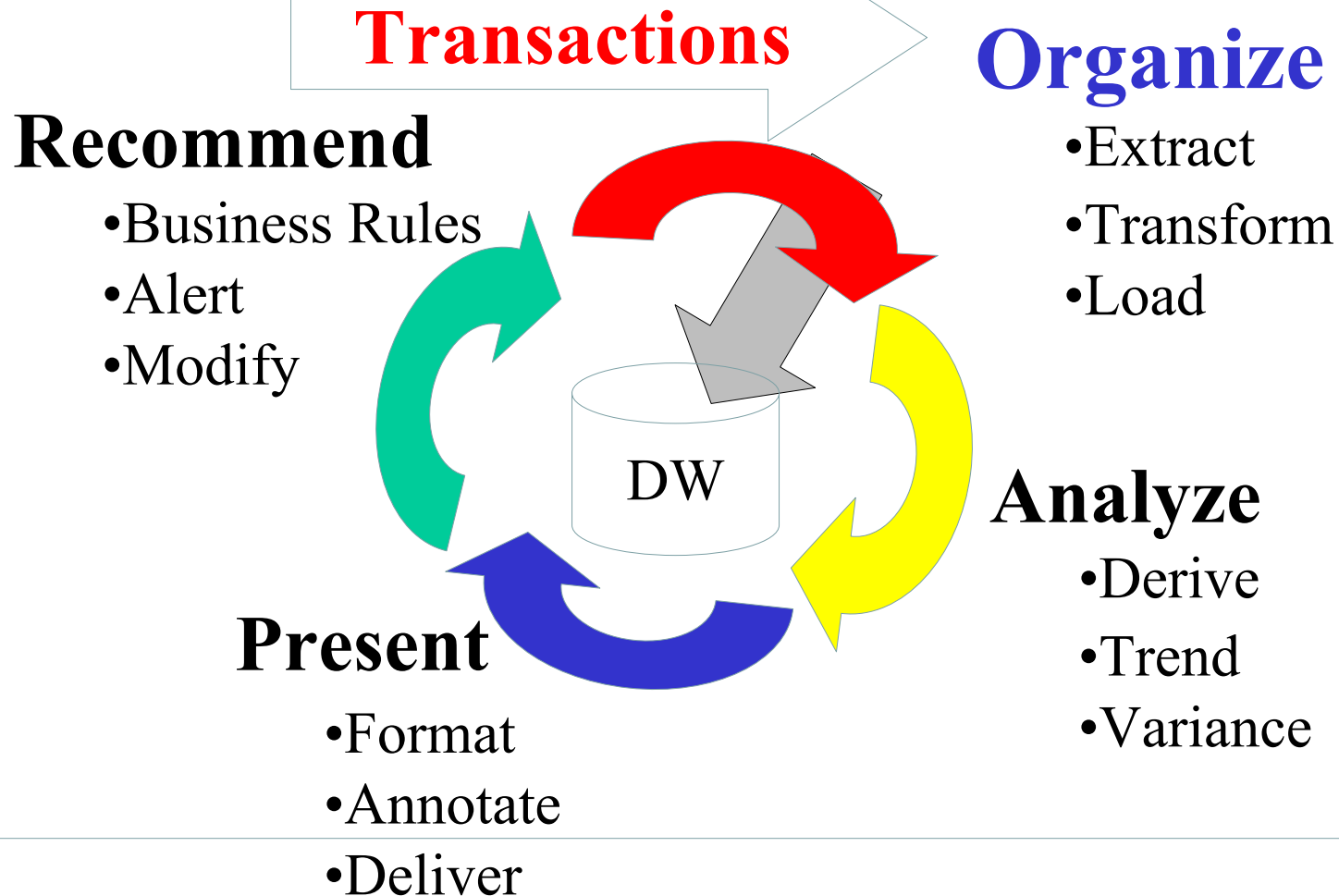
Organize

- Extract
- Transform
- Load

Analyze

- Derive
- Trend
- Variance

The BI Lifecycle - Organize



The BI Lifecycle - Organize



Activities:

- Plan, Audit, Cleanse, Model, Transform, Map, Load

Tools:

- Microsoft, IBM, Ascential, Brio, Informatica

Structures:

- ODS, Data Warehouse, Data Marts

Establish the Program	<ol style="list-style-type: none"> 1. Develop the Rationale 2. Target the Opportunities 3. Define the Architecture 4. Create the DW Program
Prepare the Project	<ol style="list-style-type: none"> 5. Plot the Project Tasks 6. Establish the Infrastructure 7. Procure the Tool Kit 8. Assemble the Team
Initiate the Database	<ol style="list-style-type: none"> 9. Investigate Consumer Needs 10. Triage the Source Elements 11. Model States and Dimensions 12. Design the Database 13. Acquire Source Data 14. Populate the Database
Explore the Data	<ol style="list-style-type: none"> 15. Iterate Base Table Design 16. Explore Consumer Usage Interactively 17. Tune Collection Design 18. Plan Cycles and Production Migration
Implement the Deliverables	<ol style="list-style-type: none"> 19. Prepare for Release 20. Train the Consumers 21. Initiate Support Processes 22. Migrate to Production
Expand the Environment	<ol style="list-style-type: none"> 23. Manage the DW Inventory 24. Synchronize with Evolving Business Needs 25. Evangelize Endlessly 26. Do it Again!

Step By Step™ Lifecycle



Data Warehouse Definition

Subject Oriented	Regrouped into Business Topics
Integrated	Connected by Common Domains
Consistent	Rationalized to Explain Variances
Non-Volatile	Organized for Repeatability
Time Variant	Presenting Multiple Periodicity
Historical	Retaining As-Was Detail
Dimensional	Standardized for Business Access
Adaptive	Configured for Future Needs

...collection of data for decision support.



Data Warehouse Architectures

Principles

Rules-of-the-road relating what is unique about data warehousing.

Information Architecture

A framework for managing the usage, meaning, structure, and movement of data within the enterprise.

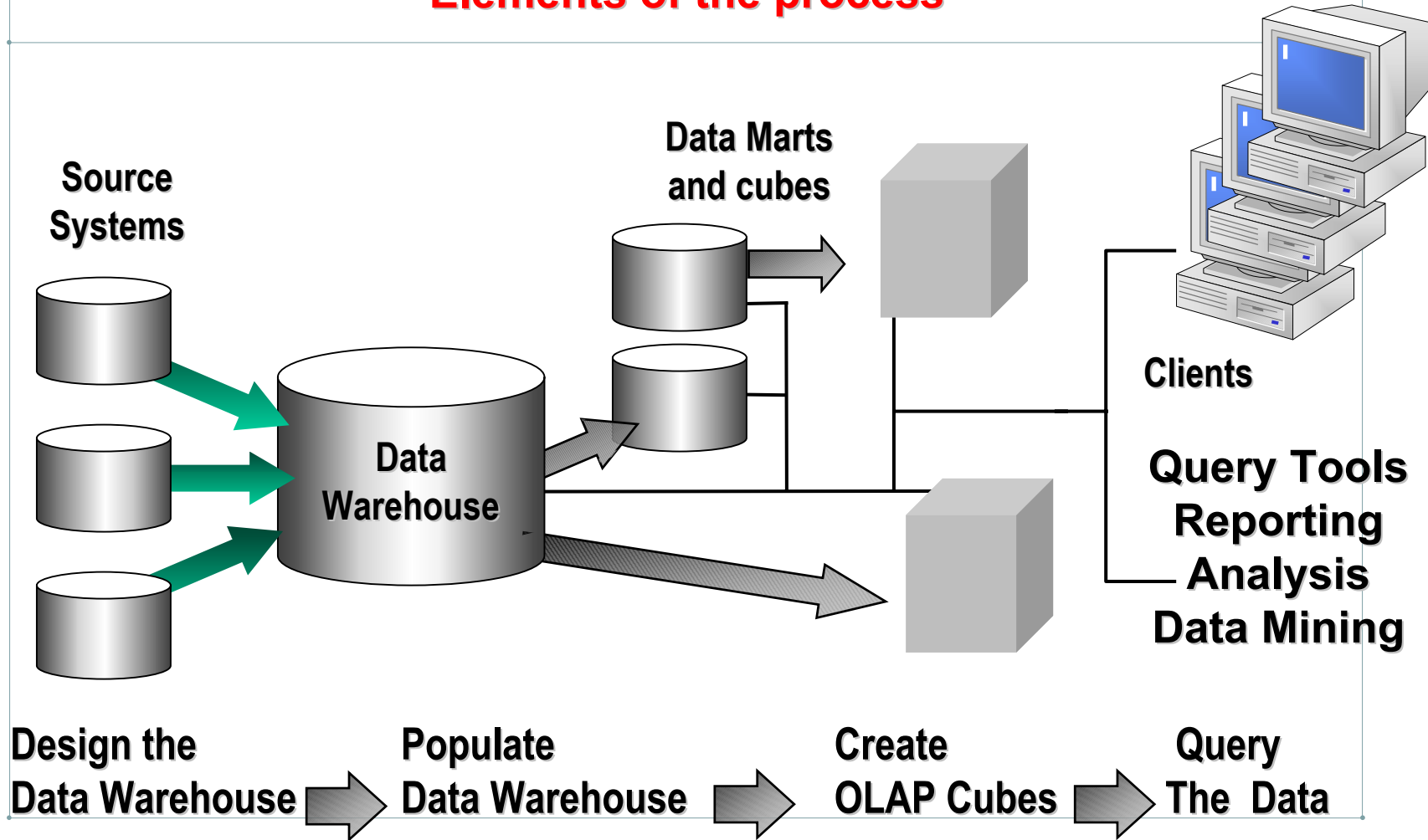
Technical Architecture

A component strategy for a data warehouse.

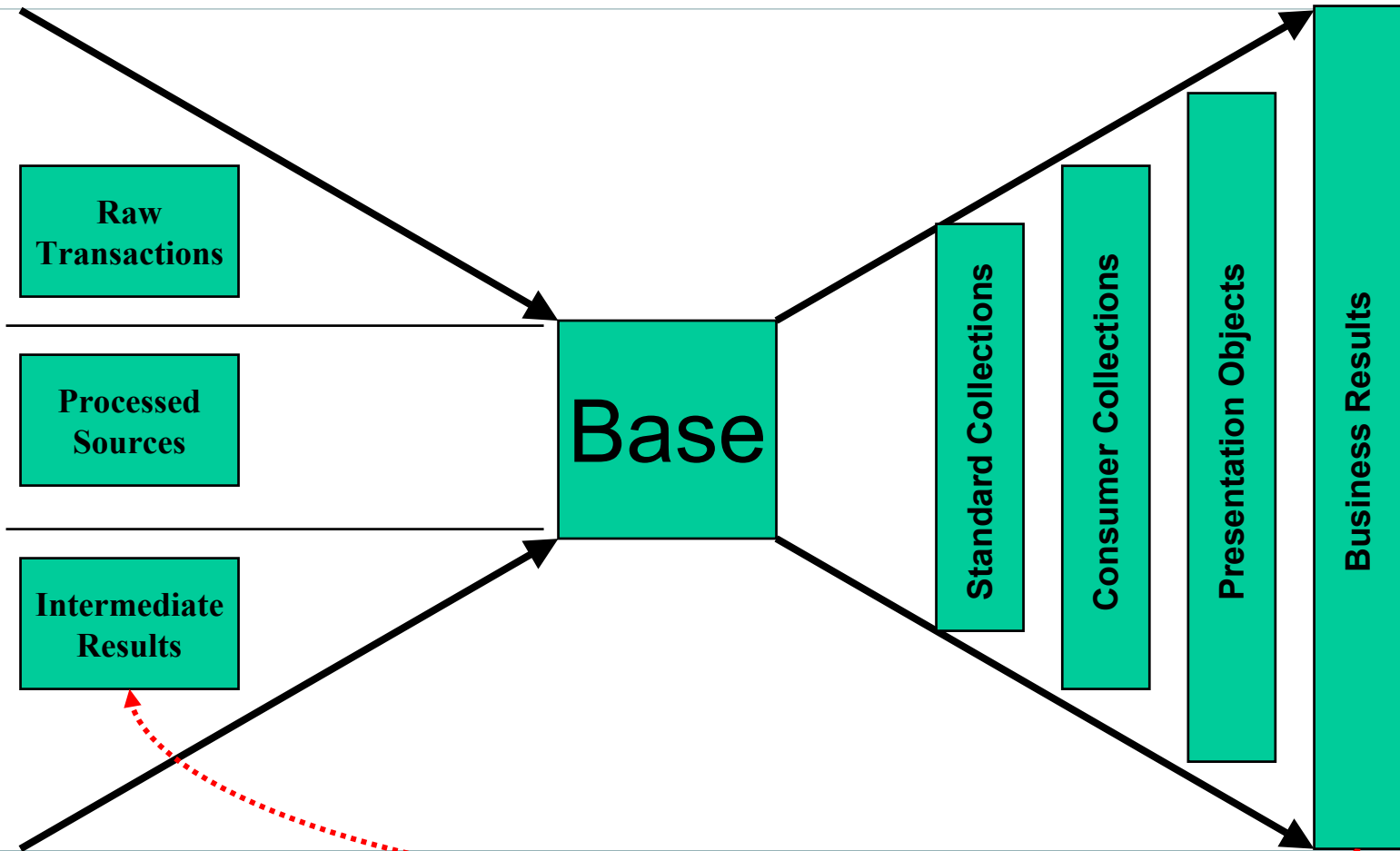


Classical Technical Architecture

Elements of the process



Collect – Integrate – Specialize



Optimal Design

- ❖ **Retains fundamental integrated base detail**
- ❖ **Provides common reference & translation tables for integration**
- ❖ **Uses data-driven quality management**
- ❖ **Retains as-is and as-was for consistency**
- ❖ **Creates the right number of collections**
- ❖ **Supports a diversity of data structures**
- ❖ **Captures intermediate results in the information supply chain**



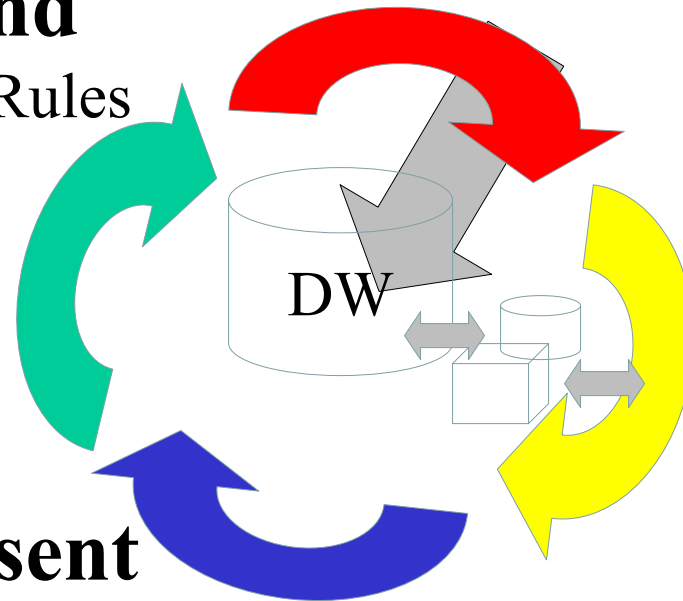
The BI Lifecycle - Analyze

Recommend

- Business Rules
- Alert
- Modify

Present

- Format
- Annotate
- Deliver



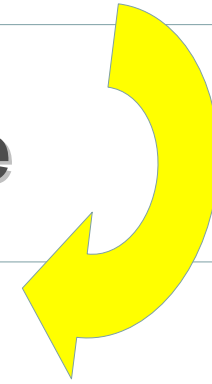
Organize

- Extract
- Transform
- Load

Analyze

- Metrics
- Trends
- Variance

The BI Lifecycle - Analyze



Activities:

- Query, Reporting, Statistics, OLAP, Data Mining

Tools:

- Microsoft, IBM, Brio, Hyperion, SAS, SPSS
Informatica

Structures:

- Flat, Norm, Denorm, Star Schema, Cubes (R/H/M)

Understand the Tool Categories

Report

Driven by output image

Weak access specification

Non-interactive usage

Query

Driven by access specification

Output format options secondary

Interactive but non-exploratory

Analyze

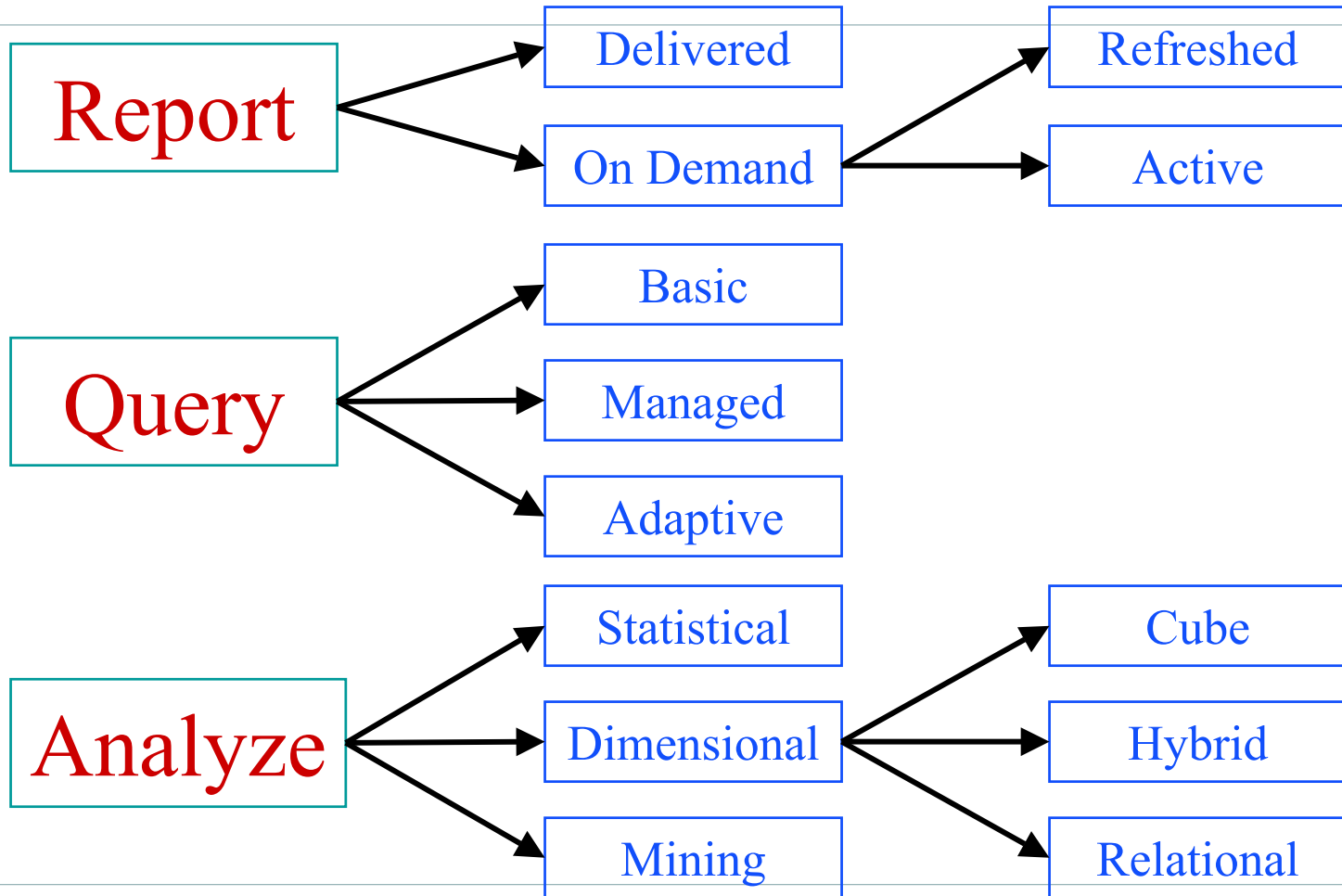
Driven by exploratory paradigm

Deterministic access path

Output format not a design concern



Understand the Tool Sub-Categories



What is OLAP?

OnLine Analytical Processing

- **“It’s a cool way of cheating that enables you to get queries answered incredibly fast”**
 - John Miller, SQL Server Program Mgr
- **OLAP pre-aggregates data across all dimensions**
 - Example: by MO, QTR, YR AND by Country, State, City ...
- **Basic argument:**
 - “When the need is known and repetitive, why summarize on the fly?”



Why Use OLAP With DW?

OLAP is an enabling technology that supports dynamic analysis

- Intuitive multidimensional model provides drill-down, slice & dice, and drill-through to the details when needed (if supported)
- Fast response times against huge underlying data volumes
- Offers complete syntax for expressing analytical queries supporting complex business logic
- Optimizes the use of network resources

ACCLIMATE

The OLAP Design Methodology for Effective Solutions

Assemble the Team

Conduct FSR Interview

Conduct IS Analyst Interview

Leverage DW infrastructure

Identify OLAP Engine and Presentation Tools

Model Presentation Modes Collaboratively

Amend OLAP Design for Prime Time

Train the Consumers

Exploit the OLAP Solution



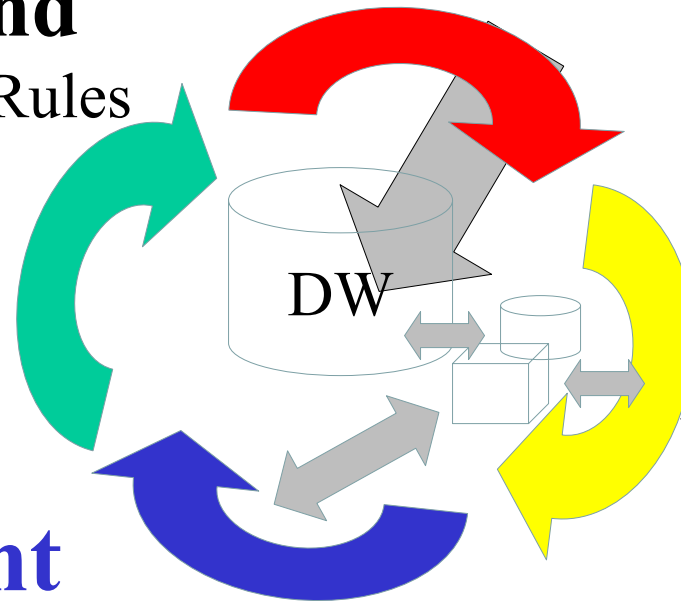
The BI Lifecycle - Present

Recommend

- Business Rules
- Alert
- Modify

Present

- Format
- Annotate
- Deliver



Organize

- Extract
- Transform
- Load

Analyze

- Metrics
- Trends
- Variance

The BI Lifecycle - Present



Activities:

- Format, Annotate, “Chart”, Publish, Deliver

Tools:

- Microsoft, Proclarity, Brio, Crystal, Hyperion,

Structures:

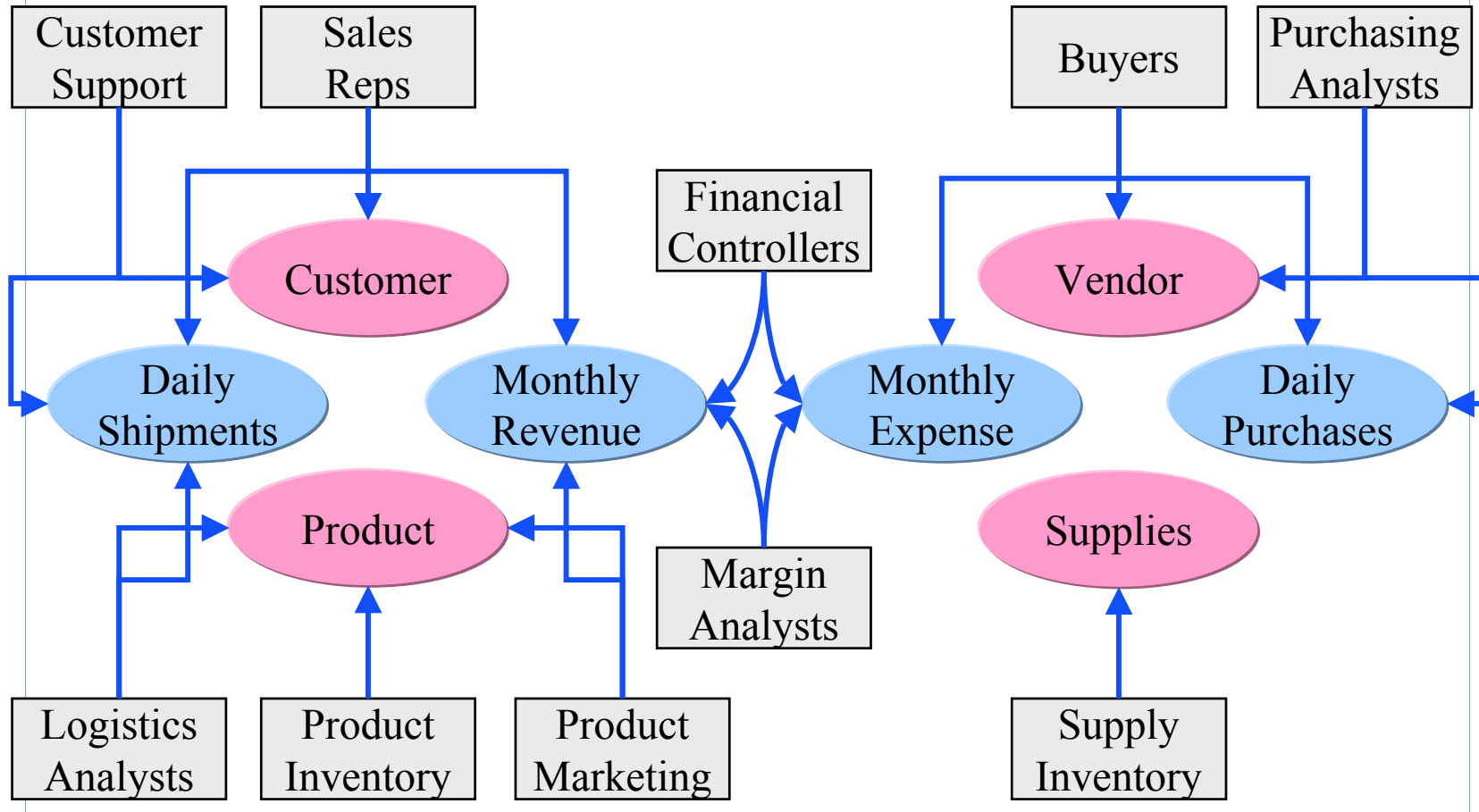
- Flat, Relational, Star Schema, Cubes (all forms)

Seven Habits for Designing Highly Effective OLAP Solutions

- **Visualize**
- **Design with the end in mind**
- **First things first**
- **Focus on the Customers**
- **Listen first, then execute**
- **Collaborate**
- **Review, Analyze, Iterate**



Constituencies: Expand the Use

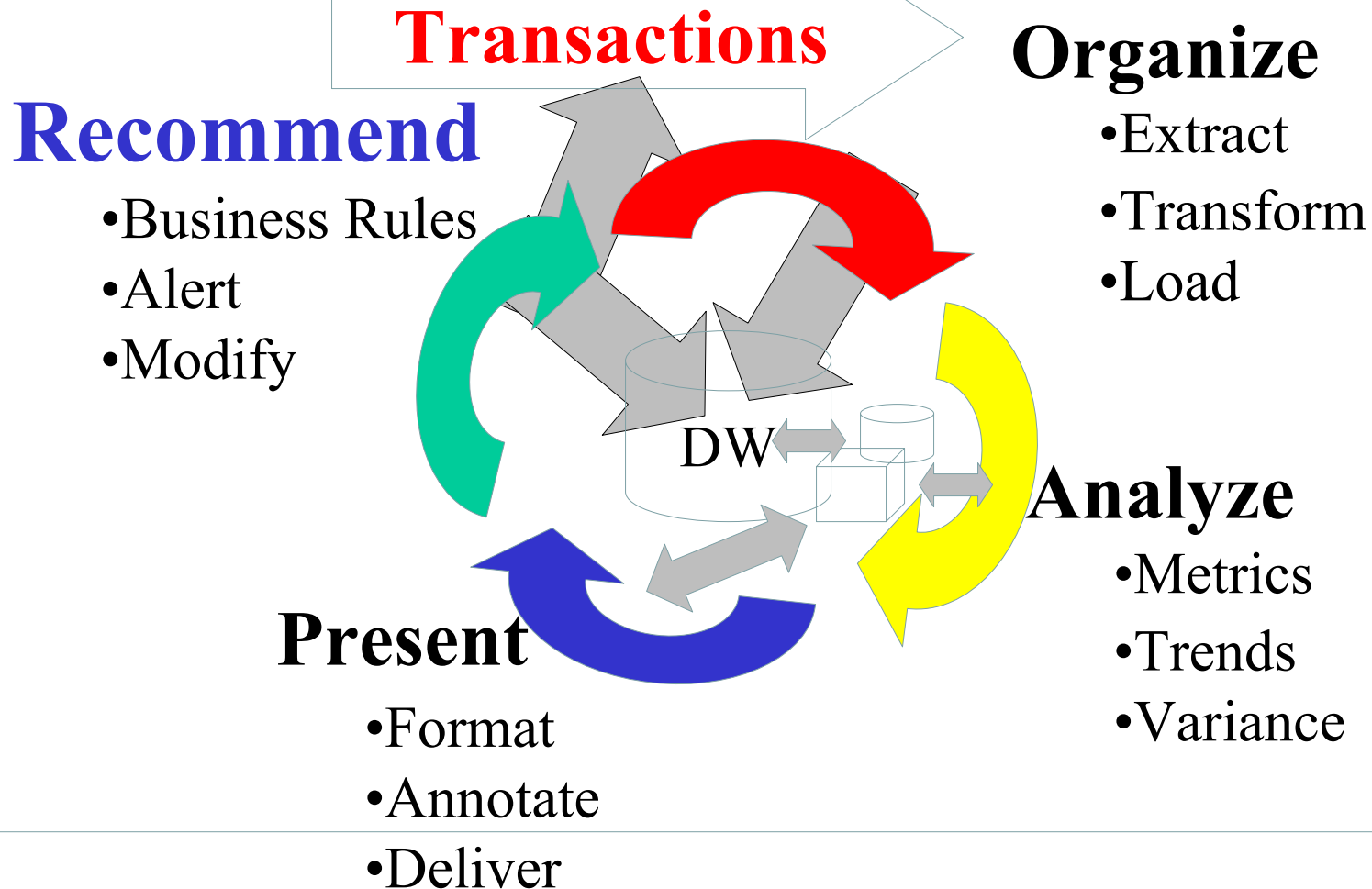


Know Your Consumers

<i>Value-Added Distributor</i>	Builder	Creates custom solutions
	Provider	Develops queries and provides data
	Mentor	Helps indirect consumers learn the tools
<i>Direct Information Consumer</i>	Hunter	Validates a vision
	Miner	Searches for insights
	Planner	Sets new targets
	Forecaster	Projects the future
	Analyst	Seeks the cause
	Tracker	Scans for targets
	Clerk	Generates results for others
<i>Indirect Consumer</i>	User	Uses data but not data access tools
	Skeptic	Does not do data (or so they say)



The BI Lifecycle - Recommend





Advanced Analytics

Michael Haisten

VP – Business Intelligence

Norman Comstock

Director – OLAP Solutions

Power Analytics Are:

- Structurally Diverse**
- Functionally Integrated**
- Visually Compelling**
- Broadly Scalable**
- Dimensionally Sophisticated**



Structurally Diverse

- Flat Files**
- Relational Tables – Normalized**
- Relational Tables – De-normalized**
- Relational Tables – Star Schema**
- R-OLAP Structures**
- H-OLAP Support**
- M-OLAP Cubes**



Functionally Integrated

Capabilities

Query

Reporting

Graphs & Charts

Animation

KPI Examination

Trending & Projection

Scenario Analysis

Pattern Identification

Features

Rich Embedded Functions

Macro Capability

Procedural & Declarative

Functional Extension

DBMS Integration

Presentation Diversity

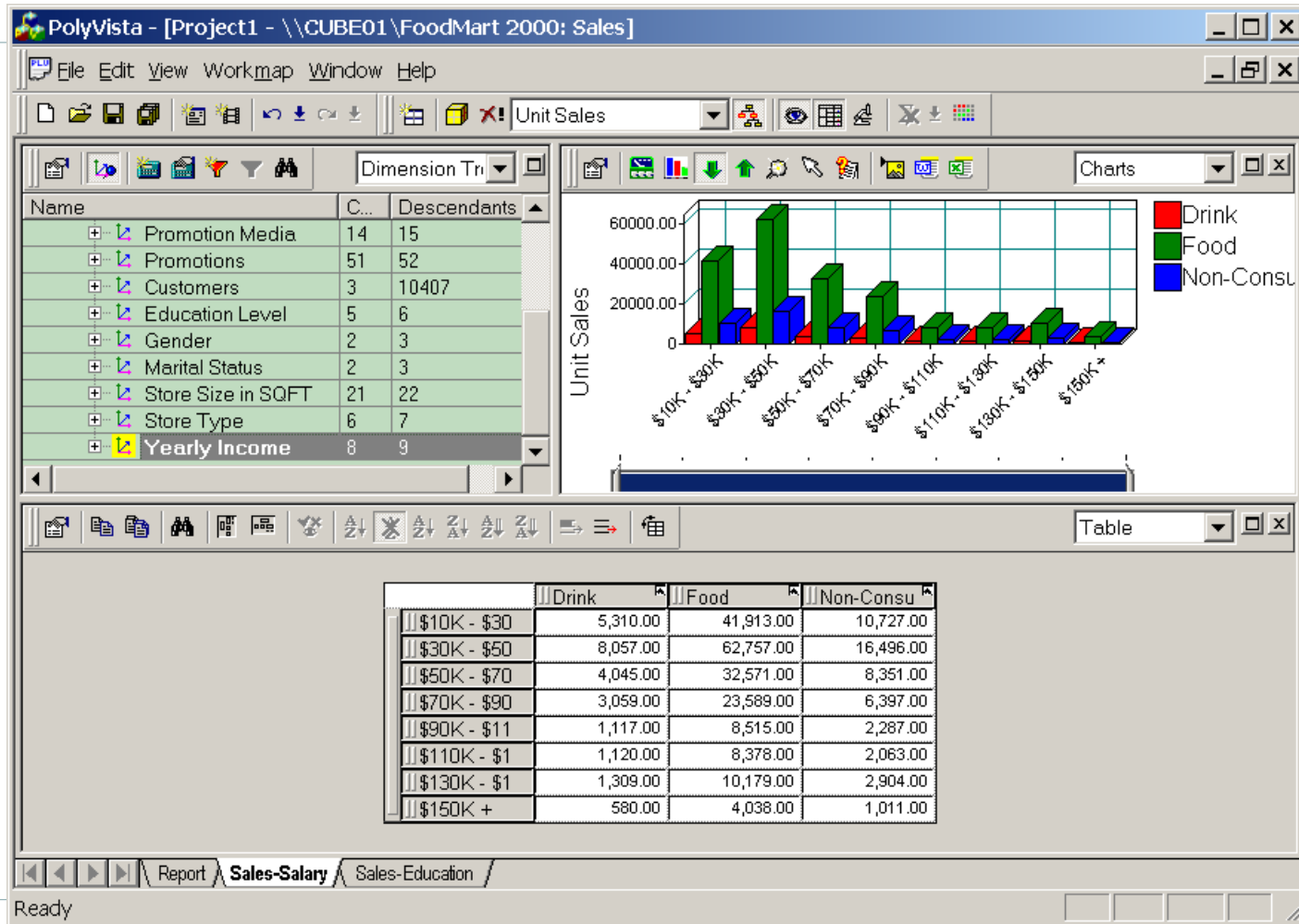


Visually Compelling

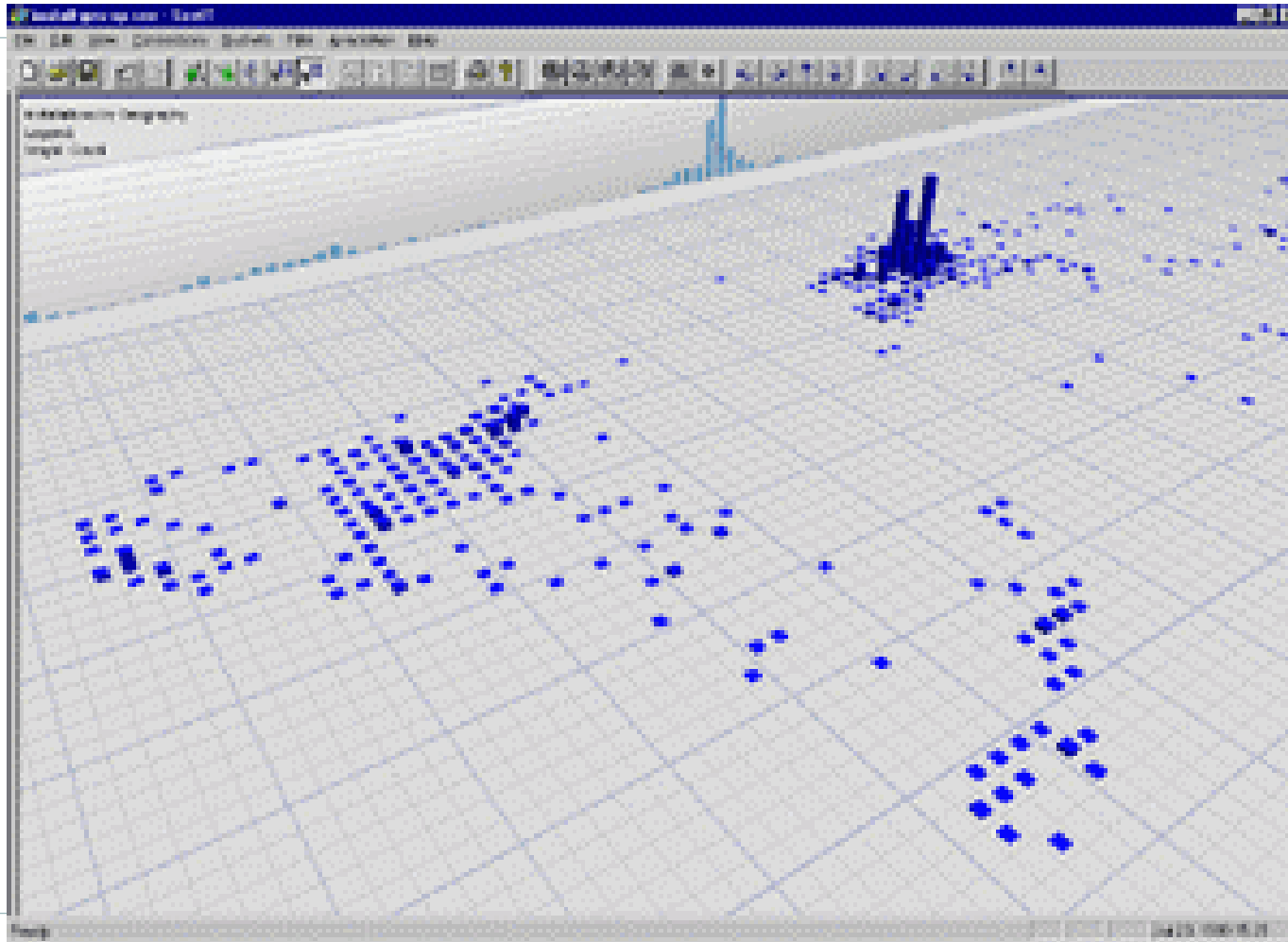
- Graphical Display**
- Graphical Augmentation**
- High Synthesis**
- Intuitive Navigation**
- Combined Presentations**



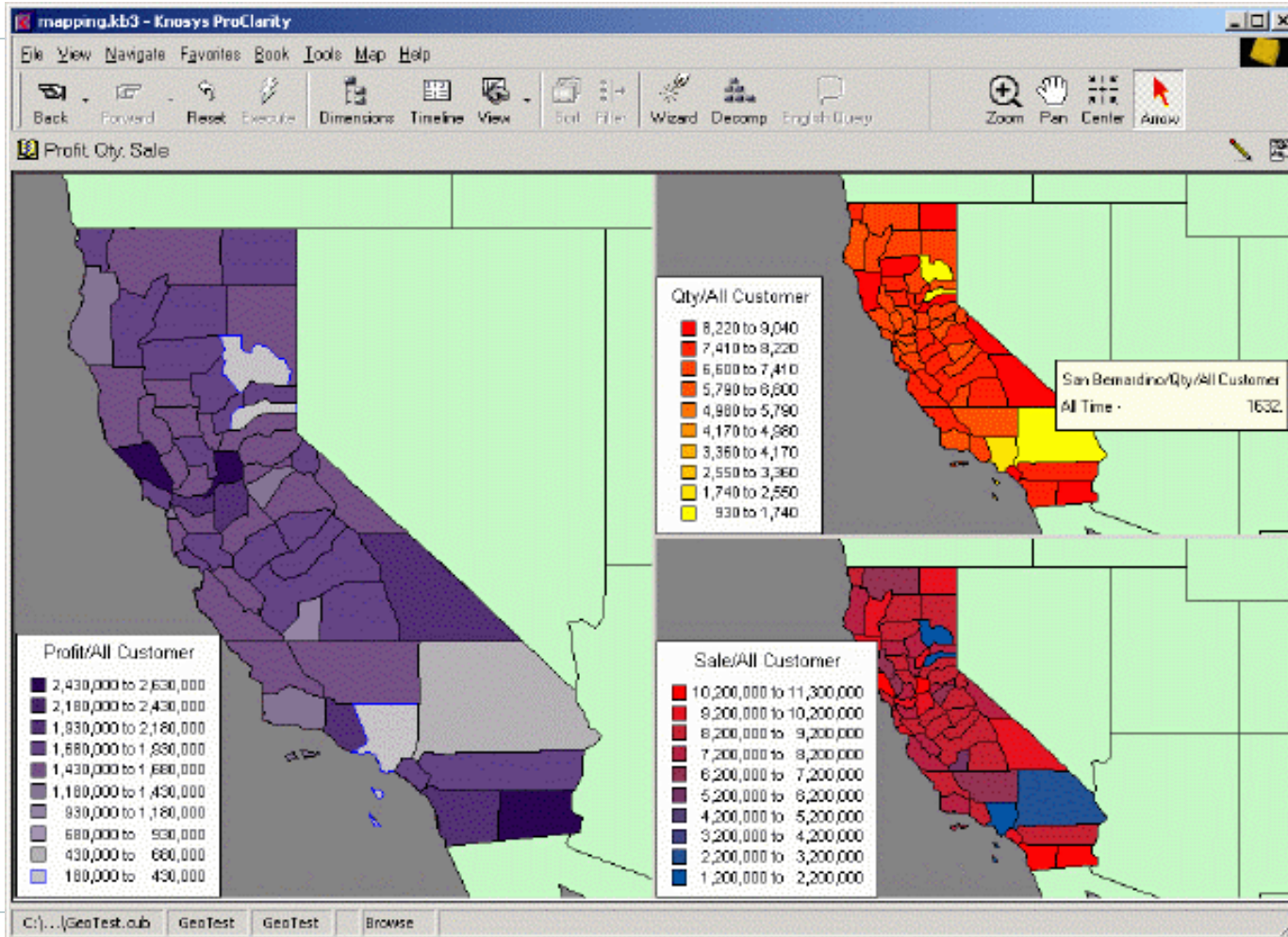
Reports, Charts, & Graphs



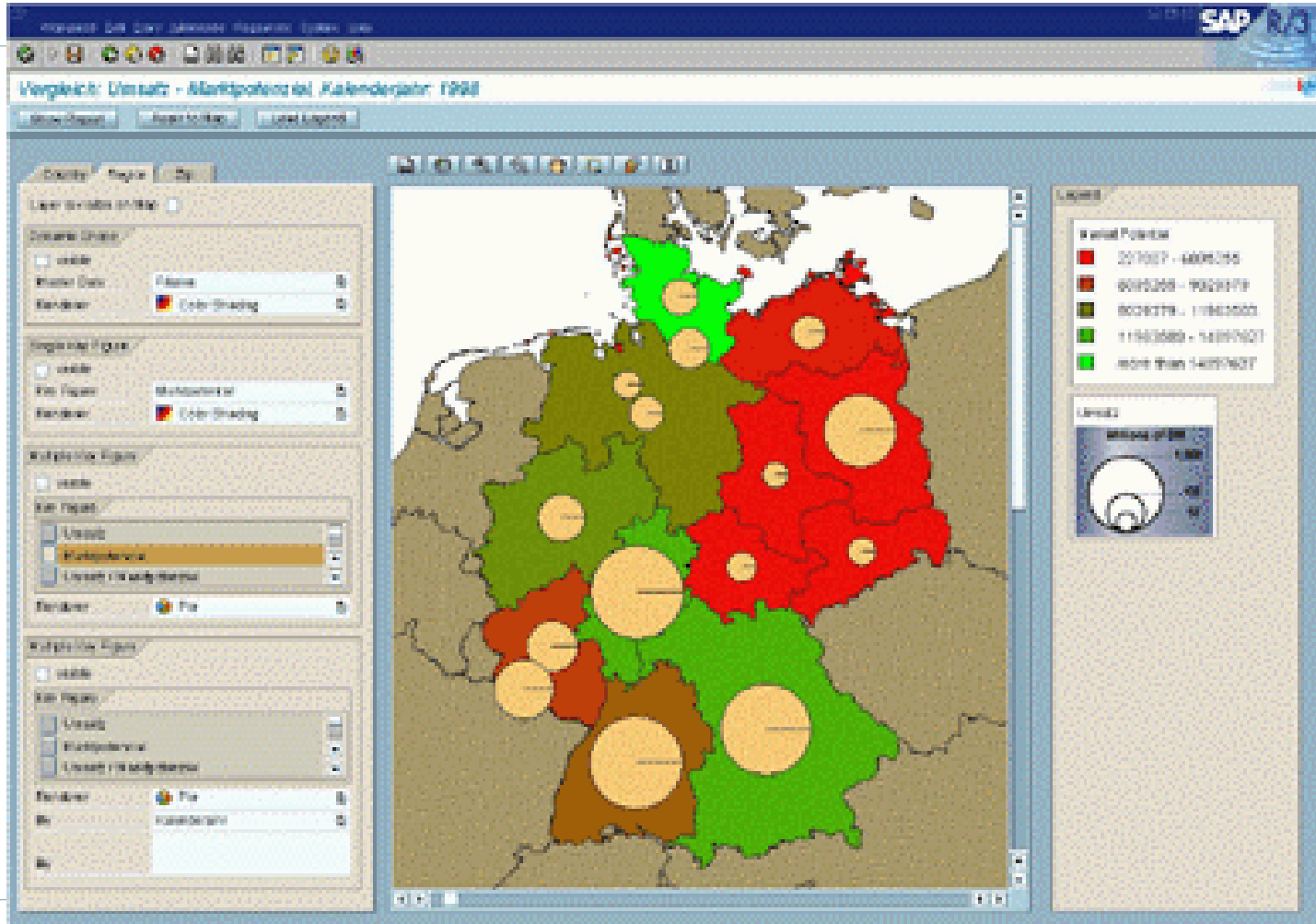
3D Visualization



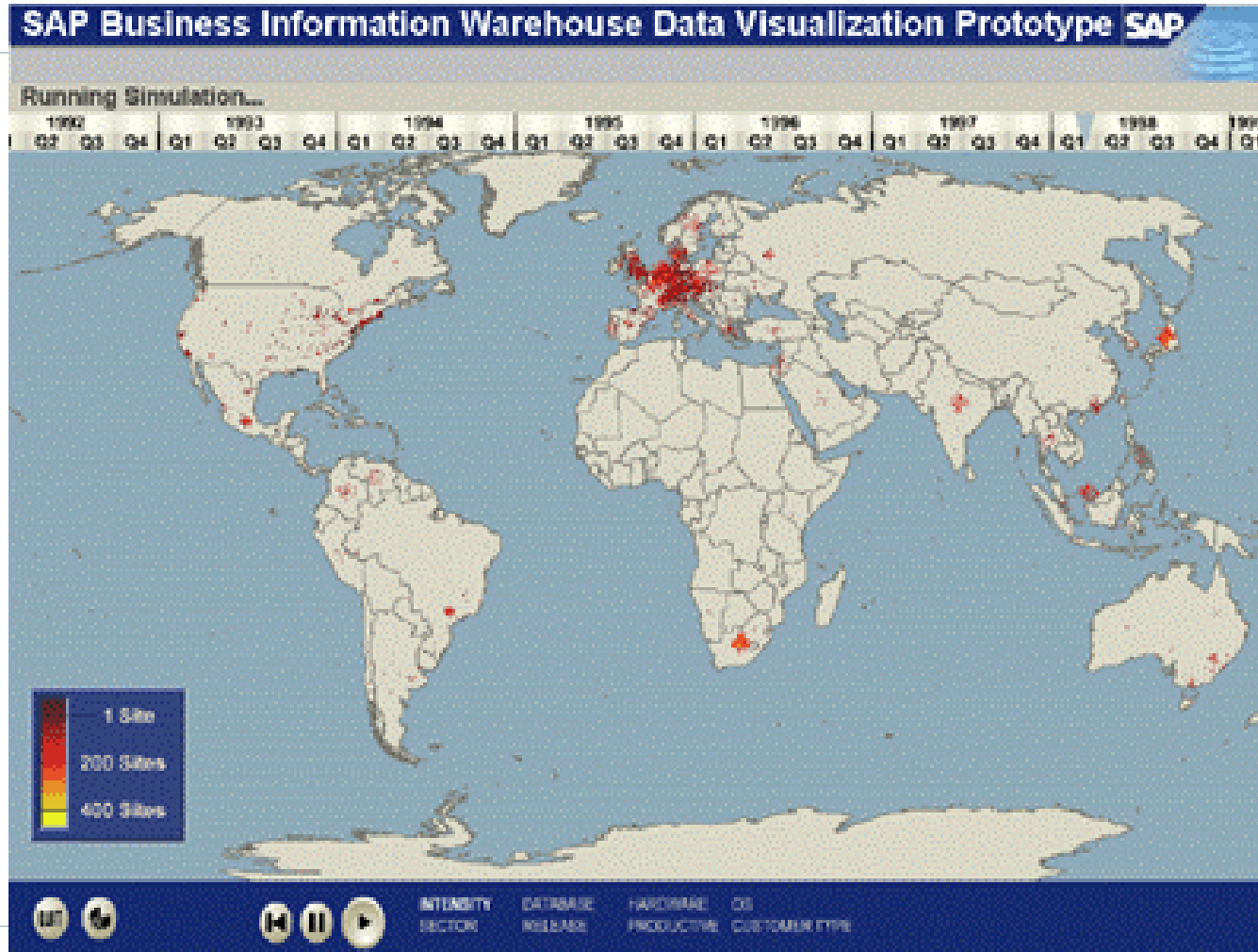
Geographic Spatial Mapping



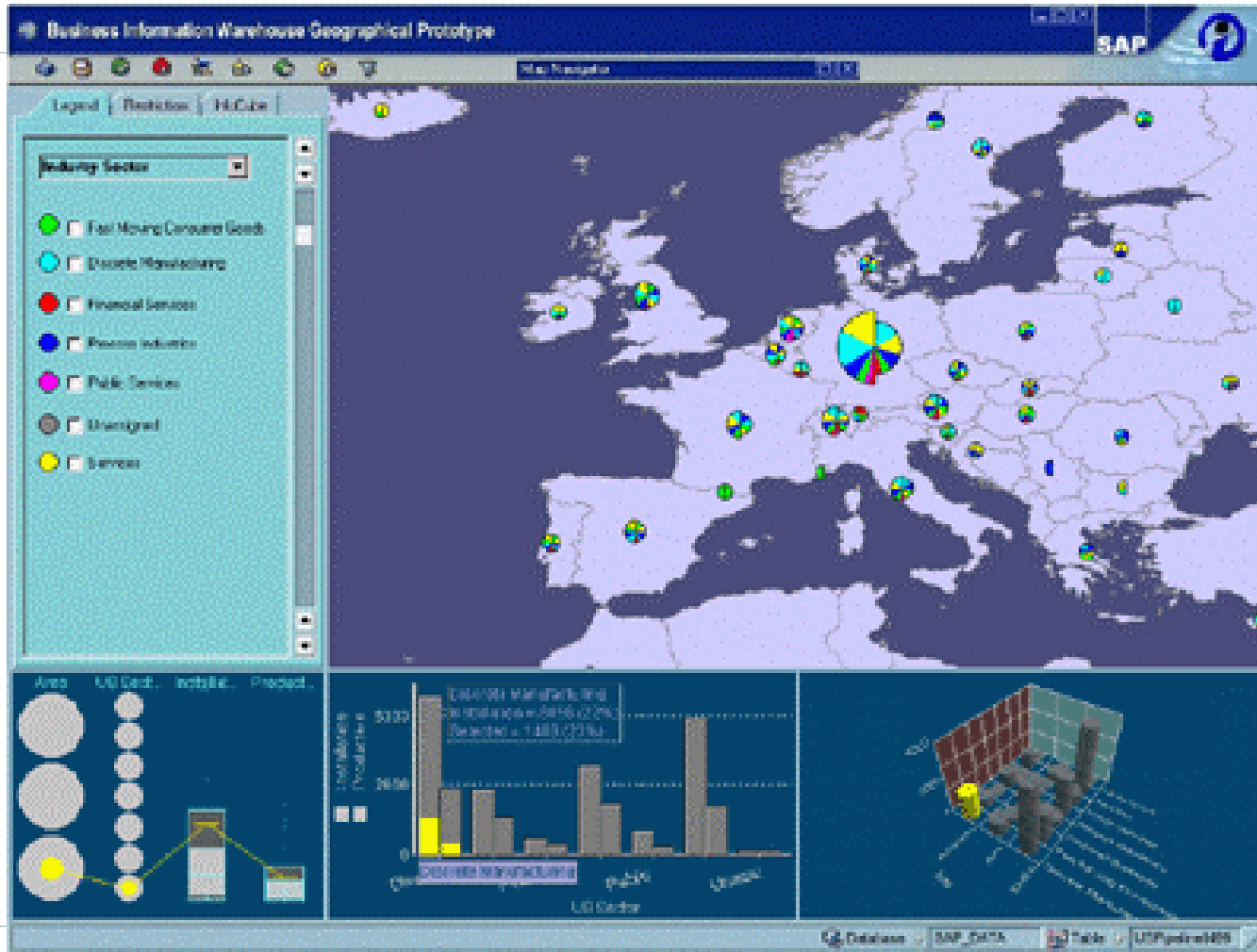
Geographic Data Presentation



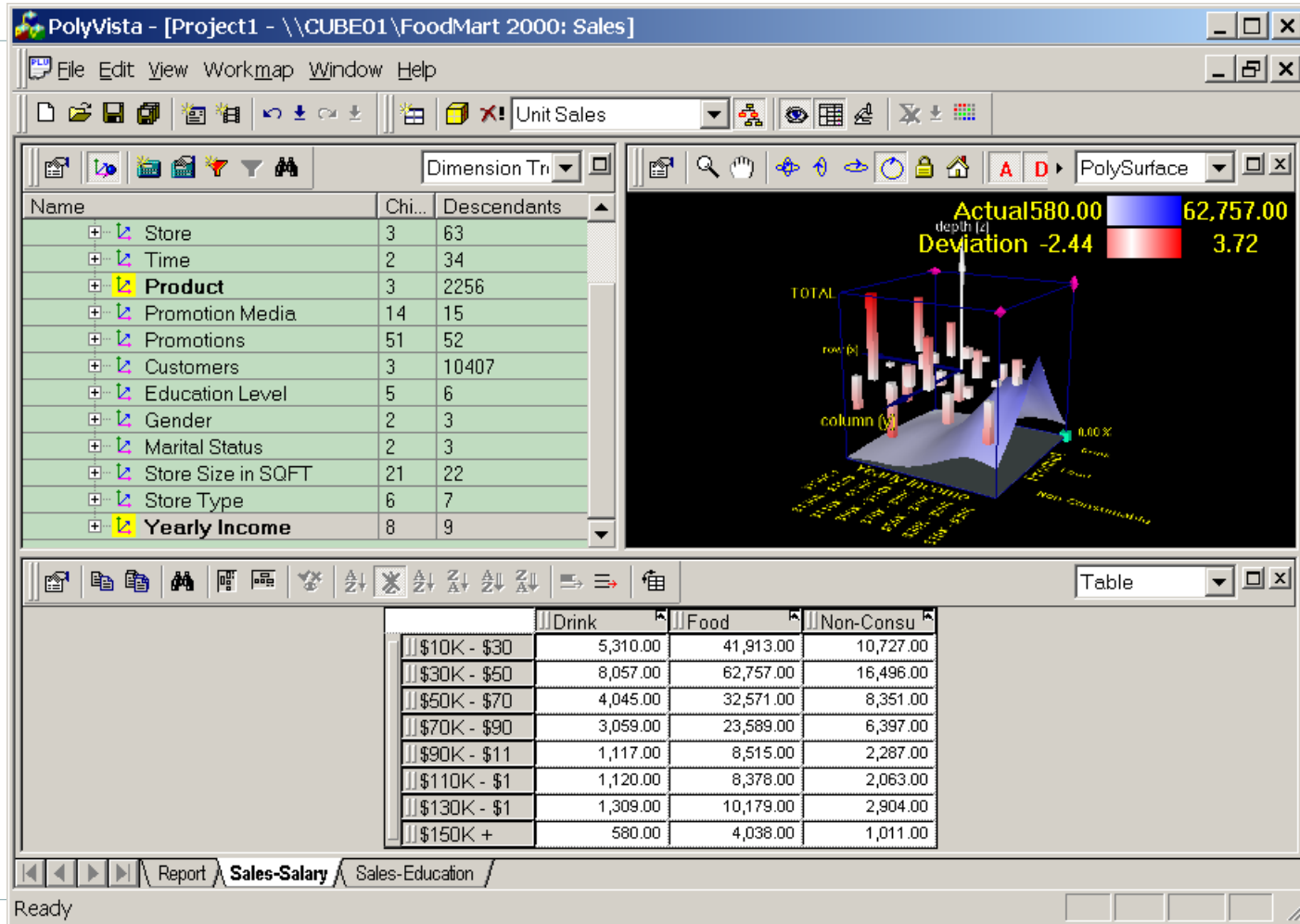
Animation over Time



Combined Visualization



Combined Visualization



Data Mining

The screenshot displays the PolyVista data mining software interface. The main window title is "PolyVista - [Project1 - \\CUBE01\FoodMart 2000: Sales]". The interface is divided into several panes:

- Dimension Tree:** A tree view on the left showing dimensions like Sales, Measures, Store, Time, Product, Promotion Media, Promotions, Customers, Education Level, Gender, and Marital Status. The "Gender" dimension is highlighted.
- PolySurface:** A 3D surface plot on the right showing a surface over a grid. It displays "Actual 1,603.00" and "Deviation -14.26".
- Discovery:** A pane at the bottom left showing a "Cross-Tab Description" table with a "Score" column. The top row is highlighted in green.
- Table:** A data table on the bottom right showing sales data categorized by income ranges and gender.

Cross-Tab Description Table:

Cross-Tab Description	Score
Gender=F, Marital Status=M, Yearly Income=\$70K - \$90K, Measure = Unit Sales	14.90
Marital Status=M, Time=Q1, Yearly Income=\$70K - \$90K, Measure = Unit Sales	10.46
Gender=M, Marital Status=M, Yearly Income=\$30K - \$50K, Measure = Unit Sales	10.22
Gender=M, Marital Status=S, Yearly Income=\$130K - \$150K, Measure = Unit Sales	10.14
Gender=M, Yearly Income=\$30K - \$50K, Measure =	10.14

Data Table:

	F	M
\$10K - \$30	14,057.00	14,200.00
\$30K - \$50	18,886.00	23,374.00
\$50K - \$70	11,809.00	10,194.00
\$70K - \$90	9,388.00	7,910.00
\$90K - \$11	3,243.00	2,522.00
\$110K - \$1	2,914.00	3,248.00
\$130K - \$1	3,352.00	3,409.00
\$150K +	1,687.00	1,603.00

Broadly Scalable

- Platforms**
- Volume**
- Dimensions**
- Members**
- Cubes**



Dimensionally Sophisticated

Advanced OLAP Techniques:

- **Distinguishing between dimensions & attributes**
- **Identifying dense vs. sparse dimensions ***
- **Selecting combinations of dense & sparse for performance ***
- **Determining the appropriate number of levels in a hierarchy**
- **Sharing leaf-node members to create alternate hierarchies**
- **Classifying flat dimensions with many members**
- **Using member aliases (logical & physical name)**
- **Building in write-to members as variables for flexible calculations**
- **Determining calculation pass order**
- **Determining calculation solve order**



Advanced Analytics with Microsoft Analysis Services

- ✓ **Monolithic scale of MOLAP cube (T3 Case)**
- ✓ **Systematic partitioning to expedite cube processing**
- ✓ **Complex and conditional cross-dimensional calculations (via MDX)**
- ✓ **Rich attribute analysis without propagating new dimensions**
- ✓ **Near real-time analysis**



Q & A



***Putting Intelligence Back into
Business Intelligence.***

**Daman Consulting
1250 Capital of Texas Highway South
Building One, Suite 340
Austin, Texas 78746**

(512) 329-6646

www.damanconsulting.com

