VISUAL ANALYSIS FOR A SENSE-AND-RESPOND ENTERPRISE

Shubir Kapoor, Donna Gresh, Josef Schiefer, Pawan Chowdhary, Steve Buckley
IBM T.J. Watson Research Center
1101 Kitchawan Road
Yorktown Heights, NY 10598, USA
Email: { shubirk | gresh | schiefer | chowdhar | sbuckley}@us.ibm.com

Abstract

A Sense-and-Respond enterprise is one that is constantly evolving to keep up with the rapidly a changing business environment. While these enterprises are "data rich," they are often at the same time "information poor." There is a critical need to:

? *Display large amounts of data* – to see the both the "needle" and the "hay."
? *Make complex patterns stand out* – to focus on relationships that are relevant.
? *Interact with the data* – to isolate any interesting subset as the analysis proceeds.
? *Correlate across many variables* – to identify complex interrelationships.
? *Be easy to understand* – to make learning quick, to make the information usable by a wide audience and to facilitate communication among decision-makers.

Monitoring the vast amounts of data is more than just providing real time alerts. It about saying “There is something going on here and here is all the information that will help you find the needle in the haystack and make a well informed timely decision”. In this paper we introduce data visualization techniques that are well-suited to finding the needle in the haystack because it can easily present large volumes of data, provide interactivity to explore the data, make visual patterns easy to see and make multivariate analysis simple and easy to comprehend.

Keywords: Information Visualization, Data Analysis, Sense-and-Respond.

1 Introduction

The widespread use of the Internet and related technologies in various business domains has accelerated the intensity of competition, increased the volume of data/information available, and shortened decision-making cycles considerably. Consequently, strategic managers are being exposed daily to huge inflows of data and information from the businesses they manage and they are under pressure to make sound decisions promptly. Typically, in a large organization, many distributed, heterogeneous data sources, applications, and processes have to be integrated to ensure delivery of the best information to the decision makers. In order to support effective analysis and mining of such diverse, distributed information, a data warehouse (DWH) collects data from multiple, heterogeneous (operational) source systems and stores integrated information in a central repository.

Since market conditions can change rapidly, up-to-date information should be made available to decision makers with as little delay as possible. A well-developed strategy is vital, but its ultimate value to an organization is only as good as its execution. As a result, deployment of business intelligence solutions for *operational* and *tactical* decision-making is becoming an increasingly important use of information assets. For a long time it was assumed that data in the DWH can lag at least a day if not a week or a month behind the actual operational data. That was based on the assumption that business decisions did not require up-to-date information but very rich historical data. In today's competitive business climate, companies can't afford to wait days or weeks for information that can reduce costs and improve business processes.

Companies need to track business processes -- such as order processing, quality assurance, inventory, logistics, compliance, etc. in real time, to improve operational efficiency as business events are happening. They are also looking for answers to the tough questions: How can we improve revenues? What characteristics do our most profitable customers share, and how can we serve them better? Where are we spending money and seeing the best returns? Executives today are more challenged than ever to make quick, well informed decisions that address growing business issues and regulatory standards. To answer these questions, companies need a window into the health of the organization and the tools to act. Every organization is made up of a complex combination of data, assets, processes and systems.

The challenge lies in integrating, monitoring and maintaining them all and allowing users to manipulate the data and relationships in a graphical form gives unparallel opportunities to make full use of resources. As a result, more powerful and timely business analytics and visualization solutions are needed to help produce reliable results, drive confident decision-making, and gain insight into valuable business data that uncovers unforeseen trends and previously unarticulated relationships. However, these discoveries are not useful unless applied quickly to the decision-making process. This is where the use of advanced visualization techniques for data analysis comes into play and helps build a Sense-and-Respond Enterprise. It allows analysis results to be shared across the enterprise quickly, saving deployment cycles and lowering costs, while enhancing ROI.

2 Sense-and-Respond Enterprise

A Sense-and-Respond Enterprise is, in the long run, in a much better position to provide on-demand services than a traditional Make-and-Sell organization. Sense-and-Respond is a general approach for adaptive business management first popularized in IBM by Stephen Haeckel of the Palisades Advanced Business Institute [1]. Haeckel describes the transformation from a Make-and-Sell organization to a Sense-and-Respond organization. He advocates a new form of strategic planning based on roles and responsibilities. In his view, organizational hierarchy is replaced by a dynamically configured network of modular capabilities. Governance is performed on the basis of context and coordination by people in roles accountable for outcomes rather than by command and control (Figure 1).

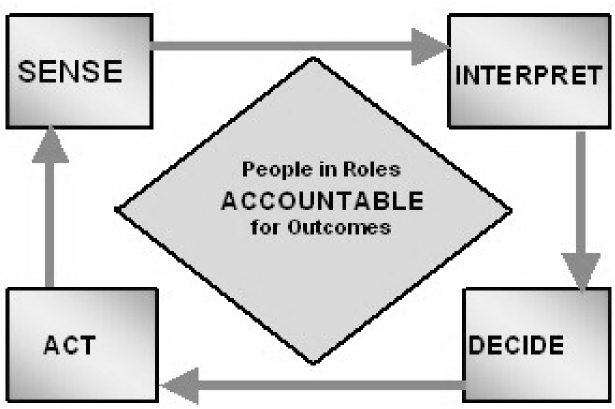


Figure 1: Stephen Haeckel's Sense-and-Respond loop.

Decision makers typically use exception-based analysis on published metrics to identify opportunities, then dig deeper into the data to understand the causes of those opportunities. From there, they model the business situation so that they have a framework against which to evaluate different decision alternatives. After selecting an alternative the user acts accordingly to the decision made. Figure 1 shows such a decision cycle. The Sense-and-Respond loop involves 4 sub processes for the decision making: *sense*, *interpret*, *decide*, and *act*. The *sense* process monitors and collects desired data (based upon the intents) from the business environment. The output from the *sense* process is in the form of metrics or key performance indicators, which provide the information foundation for the interpret and decide processes. After the gathering and preparation of data from various sources, the *interpret* process inspects and analyzes this data based on the business intents and constraints. The detection phase usually generates new business situations that are further explored. The data analysis helps to predict the performance and assess the risks of the available options for responding to the business environment. It facilitates determining the root causes of the identified business situations. Key for the determination of causal factors is the ability to identify inherent relationships and dependencies between variables that drive the situational or exceptional performance. The outputs of the *interpret* process are alternatives to improve the current business situation and the guidance for the decision makers to select the best alternative. The *decide* process selects the best option and also determines the most appropriate action for a response to the business environment. The *act* process executes appropriate business actions based on the decision that has been made. This response will either change the state of the business environment or notify other agents (humans or programs) who may be interested in the outcome and result of the decision making. The *sense* and *act* processes need to interact with the target business environment in order to obtain data/metrics and to trigger actions, respectively.
This papers focuses on visualization techniques for the *interpret* and *decide* processes. We present visualization techniques that support business analysts and decision makers for these two processes in order to 1) discover hidden business patterns/irregularities and find their root causes and 2) making well-informed decisions for effective responses to the business environment.

3 Visual Techniques

Visualization provides much more than making patterns easy to see. Visualization should be highly interactive, permitting the user to perform multidimensional analysis across many variables at once. Multivariate data is particularly challenging because it is difficult to see relationships across many variables simultaneously. Combining visuals to see multiple attributes at once plus using interaction to slice and dice across all the dimensions simultaneously provides a very powerful means to achieve multivariate correlation. Data visualization makes the slice-and-dice process much more intuitive so than other techniques, including pull-down menus (which require pre-existing structure), pivot tables and mining algorithms. Combining real-time and near real-time data analysis with advanced visualization provides a progressive, in-depth view of an organization's infrastructure and processes over long intervals.

This section describes some of the visual techniques used within the Sense-and-Respond architecture developed by IBM Research [2]

3.1 Background / History

Visualization has become an important component of scientific data analysis over the last 20 or so years. The beginning of "scientific visualization" is often dated to an NSF workshop on Visualization in Scientific Computing [3]. Over the years, a variety of systems and tools have become widely available to see, for example, the patterns of wind flow over a simulated aircraft wing, or to plot variations in ocean water temperature However, despite a rich literature in methods of "information visualization" [4], visualization of unstructured data such as tables of marketing information, has been slower to be adopted in business environments, beyond the use of basic graphs and charts.

The primary goal of any visualization is to see patterns more easily than would be possible by simply looking at tables of numbers. In the case of "scientific visualization," this is typically achieved by arranging the data in a physically meaningful way (i.e. in simulated two or three dimensions), and coloring and shading the data to represent magnitude, velocity, or other variables of interest. The challenge of visualizing unstructured data is that there is no underlying spatial organization of the data as there typically is for scientific data, thus the choice of spatial metaphor is less obvious.

In the information visualization literature, a commonly used technique is multiple views of the data with "brushing and linking.". Individual visual presentations of the data are typically created for subsets of the variables, where the presentation style depends on the characteristics of the data. For example, numerical data are appropriately represented by scatter plots or histograms, while categorical data (that is, data represented by, for example, a small number of string values) may be represented by a proportionally sized rectangle or a textual list. In the "brushing and linking" paradigm, individual data points in one view can be interactively colored, and that color is maintained and propagated to other views of the data. In this way, relationships between variables can be more easily seen

3.2 Benefits

In the business application discussed here, the data consist of various "events" which are of interest to the user. While the primary display mechanism for these events is a textual "alert" to notify the user of a business event needing a response, there remains a need for the user to be able to delve more deeply into the patterns, causes, and relationships of the historical record of events. As humans are primarily visual beings, the pattern recognition process can be much more efficient when presented graphically than when simply presented in tabular form. In this way, the user can potentially make decisions altering the business operations based on the additional understanding provided by the visual analysis.

3.3 OPAL

One component of the Sense-and-Respond dashboard is the capability of visually analyzing the overall character of historical data to spot trends, outliers, and other patterns. For the application discussed here, we used the Opal library of information visualization to create customized panels for the various measures under consideration. Opal provides a variety of possible views of multidimensional data, and each view is automatically linked by the Opal infrastructure so that when color is applied to a point in one view, it is automatically propagated to all other views. Thus, for example, in a scatter plot of two variables, a user may decide to color all large values in one of the variables red. The result would be that in any other view of the data, those same points would also appear red. Thus it might be easily discerned that large values in one variable are generally also large in some other variable.

Opal is a Java library, developed in-house at IBM Research. For the application discussed here it was used to create a set of applets to provide the required visual presentations and interactions. The applets are populated with the desired data, with the user having the ability to filter the data by the time horizon in which he or she is interested.

3.4 Root Cause Analysis with OLAP

Root cause analysis is a step by step method that leads to the discovery of a problem's first or root cause. An unwanted business situation which wastes time and resources can happen for a number of reasons. There is a definite progression of actions and consequences that lead to the failure or unwanted situation. A root cause analysis investigation traces the cause and effect trail from the end failure or problem back to the root cause.

OLAP (online analytical processing) is analytical processing that enables a user to easily and selectively extract and view data from different points-of-view [5]. Whereas a relational database can be thought of as two-dimensional, a multidimensional database considers each data attribute (such as product, geographic sales region, and time period) as a separate dimension that define the context for a fact. OLAP software can locate the intersection of dimensions (all products sold in the Eastern region above a certain price during a certain time period) and display them. OLAP has the ability to answer "why?" and "what if" questions about business situations which is a mandatory requirement for root cause analysis and also essential for the discovery of previously unseen relationships between data items.

For exploration of data, OLAP provides navigational operations like select, drill-down, roll-up and pivot conforming to the multidimensional view of data. OLAP systems have fast response times so that the users can interactively invoke sequences of such operations. This is a very crucial capability for performing root cause analysis because the data analysis can be driven by the flow of thoughts of the analyst. A typical OLAP analysis sequence proceeds as follows: the user starts from an aggregated level, inspects the entries visually (perhaps aided by some graphical tools), selects a subset to inspect further based on some intuitive hypothesis or needs, drills down to more detail, inspects the entries again and either rolls up to some less detailed view or drills down further and so on. The analysis operations deliver "just-in-time" information to the users that is needed for effective decision-making for future actions. Just-in-time information is computed data that usually reflects complex relationships and is calculated with acceptable latency for the users.

In order to use OLAP for Sense-and-Respond, a flexible data management solution for preparing data for the analytical processing is required. A semi-generic data model for a data warehouse ensures that OLAP systems can respond to changing business requirements as needed for effective decision making.

The purpose of the Sense-and-Respond data warehouse is to provide a data foundation for a process-driven decision support system to monitor and improve business processes continuously and in near real-time. It is a global information repository, which is seamlessly integrated with an organization's business processes and enables users to access comprehensive information on business processes very quickly, at different aggregation levels, from different and multidimensional points of view, over a long period of time, using a huge historical data basis prepared for advancing the business operations with more intelligence and also to effectively support the management of business processes.

Sense-and-Respond systems assess business situations by monitoring KPI's (Key Process Indicators). A data model for the data warehouse has been designed which captures the KPIs with context information. Figure 2 illustrates an example for a data model for capturing the performance data of a transportation management system. The data model has been divided into two parts:

1) An industry specific part which is modeled as a star schema [6] and therefore includes fact tables and dimension tables. The star schema facilitates building OLAP cubes for root cause analysis. In order to make fact tables more extendible and manageable, we use generic metrics in the fact tables that are described by attributes in the metric type dimension (MetricTypeDim).
2) Generic data for Sense and Respond that stores the metric type definitions, commitment profiles, situation, and user definitions. The table structures that captures this type of data is the same for all Sense-and-Respond solutions.

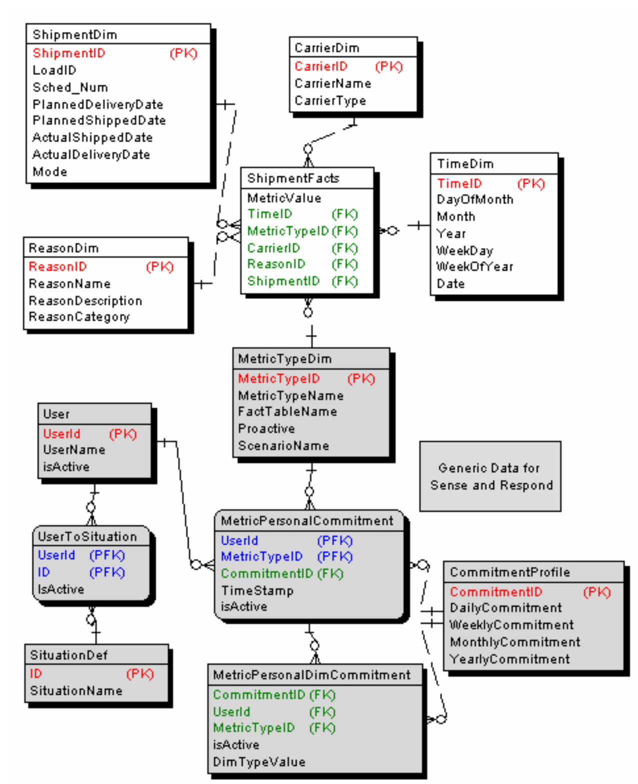


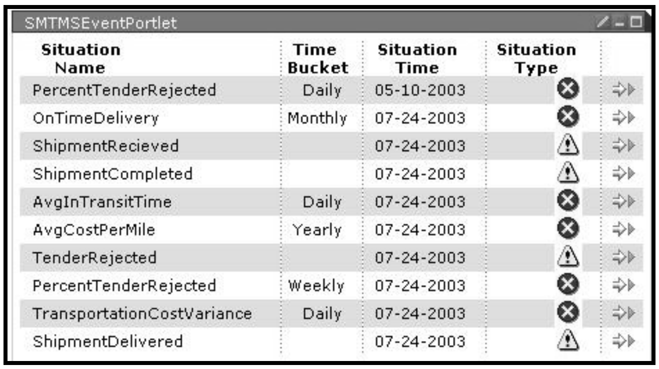
Figure 2: Data Warehouse Model.

As you can see in Figure 2, we use metric types to link situations and commitments of a Sense-and-Respond loop with the industry specific parts. The data warehouse design enables the addition of new metric types to existing fact tables (for example, when the solution is already in production) and easy extension with new fact tables with a different metric context.

In our Sense-and-Respond solution for transportation management, we captured metrics for shipments in the MetricTypeDim table such as average in transit time, expedite ratio, on time delivery, transportation cost variance, and average cost per mile. The context for these metrics includes time, shipment, carrier and reason which are defined in the dimension tables. We implemented Java programs for loading the data into the OLAP cubes and for updating the aggregates. For the access to OLAP cubes and execution of analysis operations, we also use Java interfaces. For our solution, we also built *Analytical Agents* that use this Java interface to perform automated root cause analysis.

4 Usage Scenario

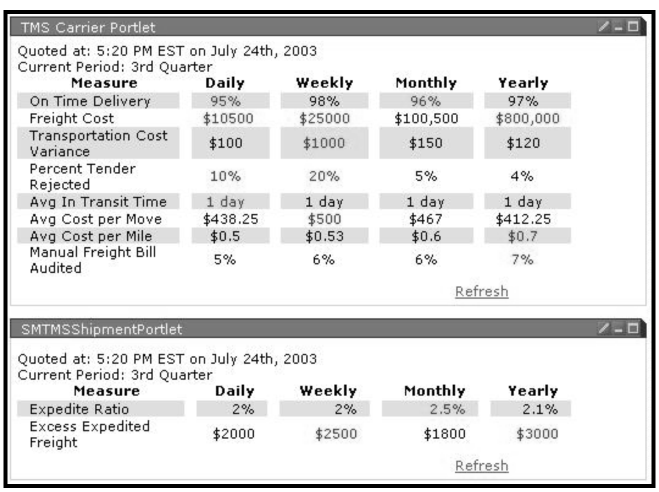
In this section we describe the visual analysis capabilities built for an on-demand outsourced supply chain management system, which we call SaRTM (Sense-and-Respond for Transportation Management). The SaRTM system provides visibility as well as proactive monitoring and notification across the entire transportation life cycle – load planning, tendering, carrier load and dispatch, in-transit execution, post-delivery accounting and financial processing. Events received from planning and execution systems are intelligently broken down by the SaRTM system to retrieve key details, correlate and aggregate data into KPI's and determine if any out-of-threshold situations have occurred or might occur in the near future. Users are alerted to situations in Dashboard portlets. Role players identified in the requirements phase of a SaR engagement receive personalized dashboards for viewing KPI's, situations, alerts and actions that are directly relevant to their line of business. These users receive alerts in 'near real time' if exceptions occur in their business (see Figure 3). For example, if a tender is rejected by a carrier, the transportation planner is notified. If more than 10% of overall tenders are rejected across all carriers in the previous month, the transportation analyst/manager is notified to take corrective action to resolve the problem. As part of this corrective action, the analyst might need to drill down and determine which carrier(s) have the most tenders rejected.


SMTMSEventPortlet

Situation Name	Time Bucket	Situation Time	Situation Type
PercentTenderRejected	Daily	05-10-2003	⊗
OnTimeDelivery	Monthly	07-24-2003	⊗
ShipmentRecieved		07-24-2003	⚠
ShipmentCompleted		07-24-2003	⚠
AvgInTransitTime	Daily	07-24-2003	⊗
AvgCostPerMile	Yearly	07-24-2003	⊗
TenderRejected		07-24-2003	⚠
PercentTenderRejected	Weekly	07-24-2003	⊗
TransportationCostVariance	Daily	07-24-2003	⊗
ShipmentDelivered		07-24-2003	⚠

Figure 3: Dashboard alerts in SaRTM.

The dashboard also provides a snapshot of the business for visibility purposes. As part of this snapshot, it indicates the value of the KPI's across all time buckets. KPI's that have crossed their thresholds are indicates in a separate color. Figure 4 shows an example of dashboard visibility.


TMS Carrier Portlet

Quoted at: 5:20 PM EST on July 24th, 2003
Current Period: 3rd Quarter

Measure	Daily	Weekly	Monthly	Yearly
On Time Delivery	95%	98%	96%	97%
Freight Cost	$10500	$25000	$100,500	$800,000
Transportation Cost Variance	$100	$1000	$150	$120
Percent Tender Rejected	10%	20%	5%	4%
Avg In Transit Time	1 day	1 day	1 day	1 day
Avg Cost per Move	$438.25	$500	$467	$412.25
Avg Cost per Mile	$0.5	$0.53	$0.6	$0.7
Manual Freight Bill Audited	5%	6%	6%	7%

Refresh

SMTMSShipmentPortlet

Quoted at: 5:20 PM EST on July 24th, 2003
Current Period: 3rd Quarter

Measure	Daily	Weekly	Monthly	Yearly
Expedite Ratio	2%	2%	2.5%	2.1%
Excess Expedited Freight	$2000	$2500	$1800	$3000

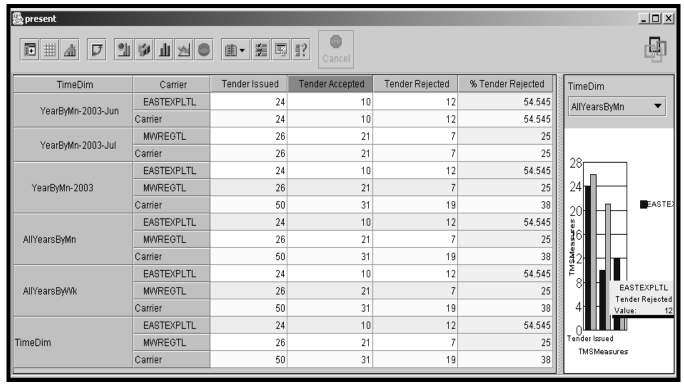
Refresh

Figure 4: Dashboard visibility in SaRTM.

Drill-down information is provided by creating OLAP cubes over the data warehouse. Coupled with advanced visualization techniques, OLAP helps users perform root cause analysis and identify trends.

4.1 OLAP Analysis

OLAP cubes and fact tables along with an OLAP thin client help the user drill down into each KPI to determine the root cause of a problem. Figure 5 illustrates a drill-down provided for the tenders rejected KPI.


present

TimeDim	Carrier	Tender Issued	Tender Accepted	Tender Rejected	% Tender Rejected
YearByMn-2003-Jun	EASTEXPLTL	24	10	12	54.545
	Carrier	24	10	12	54.545
YearByMn-2003-Jul	MWREGTL	26	21	7	25
	Carrier	26	21	7	25
YearByMn-2003	EASTEXPLTL	24	10	12	54.545
	MWREGTL	26	21	7	25
	Carrier	50	31	19	38
AllYearsByMn	EASTEXPLTL	24	10	12	54.545
	MWREGTL	26	21	7	25
	Carrier	50	31	19	38
AllYearsByWk	EASTEXPLTL	24	10	12	54.545
	MWREGTL	26	21	7	25
	Carrier	50	31	19	38
TimeDim	EASTEXPLTL	24	10	12	54.545
	MWREGTL	26	21	7	25
	Carrier	50	31	19	38

Figure 5: OLAP drill-down for tenders rejected.

4.2 Advanced Visualization

As a complement to the OLAP drill-down capabilities, the dashboard also provides advanced visualization capabilities. This functionality is built on a visualization library, Opal, that was developed by IBM Research. Opal provides a variety of ways to view multidimensional data, as well as capabilities for transforming and manipulating data. In SaRTM, Opal is being used in a client/server configuration. This allows the data to remain on the server, while users view and interact only with the resulting images.

Users interact with data by coloring subsets of points using the mouse. All views of the data currently shown are then automatically linked using the selected color. The dashboard visualizations give the user the ability to determine correlations and relationships.

As an example, consider a violation of the transportation cost variance KPI. This would imply that one or more carriers have exceeded their defined thresholds. While it is of interest to know which carriers were involved, the user may also be interested in seeing, for example, the distribution of cost variances of the violating carriers with respect to other carriers, or may be interested in determining whether certain locations are associated with cost variance more than other locations.

Figure 6 illustrates this (simulated) scenario. A scatter plot of the cost difference vs. the planned cost is shown for the selected time bucket of "month to date". The user has interactively colored all of the points with a large cost difference red, and all the points with a smaller cost difference green. The companion images show the proportion of shipments per carrier, per from-location, and per to-location which have been colored. The user can see, for example, that one carrier (SEEXPLTL) accounts for all of the red (high cost variance) points, and two other carriers (MWREGTL and NATIONWIDETL) have much higher than typical numbers of shipments with cost variance. At the same time, the user can see which locations are generally associated with the cost variances. Alternatively, the user could have colored a particular location, and seen the distribution of cost for that location. This analysis might also reveal trends in the business which would not be as obvious by looking at a textual representation of the data. Figure 6 also reveals that shipments whose planned cost was low have a higher probability of cost variance that the higher planned cost shipments. Analyzing the data over multiple time horizons might reveal different trends which could result in making significant changes in the way the business operations are executed. In this way the overall character of the data can be easily investigated.

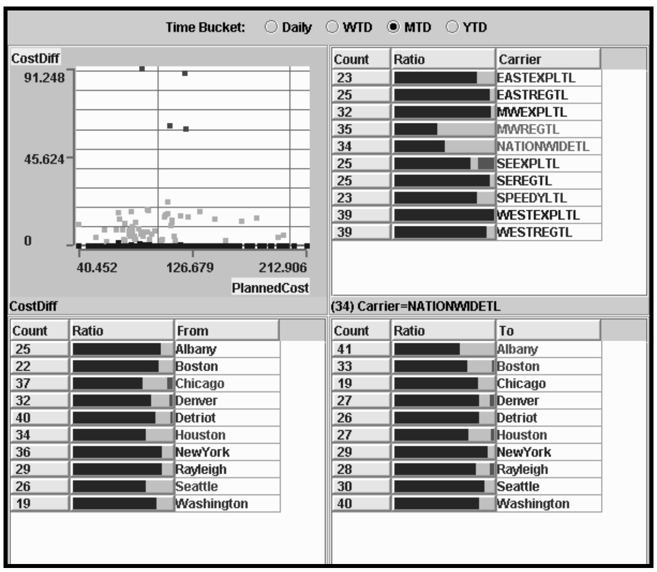


Figure 6: Transportation cost variance analysis.

5 Conclusion

In large organizations, huge amounts of data are generated and consumed by business processes. Business managers need up-to-date information to make timely and sound business decisions. In this paper, we introduced visualization techniques in context of a Sense-and-Respond system that help business users in interpreting business information and making faster and more well-informed decisions. By leveraging data collection, aggregation and correlation capabilities, our data visualization techniques provides the ability to target precise investigations and pinpoint specific trends and linkages over time through many visual display options including: pie charts, line charts, bar graphs, time tables, scatter plots and histograms. These types of anomalies alert organizations to irregular behavior so they may proactively identify potential threats and take action immediately.

The SaRTM system described in this paper is the second pilot of the Research Sense-and-Respond architecture. The first pilot is currently in progress at the IBM Microelectronics plant in Burlington, Vermont.

6 Acknowledgement

The authors would like to acknowledge the following people: Scott Simmering, the executive sponsor of this work; Karen Butner, Jan Browder, Kevin Stang, Henry Chang, Markus Ettl and Grace Lin who made significant contributions to the Research Sense-and-Respond architecture and Transportation Management System design.

References

[1] Haeckel, S. and A. Slywotzky 1999. Adaptive Enterprise: Creating and Leading Sense-and-Respond Organizations, *Harvard Business School Press, Cambridge, MA.*

[2] S. Kapoor, S. Buckley, "A Sense-and-Respond Approach for On-Demand Outsourced Supply Chain Management", *2003 Academy Conference on E-Business On Demand*

[3] McCormick, B., DeFanti, T., and Brown, M., "Visualization in Scientific Computing," *Computer Graphics, Vol. 21, No. 6, Nov. 1987, pp. 1-14.*

[4] Card, S., Mackinlay, J., and Shneiderman, B., "Readings in Information Visualization: Using Vision to Think," *Morgan Kaufman Publishers, 1999.*

[5] Codd, E.F., S.B. Codd, and C.T. Salley, "Providing OLAP (On-Line Analytical Processing) to User-Analysts: An IT Mandate", *E.F. Codd & Associates, 1993.*

[6] Kimball, R., Reeves, L., Ross, M., Thornthwaite, W. "The Data Warehouse Lifecycle Toolkit", John Wiley & Sons, 1998.