



TIBCO Live Datamart: Push-Based Real-Time Analytics

ABSTRACT

TIBCO® Live Datamart is a new approach to real-time analytics and data warehousing for environments where large volumes of data require a management by exception approach to business operations. TIBCO Live Datamart combines techniques from complex event processing (CEP), active databases, online analytic processing (OLAP), and data warehousing to create a live data warehouse against which continuous queries are executed. The resulting system enables users to make ad hoc queries against tens of millions of live records, and receive push-based updates when the results of their queries change.

TIBCO Live Datamart is used for operational and risk monitoring in high-frequency trading environments, where conditional alerting and automated remediation enable a handful of operators to manage millions of transactions per day, and make the results of trading visible to hundreds of customers in real time. TIBCO Live Datamart is also used in retail to track inventory, and in discrete manufacturing for defect management.

INTRODUCTION

Business intelligence, analytics, and data warehouse technologies are deployed in every area of business to support strategic decision-making. These technologies allow users to access large volumes of business data, query the data, and visualize results. Other systems are limited by the underlying data management infrastructure, which is nearly always based on periodic batch updates, forcing the system to lag behind real-time events by hours and often a full day. Other systems nearly always require pull-based querying, so new results are delivered only when a client requests them.

Out-of-date results generally require operational decision support that's tightly coupled with the applications in use by operational staff. However, tightly coupled decision support is less flexible, and less able to integrate data across multiple applications. In heterogeneous environments, operational intelligence requires visibility across dozens of production systems. TIBCO Live Datamart addresses this need by removing the two major limitations of traditional analytics: it delivers continuously updating live data and it provides push-based query results.

OPERATIONAL INTELLIGENCE

In modern financial markets, trading happens in milliseconds, and trades happen thousands of times per second. When something goes wrong with infrastructure, bad trades can pile up in seconds, with millions of dollars of exposure. Problems with trade flow like stuck orders or underperforming orders can be the result of faulty internal systems, or cut across several trading firms, markets, geographies, and even asset classes.

When protecting a firm's interests, customer relationships, and liabilities, things like rapid analysis, investigation, notification, and remediation of problems is critical. If you can detect a problem quickly and act fast enough, you can often avoid loss and even find new revenue opportunities. Furthermore, the firms with the timeliest information about adverse conditions create goodwill in their customer base.

In order to understand problems as they occur, operators need analytics-driven notification systems that identify problem signatures, put those notifications in context, and empower the operator to analyze data across multiple systems in seconds. Ideally, the same system that empowers operators to understand problems can be exposed to customers, giving them visibility into the infrastructure on which they are trading.

TIBCO LIVE DATAMART SOLUTION

TIBCO Live Datamart is built on top of the TIBCO StreamBase® Event Processing Platform and benefits from the CEP capabilities of that platform. The major components of TIBCO Live Datamart (see Figure 1) are as follows:

- Connectivity
- Continuous query processing
- Aggregation
- Alerting and notification
- Client connectivity

Connectivity is based directly on other TIBCO products, such as TIBCO StreamBase and TIBCO ActiveSpaces®. In addition, any TIBCO Live Datamart server can connect to any other TIBCO Live Datamart server in a federated system and pass through the remote tables as its own.

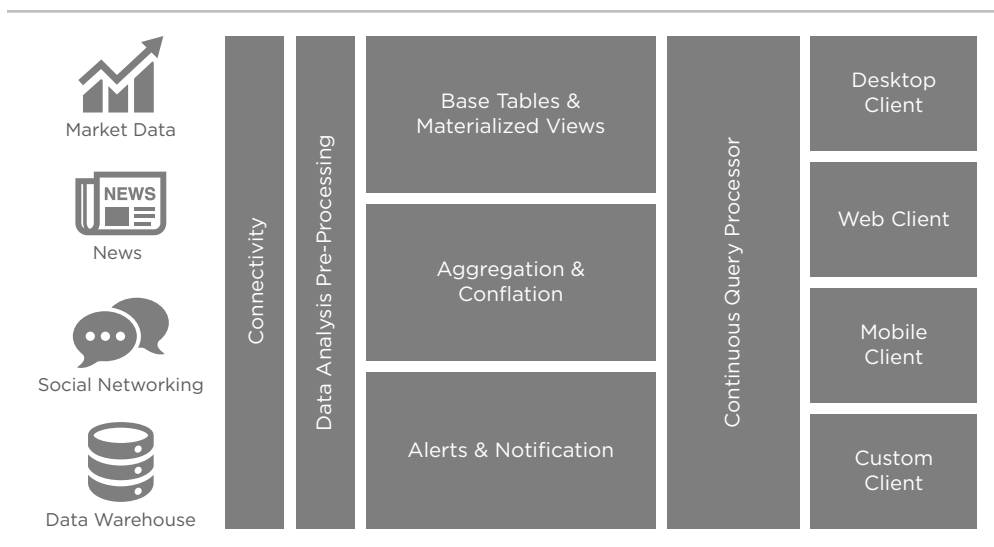


Figure 1: TIBCO Live Datamart Architecture

The Live Data Table serves as the unit of organization of every TIBCO Live Datamart project. The live data table continuously updates as new data is sent to the server. A live data table thus differs from a traditional data table in that its contents are always up-to-date.

The TIBCO Live Datamart Continuous Query Server (CQS) is at the heart of the system, efficiently managing thousands of concurrent queries against a live data set. The goal of the CQS is to take a query predicate, and the current contents of a live data table, and return a *snapshot* containing the rows that match the predicate. The processor then uses push-based *continuous updates* as the values in the table change. Updates include:

- Inserts
- Deletes
- In place updates of rows
- Updates that bring new rows into the view
- Updates that take rows out of the view

A key design principle of the CQS is to efficiently support large, pre-joined, de-normalized tables typically found in data warehousing and manage data in a main-memory database.

The CQS operates by indexing both queries and data, and by distributing records across multiple CPU cores. Snapshot queries are executed using indexed in-memory tables, with multiple threads simultaneously scanning different ranges of the index. Continuous updates are delivered using indexing, but in this case the queries are indexed, not the data. Given a query predicate such as “*symbol=‘IBM’ && quantity > 1000*”, the query might be placed into a table of registered queries indexed by symbol, under the key “IBM.” When an update to a record comes in, the pre- and post-image of that record are checked against this table to see if it matches the key “IBM.” Given a match, the rest of the predicate (i.e. “*quantity > 1000*”) is checked and an update is dispatched to clients of that query.

- a) Dynamic Aggregation
- b) Action
- c) Live Visualization
- d) Alerts

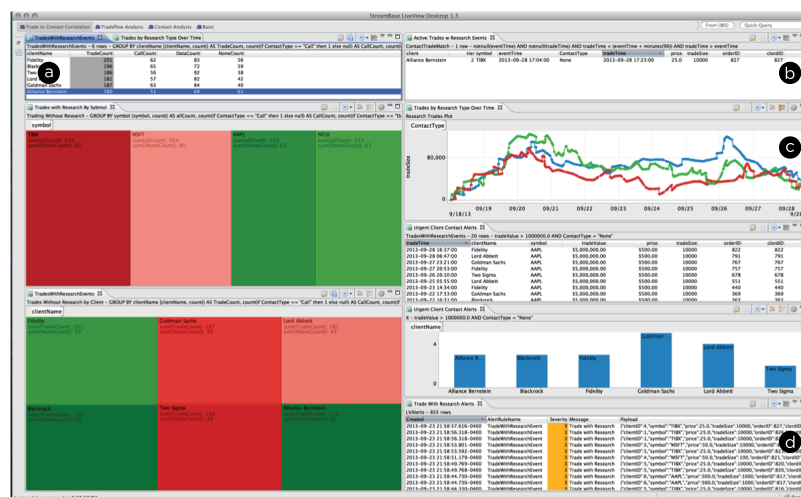


Figure 2: LiveView Desktop Interface

Materialized Views are the basis of complex analysis in TIBCO Live Datamart. Materialized views in TIBCO StreamBase LiveView are pre-computed, and then continuously or periodically updated using calculations based on one or more live data tables or on other materialized views. Materialized views include aggregations, joins, or alerts.

Aggregations are author-time or ad-hoc. *Author-time aggregations* can be configured into a live data table based upon common aggregations that users will want to query. A single table can be configured with multiple aggregations. Ad hoc aggregations are issued in a query. Both types of aggregation leverage the full power of the built-in query language, LiveQL, which includes not only SQL-like expressions, but additional query modifiers for streaming data.

Joins, such as joining pricing information with positions, can be driven by any CEP logic, which includes integrations with other languages like Java, Matlab, or R, leading to nearly infinite flexibility.

Alerts enable management-by-exception. The system will alert operators of opportunities or impending problems, allowing them to spend most of their time responding to these alerts.

Alerts are defined by conditions against a base table, and are another kind of materialized view, with rows matching the conditions being persisted in alert tables until the alert is cleared in some way. Some alerts, such as stuck orders, will clear automatically when the problem is resolved. Other alerts, such as an excessive number of market data anomalies, can only be cleared manually.

Alerts then drive notifications, messages delivered to a user through some interface. Notifications can come complete with context and recommended remediation. Context is enough information for the operator to understand the notification, generally the results of queries that the operator would be expected to need. Context might also include historical as well as live information.

Alerts can also feed into case management. Feedback from the user interface or other system can notify the server when an alert is being handled, who is handling it, and when it is resolved. Resolution conditions can be controlled manually by the user, or with automation, such as when the system detects the alert no longer applies. Alerts are auditable and may feed other systems for case or ticket tracking.

View Desktop and Web environments connect through the TIBCO StreamBase LiveView RESTful client layer. This layer is responsible for authentication and authorization, dispatching queries to appropriate query processors, managing registration of queries, and buffering results for delivery to occasionally connected clients.

This connectivity then drives a TIBCO StreamBase LiveView® Desktop environment as well as a StreamBase LiveView web dashboard, based on HTML5.

PERFORMANCE AND SCALABILITY

Delivering live data to operators requires several key performance metrics. The data volumes must be handled, and all computation completed, so that a complete and consistent picture is presented. The latency must be minimal, so that data is truly live. Furthermore, the queries required by users must be serviced efficiently and accurately in the face of changing data volumes, changing analysis, and an increasing number of sophisticated operators.

Single node performance is required to make efficient use of individual machines. LiveView performance is driven by the capacity of the continuous query processor and in particular the ability of the processor to scale across multiple cores. Both snapshot and continuous queries are partitioned and executed in parallel to get results faster and minimize contention between queries.

Performance results on a single box (24 core, 128 GB RAM server, 2011 vintage) are as follows:

Row Data	Financial, 100 fields, 2 KB
Total Data	30 million records
Queries	1,000 queries from 100 users
Throughput	50,000 updates per second
Latency	100 milliseconds

Scale out performance is based on sharding. Because the architecture of TIBCO Live Datamart uses single table queries and simple predicates, these queries naturally shard. Similarly, the event-based architecture for data preparation makes it easy to spread data across multiple nodes, or to create replicas of data for easy scale out.

TIBCO Live Datamart deployments can scale to tens of nodes today, with even more planned for the future. Partitioning can be by data set, by user, or even by breaking up a single data set or single user across multiple machines. This offers maximum flexibility in environments where data volumes vary and scaling can be critical.

BENEFITS OF LIVE INTELLIGENCE

In deployments of the TIBCO Live Datamart platform, three principal benefits have been realized: improved responsiveness, management by exception and correlation, and trust through radical transparency.

Improved responsiveness isn't just about analytics-driven, proactive alerts. It also comes from the context of the condition, and the operator's ability to use querying and analysis to understand the problem. When views are updated live, an operator's perspective on a situation is always in sync with the actual systems in question. This means they can see the results of their interventions as they go. This prevents over-correction and allows the operator to move on to the next problem as soon as the situation is resolved.

Management by exception uses correlated, real-time events from a number of large systems to free operators from watching well-behaved systems. This means they can take initiative on systemic situations that may have gone undetected without live, real-time analytics. For example, if the main systems are correctly functioning, operators might identify client orders that are underperforming in the market and recommend changes to improve their execution. Freeing skilled operators from low-value monitoring allows them to take initiative, improving customer relationships and profitability.

RELATED WORK

There are several alternatives to building a system as we have described here. They include:

Traditional analytics and data warehousing, the limitations of which are described above, are not able to keep up with the volumes of data or the expected latency. Their architectures are designed for static data sets, rather than live, changing data.

General purpose CEP can be used to build out similar alerting logic, driving notifications of problems. This system is effective, and in fact the StreamBase event processing platform has been used in this way. However, such systems lack support for ad hoc queries, and for managing continuous result sets as required to maintain up-to-date visibility into operational problems.

Messaging-based solutions using publication-subscription concepts have been another approach. In these systems, data items are divided among named topics, or topic names are derived from primary keys. A separate query server may be used to tell clients which topics to follow. This architecture often limits the level of indexing and complexity of predicates, forcing clients to consume excessive data. It can also fail to provide end-user visualization and ad-hoc query capability.

Custom software on a data grid platform, such as Oracle Coherence, has been the most effective alternative to date. A system developed and publicized by the Royal Bank of Scotland has similar functionality to TIBCO StreamBase LiveView and was developed in this way. However, the system is not as general purpose and requires a great deal of custom software development.

CONCLUSIONS

Traditional analytics and data warehouse systems have operated by design with out-of-date data. In TIBCO Live Datamart, we show that a compelling system can be developed without these limitations, giving users access to live, accurate data, with associated benefits in trust and responsiveness.



Global Headquarters
3307 Hillview Avenue
Palo Alto, CA 94304
+1 650-846-1000 TEL
+1 800-420-8450
+1 650-846-1005 FAX
www.tibco.com

TIBCO Software Inc. is a global leader in infrastructure and business intelligence software. Whether it's optimizing inventory, cross-selling products, or averting crisis before it happens, TIBCO uniquely delivers the *Two-Second Advantage*®—the ability to capture the right information at the right time and act on it preemptively for a competitive advantage. With a broad mix of innovative products and services, customers around the world trust TIBCO as their strategic technology partner. Learn more about TIBCO at www.tibco.com.

©2014-2015, TIBCO Software Inc. All rights reserved. TIBCO, the TIBCO logo, TIBCO Software, ActiveSpaces, StreamBase, and StreamBase LiveView are trademarks or registered trademarks of TIBCO Software Inc. or its subsidiaries in the United States and/or other countries. All other product and company names and marks in this document are the property of their respective owners and mentioned for identification purposes only.
01/20/15