COMPARISON OF SMARTDATA FABRIC® WITH TIBCO® DATA VIRTUALIZATION   REVISION 2.0

Introduction

Tibco® Data Virtualization (fka Cisco Data Virtualization and Composite Software) is probably the closest competitor WhamTech SmartData Fabric® has. From their sales and marketing material, the Tibco solution seems very similar to WhamTech, however, Tibco's approach is fundamentally different from WhamTech's. The reason for Tibco being considered closer to WhamTech than other data virtualization, federation and integration vendors, is that Tibco assumes a distributed approach similar to WhamTech, versus a centralized approach of most, if not all, other data virtualization vendors.

Fundamental Difference in Approaches

Tibco is a combination of a middleware product and conventional federated adapters, similar to other conventional data virtualization vendors, e.g., Denodo®, IBM® InfoSphere®, Oracle® Fusion Middleware and Red Hat® JBoss® Data Virtualization. Queries are submitted by applications to the middleware, from the middleware to the conventional federated adapters, and lastly submitted to and executed by data source systems. Query success and performance are therefore limited by data quality and data source system capabilities, and query loads are imposed on and absorbed completely by data source systems. As with many data virtualization vendors, that recognize the dependency on data source systems, Tibco can improve data quality on results and add value by caching result sets and making them available to subsequent queries for specific results data.

WhamTech SmartData Fabric® EIQ Adapters™, in contrast to Tibco and all other data virtualization, federation and integration vendors, query WhamTech's own independent indexes, indexed views and business objects that are built and maintained, usually, in near real-time using cleansed, transformed and standardized data. This obviates the need for much of the work that conventional federated adapters and middleware need to perform to overcome both data and data source system deficiencies. And alleviates almost all loads on data source systems, significantly improves query success and performance, and allows far more capabilities.

Tibco in General

Tibco is primarily a federated data access system with middleware, conventional adapters and results cache:

G1.   Tibco relies on conventional federated adapters to submit queries to data source systems for execution.

G2.   Tibco builds materialized indexed views of past query results data in cache for subsequent queries – elements of a distributed data warehouse/data mart.

G3.   Tibco establishes PK-FK relationships within data sources and across data sources based on schema-level keys, e.g., LAST_NAME in one data source = FAMILY_NAME in another.

G4.   Tibco performs ETL operations to applications, data warehouses and data marts that can include cleansing, transforming and standardizing data AFTER results data retrieved.

G5.   Tibco OEM their products to Informatica, among other vendors.

Tibco Advantages Compared to WhamTech

A1.   Tibco is an established vendor with a large customer base.

WhamTech has worked in different type of projects, including specialist products for Web and enterprise search, and eDiscovery, and for large federal system integrators on DOD and intelligence community projects. While WhamTech has had commercial customers in the past and is still working with large federal system integrators, in the last few years, WhamTech has primarily focused on commercial markets and is fast-gaining a reputation for being disruptive, particularly in insurance, banking and healthcare where it is difficult or not possible to copy all data to a single data store, or by copying, introduce latency.

A2.   Tibco probably has 80% solutions for standard applications such as SAP, Siebel, Oracle E-Business Suite, PeopleSoft and salesforce.com.

WhamTech is building its portfolio of data sources it can work with, including most common data warehouses/databases, mainframe data files, office files, Web pages, application data, email, social media, Big Data, streaming and IoT devices. WhamTech can work with almost any data source and almost any readable data, regardless of the data source or data within a data source being structured, unstructured or semi-structured.

A3.   Tibco does not need to build and maintain indexes for data sources, as it relies on data sources to execute queries through conventional federated adapters. Once results are obtained, they can be cached and indexed for subsequent queries in materialized indexed views.

WhamTech usually does have to build and maintain its own indexes and execute its own queries, but that is a positive capability for many reasons, as listed in the next section "Tibco Disadvantages Compared to WhamTech".

A4.   Tibco does not rely on data source systems once cached result materialized indexed views are built, as long as the results data does not change in the data source, but that introduces the problem of cache management. See separate white paper "#WhamTech Competitor Comparison vs. Denodo, Red Hat, Tibco and Other Conventional Federated Data Access Vendor's Use of Cache".

WhamTech generally returns to data sources to obtain results, but can store data locally and/or cache latest results data if needed. WhamTech prefers to leave data in sources, as this is the main benefit, and maintains indexes instead of the more difficult task of maintaining a results cache. While locally stored data is simple to maintain, results cache is not, depending on how source data is changing. Results cache is particularly difficult to maintain in the Big Data world of large volume and high velocity of data, unless results consist of historic data that does not change. Again, see the above-referenced white paper for a more in-depth discussion of the reasons for and the problems of cache.

Tibco Disadvantages Compared to WhamTech

D1.   Tibco does not solve the initial data source access problem that plagues conventional federated data access – potentially, poor data quality, governance and security, and limited indexes and query processing. This directly impacts query success on data sources, query performance and scalability.

WhamTech discovers, profiles, identifies, classifies, secures, cleanses, transforms and standardizes data for indexes and results, has its own independent indexes and scales for high performance distributed parallel query processing on these indexes.

D2.   Tibco makes copies of potentially confidential/sensitive data in indexed cache/databases.

WhamTech leaves most data in sources until results need to be retrieved, subject to the appropriate role-based access control, and dynamic data masking, tokenization and encryption.

D3. Tibco imposes significant query loads on data sources, and query success and performance are completely dependent on data source systems.

WhamTech almost completely absorbs the query processing load. There is a small data source system load for sequential reads of final and sometimes interim results data, but this is minimal (< 5%) compared to full query processing. WhamTech has built out systems capable of simultaneously indexing and query processing 10s of billions of records and 10s of TBs per day.

D4. Tibco has no obvious mechanism to update the results cache, although it is assumed that some mechanism is in-place.

WhamTech, in a near real-time index update architecture, has the latest index data and results data available, with several options to account for any latency differences.

D5. Tibco has no obvious capabilities for derived value indexes and indexed views such as pre-joins, pre-aggregations, pre-calculations, pre-joins, fuzzy match, text search, information geometry (unstructured data categorization), entity extraction, and Link Indexes for link mapping, link analysis/graph database and link visualization.

WhamTech offers all of the above capabilities. One of the benefits of these derived value indexes and indexed views is that these can be maintained in near real-time and automatically used by WhamTech SmartData Fabric® EIQ Adapters when part of a query. This avoids relying on potentially stale results cache or the adapter requesting raw data from a source or sources and performing on-the-fly joins, aggregations, calculations or view materialization, which can considerably slow down reporting, BI and analytics.

WhamTech future development will allow the automatic creation of some of the derived value indexes and indexed views when their use would be beneficial to a query. Such indexes and indexed views could be deleted if not reused in a certain timeframe.

D6. Tibco has so-called real-time monitoring, but that represents a constant polling of data sources, and therefore query load, to check if any relevant changes have occurred.

WhamTech usually relies on Changed Data Capture (CDC) through transaction logs to update indexes, and monitors these updates to automatically update operational/real-time monitoring dashboards, and/or send alerts/notifications to subscriber/subscribing applications. The combination of monitoring, event processing, stored procedures at the SmartData Fabric® level and Business Process Management (BPM) software allows for extraordinary capabilities usually addressed by additional separate applications.

D7. Tibco has no form of Link Indexes that allow record-level relationship data to be combined with content data.

WhamTech uses combinations of queries on content indexes and Link Indexes to pre-calculate internal and external joins, degrees of separation queries, link mapping, link analysis and ontological data model representations, as well as enabling virtual solutions, such as CDI-MDM/MPI, CRM and other solutions.

D8.   Tibco seems to present potentially complex data models to applications, requiring significant schema mapping work.

WhamTech's approach is to import reference data models, such as ACORD, HL7, FHIR, NIEM and XBRL – same data from multiple data sources are mapped to the same reference data views and business objects. The reference data view can, in-turn, be accessed as though (i) data objects (Web/data services and ontological view), (ii) virtual relational view and (iii) Big Table/single flat file view. Various associated query languages can be used with conversion to SQL, e.g., OQL, SPARQL, PL-SQL and NoSQL.

D9.   Tibco, in the worst-case scenario when a data source is unavailable, can only access cached results data.

WhamTech can invert any and all data from indexes, assuming that all needed data is indexed. WhamTech indexes and/or own or third-party storage can also be used for archive and/or time-slice data. This last approach is used to store highly compressed and aggregated IoT device data that is accessed using standard drivers, APIs and Web/data services, and SQL.

D10. Tibco is probably very expensive to scale using such an approach, plus, installing and configuring such a system to overcome data and data source system deficiencies are probably manually intensive and therefore very expensive as well. Typically, at least five times as much cost over an extended period is needed to achieve the basics associated with conventional data virtualization, federation and integration solutions, because of all the data and query issues that are attempted to overcome.

WhamTech inherent approach overcomes many of the time consuming and therefore expensive processes needed for Tibco to work.

Denodo, Tibco and other DV vendors, and WhamTech Working Together?

This question has been asked a few times. Conventional DV vendors could replace conventional federated adapters with WhamTech's EIQ adapters, but that might require an internal shift away from their own perceived competitive advantage, which is usually focused on query optimization. WhamTech has experienced this before with a large middleware company that realized that WhamTech obviates the need for much of what they leverage and market, leading to an internal not invented here (NIH) rejection, even with customers expressing dissatisfaction with their product and approach. One such large middleware company sold for a relatively small amount shortly after WhamTech was in discussions with them.

For more information, please contact: Gavin Robertson, CTO & SVP, gavin.robertson@whamtech.com, (972) 991-5700 x706