

EVENT DRIVEN DATA INTEGRATION AN ANALYSIS

Rajeshkumar Myilsamy

St.peter's university, Chennai

rajeshborn4u@gmail.com

Abstract: Business happens in real time and is always event-driven. Understanding, capturing and integrating events provides a competitive advantage through enhanced customer service, greater business agility and greater business visibility. The opportunity exists today to become event-driven and event-aware to gain a distinct competitive advantage. The technology exists today to capture and integrate business events, regardless of the complexity of the event or technical environment. In this paper we propose a architecture for Real Time Data Integration using J2EE technology.

Keywords: Architecture, Complexity, Capturing, Event Driven, and Integration

INTRODUCTION

The main objective of the Real time data Integration is delivering trusted information wherever and whenever needed, in line & in content to specific people, applications & processes that allow organizations to gain a competitive advantage in their market places. Companies need to respond more quickly, gain operational efficiencies and deliver superior customer service by keeping a real time pulse on their business. A real time data integration strategy ensures accurate data flows across the enterprise so that organizations can make quick decisions on pricing, shelving, service and product mix, based on the latest information. Managers can analyze critical business data throughout the day to target their marketing efforts, improve up-selling strategies' and better service their customers, no matter what service they are in. We are given a complete and wholesome perspective of an organization, only when all its entities come together. It is impossible for individual entities to display this togetherness. There are two types of data Master data & transactional data. Master data is static in nature. Transactional data is the data collected by the operational systems to control different transaction. Master data objects may include business entities such as customer, product, locations, suppliers, account information etc. Master data represents the business objects around which transactions are executed.

To make sound decisions and comply with governmental reporting requirements, an organization must first establish a solid data foundation. This foundation must combine historical data with current values from operational systems in order to provide a single version of the truth that can be then used to identify trends and predict future

outcomes. Data integration technology is the key to consolidating this data and delivering an information infrastructure that will meet strategic business intelligence (BI) initiatives and tactical and governmental reporting requirements. Data integration is the enabling technology for providing trustworthy information, enhancing IT and end-user productivity, and helping organizations achieve and maintain a competitive edge. Data integration enables mid-size and large

organizations to effectively and efficiently leverage their data resources in order to satisfy their analysis and reporting requirements.

LITERATURE SURVEY

The management of key organizational data has always been important. Knowing who your customers are, what products and services you offer, and what the arrangements or accounts you have with your customers and suppliers is fundamental to the operation of most organizations. Whether your organization is a bank, a retailer, or a government agency, there is a core set of such data that is used across the enterprise. It is used to open new accounts, to introduce new products to the market, and to determine what products to offer Customers. This data is called master data.

Master data is some of the most valuable information that a business owns. It represents core information about the business such as customers, suppliers, products, and accounts and the relationships between them. Each of these domains of master data represents information that is needed across different business processes, across organizational units, and between operational systems and decision support systems. In essence, master data defines an enterprise. Master data captures the key things that all parts of an organization must agree on, both in meaning and usage. For example, it is important that all parts of an organization share an understanding of what defines a customer, which customers exist, where customers are located, and what products they have purchased or have been offered.

A common understanding is useful both to prevent bad things from inadvertently happening such as a bill getting posted to the wrong address and to provide an opportunity for significant business benefits such as improving the ability to sell complementary products to customers. Master data is important in both operational and analytical environments. Many operational business processes touch master data for example, introducing a new product to the market, signing up a new supplier, and adding a new phone service to a customer account.

All of these processes touch many different application systems that must all share a core set of information about products, suppliers, and customers. For the business process to execute properly, this master data must be accurate and consistent. Analytical systems have similar requirements master data often forms the key dimensions and hierarchies used for reporting and analysis of key business data. Increasingly, analytics are also being applied within operational business processes to better monitor and optimize business transactions. Trustworthy data is a fundamental ingredient of meaning full analytics. By integrating data across disparate operational systems, an organization can increase the effectiveness of its data access and analysis capabilities. Integrating data from multiple systems creates an environment where the whole is worth far more than the sum of the individual parts.

GOAL

In this paper we present a Real time Data Integration approach to provide a Master Data for the whole enterprise wide. Here we provide a real time data integration across heterogeneous databases those span across the whole organization. The changes to the source database are logged to a custom changed data table. We use the concept of change data capture where in we capture the changes in the source database and scrape it to the target database. This can be implemented in different ways i.e., using continuous mirroring, refresh etc.

In this implementation, after the changes are captured from the source database the changed data is given to the Event Server where we generate a XML message for each changed data. This XML message is than put into a message queue, where the target side application picks the message from this queue and transforms this message into the services required for that application that modifies the target applications database that reflects the changes done on the source side.

This real time Integration may include databases that may hold data about different domains of an enterprise; at sometimes there will be need to integrate the data of different domains to carry out some strategic business decisions.

Why is Data Integration Important?

To be successful, the organization large or small must run its operations effectively and efficiently, which requires the ability to analyze operational performance. If you can't see how you're performing, how do you know you are making the right business decisions? For an organization to thrive, or perhaps even survive, operations and analysis must work together and reinforce each other. This is especially important in small-to-medium size organizations, which in order to grow and expand need to focus their limited resources and take appropriate actions to build upon their successes while quickly identifying and resolving operational problems.

Without the entire picture, it's difficult to make sound and dependable business decisions. That's because good decision making requires a complete and accurate view of data. The ability to access and integrate all of your data sources is the start to getting the complete picture and the key to not compromising your decision-making process. Though your organization needs a complete view of operations, the data you need often resides in a variety of application systems that do

not necessarily all use the same database management system. Furthermore, these application systems may only contain current data values. They may not store prior data values needed to provide historical context and to discover trends.

Data integration allows an organization to consolidate the current data contained in its many operational or production systems and combine it with historical values. And the creation of a data warehouse (or, on a more limited scale, a single-subject data mart) facilitates access to this data. Collecting and consolidating the data needed to populate a data warehouse or data mart and periodically augmenting its content with new values while retaining the old is a practical application of data integration.

Approaches for Data Integration

An organization can integrate its data through a variety of methods including:

- Enterprise-class data warehouse implementations hosting massive amounts of historical data
- Departmental or subject-level data marts focused on a single organizational unit or functional area
- Operational data stores containing current values of data extracted from several operational systems
- Enterprise information integration (EII) deployments that provide a direct, real-time view of data residing in multiple operational systems

PROPOSED ARCHITECTURAL DESIGN

This architecture has mainly 3 components the history manager module, the CDC agent & the JMS queue, which forms the interface between the event server and the target application.

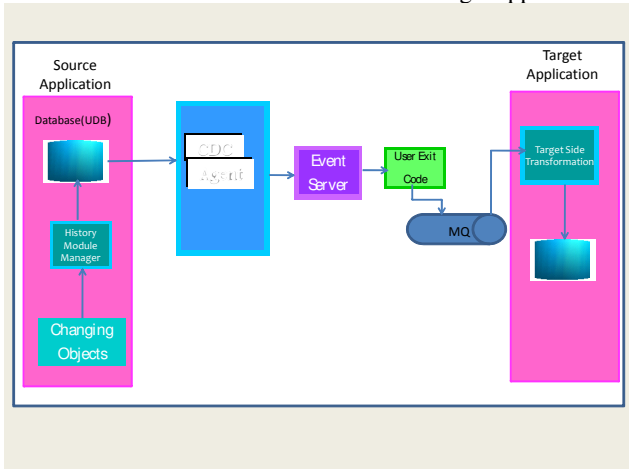


Figure 1. Proposed Architecture for the Integration

In the proposed architecture for the integration of the source and the target database the data flow from the source and the target is only considered. The source and the target may be heterogeneous databases. From the business paradigm, ensuring minimum risk requires identification and integration of only those events that directly impact agility, productivity and profitability. Change Data Capture requires a minimum amount of CPU or memory, with relatively low impact on the performance of production systems or on organizational bandwidth. This is due to its log-based Change Data Capture (CDC) technology.

It is equipped and designed to work with multiple databases, multiple applications, across multiple platforms. Change Data Capture is able to recognize and deliver events in real time, without major delays in delivering unless dictated by the business. Change Data Capture operates similarly and seamlessly across any database or platform used by the business system. It is able to recognize and aggregate complex or discrete events, regardless of the source being DB2 on Mainframe, Oracle on HP/UX or SQL Server on Windows.

METHODOLOGY

In this methodology, the changes that have to be captured for the different database objects are identified on the source database. The integration code is written on the both the sides in JAVA to reflect the changed data on either side. On the source side once any modification happens to any of the identified database object the history manager module logs those changes to the custom database table that is created by that module itself. This table is present in the same database where the source application database is. The table consists of columns such as the id of the object modified, type of the object, type of the event log level, timestamp, zipped data of the changes. Once these changes are identified these changes are converted into the XML format and logged in the var char column of the custom database table in the zipped format. The CDC agent is configured by creating subscriptions such that it continuously mirrors any updations to the custom data base table to the target i.e., the event server. The event server is mainly used as the target here because it creates an XML message for each row of the custom database table and puts into the message queue. Each column of the custom database table is mapped into as a each element of the XML message generated by the event server. Once the message is put into the queue , a module on the target application is made to read the message from the queue using JNDI,then each of the XML message is parsed to get the element in the XML message that contains the zipped data. When this data is obtained it is unzipped then we get the XML format of the changed data, which is then later transformed into the required format of the target database by necessary mappings in the target side to the changed data got from the source side.

CONCLUSION

This paper provides an approach for the Real time data Integration of the heterogeneous databases. This can be deliverable as a bundled product for the stakeholders. This provides the large enterprises to make strategic business decisions to meet the current business requirements. Data integration and data quality solutions are the keys to achieving trusted information. While some organizations choose to develop their own in-house data integration solutions, those that use packaged software solutions can benefit from the vendor's expertise and experience in working with multiple, and sometimes esoteric data sources. This also frees up their staffs for more productive tasks that help gain a competitive advantage

REFERENCES

[1] Data Integration: Creating a Trustworthy Data Foundation for Business Intelligence

[2] Enterprise Master Data Management: An SOA approach for managing core Information.

[3] How Transformation Server/ES Can Be An Integral Part Of An Event-Driven Architecture And Service Oriented Architecture

[4] How IBM Information Server Change Data Capture can be an integral part of an event-driven architecture and service oriented architecture

[5]Java Server Programming, Java EE 5, Black Book

[6]Beginning XML,4th Edition-David Hunter,Jef Rafter, Joe Fawcett, Eric van der Vlist,Danny Ayers,John Duckett

[7] An Introduction to Real-Time Data Integration , *Mark Rittman*

[8] Towards Real-Time Data Integration and Analysis for Embedded Devices, Michael Soffner, Norbert Siegmund, Mario Pukall, Veit Käöppen

[9] Striving towards Near Real-Time Data Integration for Data Warehouses, Robert M. BrucknerBeate ListJosef Schiefer