FINANCIAL SERVICES: FRAUD MANAGEMENT

A solution showcase

TECHNOLOGY OVERVIEW

FRAUD MANAGEMENT REFERENCE ARCHITECTURE

This technology overview describes a complete infrastructure and application re-architecture for an organization that is considering a big data initiative around fraud detection and prevention.

INTRODUCTION

The 2013 "Visa Says Big Data Identifies Billions of Dollars in Fraud" article[1] in the Wall Street Journal highlighted advances in fraud detection and management with big data techniques. Visa estimated that its technologies helped identify at least $2 billion worth of annual fraud, and helped the company address those vulnerabilities before money was lost.

Financial services companies – with operations spanning retail banking and credit cards – must not only implement highly sophisticated and automated business operations, but must also ensure that their business activities are transparent to business process owners, auditors, and others.

In addition to managing new sources and more streams of data, they also face new regulations that create a greater need for compliance. IT leadership teams are pressured to not only support evolving business requirements, but also improve efficiency and manageability, contain costs, and speed products to market – all while ensuring security.

Let's consider the fictional, highly representative example of a firm drowning in big data – unable to effectively store, manage, and best direct large amounts of data. This example firm understands that data can be one of the company's most important assets, and also understands data can become a liability and cost when it's not properly managed.

This technology overview describes a reference architecture for a complete infrastructure and application re-architecture in an organization considering a big data initiative around fraud detection and prevention.

BUSINESS CHALLENGE

Credit card fraud and its related costs are one of the biggest business risks facing the consumer financial industry. There are different categories of credit card fraud, including application fraud, lost or stolen credit cards, counterfeit cards, and account takeovers.

The Federal Reserve recommends that action be taken to stop the abuse in progress and that businesses incorporate risk management practices to protect against similar actions in the future[2].

1 *http://blogs.wsj.com/cio/2013/03/11/visa-says-big-data-identifies-billions-of-dollars-in-fraud*

2 *http://www.phil.frb.org/consumer-credit-and-payments/payment-cards-center/publications/discussion-papers/2002/FraudManagement_042002.pdf)*

DATA METHODS

1. Data in motion: Streaming data that is being sent in real time

2. Data at rest: A phase where data lives after it has been collected in a form that follows enterprise data architecture and governance specifications

BUSINESS REQUIREMENTS

1. Integrate and cleanse data to get a complete view of any and all transactions that could signal potential fraud.
2. Predict cardholder behavior to provide better customer service.
3. Target customer transactions that raise security flags for personalized communications.
4. Deliver personalized communications the way customers want (e.g., through web, text, and email).
5. Track potentially fraudulent events from a strategic perspective.
6. Help provide a complete picture of high-value customers to help drive loyalty programs.

DESIGN AND ARCHITECTURE

The architecture needs to consider two broad data methods:

Data in motion is streaming data produced as a result of business interactions and then sent immediately to the information architecture in real time. Examples could be the result of purchases or feedback about services and could include credit card swipes, e-commerce tickets, web based interactions, or social media feeds.

The challenge is to assimilate huge volumes of data, filter it, reason from it, and send it to downstream systems (like business process management (BPM)) for processing by employees, when applicable. Integrating events with business rules ensures changes to business rules and regulations are incorporated.

Data at rest is a phase where data lives after it has been collected in a form that follows enterprise data architecture and governance specifications. This data needs to be assimilated or federated with pre-existing sources so that the business can query it in a read/write manner from a strategic, long-term perspective.

REFERENCE ARCHITECTURE

The key technology components of this architecture are:

- Information sources that could include both machine and human actors, transmitting thousands of real-time messages per second.
- A highly scalable messaging system to help bring these transmissions into the architecture, normalize them, and send them for further processing.
- A complex event processing (CEP) tier that can process these transmissions at scale to understand their relationship to one another, as defined by business owners or developers.
- Business process workflows that are created when data matches specific patterns that indicate potential fraud. These workflows follow a predefined process modeled by the business.
- Business-relevant data that can be kept for offline or batch processing using a Java™ Data Grid or a storage platform. (e.g., deploying Hadoop-oriented workloads to understand fraud patterns as they occur)
- A scale-out deployment approach, which helps as loads placed on the system increase over time.

MESSAGING BROKER TIER

The messaging broker tier is the first point of entry in the system, and hosts a set of message queues that initiate events from the message producer tier. This broker tier needs to be highly scalable while supporting a variety of cross-language clients and protocols (e.g., Java, C, C++, C#, Ruby, Perl, Python, PHP).

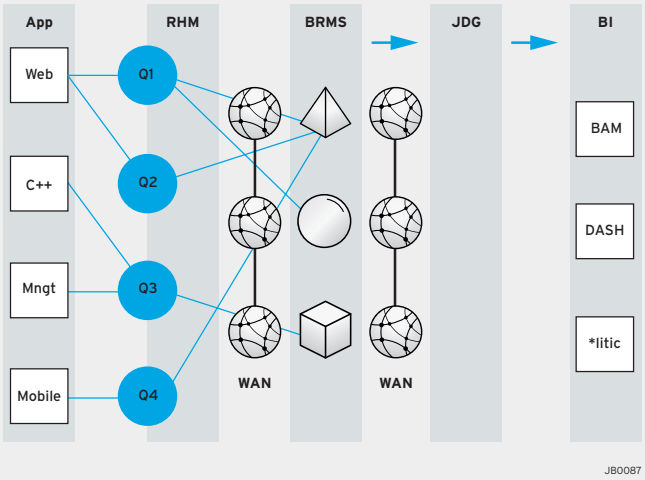


Figure 1. Complete architecture showing how transactions in the real world (shown on the left), irrespective of their channel of origin (web, mobile, physical use of a credit card at a customer outlet), can be processed to look for indicative patterns of fraud.

Using various messaging patterns to support real-time messaging, this tier integrates application, endpoints, and devices quickly and efficiently. The tier needs to be flexible so it can be deployed in various configurations to connect to customized solutions at every endpoint, payment outlet, partner, and device.

The choices that are available to implement this tier include:

- Red Hat® JBoss® A-MQ: This enterprise version includes support for the open source ActiveMQ broker.
- HornetQ in Red Hat JBoss Enterprise Application Platform: This full Java EE platform includes an embedded messaging broker capable of high-availability configuration and bridges to other messaging providers (e.g., Tibco Event Management System (EMS), or IBM Message Queue (MQ)).

CEP TIER

In this scenario, the CEP tier is an independent software module that is completely integrated with the rest of the platform. It runs on a horizontally scalable infrastructure that adds a set of features that:

- Provides full and native support for events.
- Selects a set of interesting events in a cloud or stream of events.
- Detects the relevant relationships and patterns among these events.
- Takes appropriate actions based on the relationships and patterns detected.

CEP allows the architecture to process multiple events, with the goal of identifying meaningful ones. This process involves:

- Detection of specific events.
- Correlation of multiple discrete events based on causality, event attributes, and timing.
- Abstraction into higher-level (i.e., complex or composite) events.

This ability to detect, correlate, and determine business relevance powers truly active decision-making capabilities.

BUSINESS PROCESS MANAGEMENT (BPM) TIER

The BPM tier is invoked for downstream handling as specific events are detected. BPM process and business rules can be defined by non-technical and technical users of the fraud detection platform, as shown in Figure 2.

The BPM tier spins up new processes that can be entirely automated or can include an employee (or employees) who process fraudulent events. Results of this process can include a variety of actions (e.g., a call to a customer by a call center representative, an update to a datastore that can be queried by a business intelligence application, etc.).

Business rules, CEP rules, and process definitions may be stored in a central repository and pulled into the application in real time. Therefore, assets can be changed outside of the typical software development life cycle. The rules are commonly stored in the Red Hat JBoss BRMS repository and pulled into the engine for execution at run time.

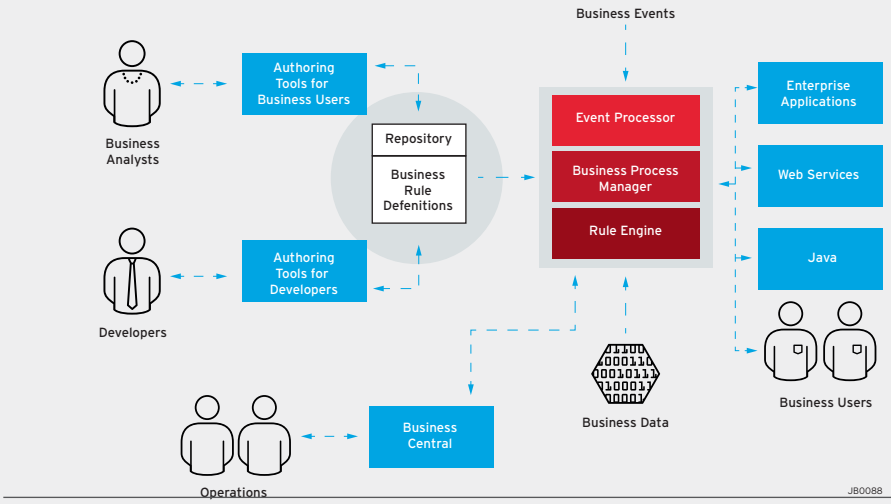


Figure 2. Business process management system (BPMS) with workflow, rules and complex event processing capabilities.

Alternatively, the assets may be embedded directly in a Java application archive (e.g., JAR, WAR, EAR) and pulled in through the Java resource classpath. The rules may also be pulled in from an external resource location (e.g., file system or web address).

STORAGE TIER

Broad needs for two distinct data tiers can be identified based on the above business requirements:

1. Because data needs to be pulled in near real time, accessed in a low latency pattern, and calculated, the design principle should be write many, read many with the ability to scale out tiers of servers.

   Java-based, in-memory data grids (IMDGs) are suitable for this use case because they support a high write rate. Red Hat JBoss Data Grid is a highly scalable, enterprise-ready implementation of a distributed data grid that gives users the ability to store, access, modify, and transfer extremely large amounts of distributed data. Further, Red Hat JBoss Data Grid offers a universal namespace for applications to collect data from different sources for all of the above functionality. Data grids can pool memory and scale across a cluster of servers in a horizontal manner.

   In addition, computation can be pushed to the tiers of servers running the data grid as opposed to pulling data to the computation tier.

   To meet the needs for scalability, fast access, and user collaboration, data grids support replication of data sets to points within the distributed data architecture. The use of replicas gives multiple users faster access to data sets and allows the preservation of bandwidth, since replicas can often be placed strategically close to or within sites where users need them. Red Hat JBoss Data Grid supports wide area network (WAN) replication, clustering, out-of-the box replication, and multiple language clients.

2. Storage for data ranging from next day to months to years, typically large-scale historical data, is the second data access pattern that needs to be supported. The primary data access principle here is write once, read many.

   This layer contains the immutable and constantly growing master data set stored on a distributed file system like Red Hat Storage Server and Hadoop Distributed File System (HDFS). Besides being a storage mechanism, the data stored in this layer can be formatted in a manner suitable for consumption from any tool within the Apache Hadoop ecosystem (e.g., Hive, Pig, Mahout).

IN-MEMORY DATA GRID

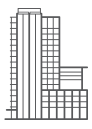
Data grids are commonly used in conjunction with BRMS to store:

- Business data that is necessary for events to correlate against.
- A stateful knowledge session between event processing batches for failover.

Red Hat JBoss Data Grid: A NoSQL datastore that provides the flexibility to store any type of data.

CONCLUSION

Applications that deal with dynamic business events like fraud detection need to be architected, designed, and developed with scalability, flexibility, and performance in mind. Enterprise open source provides robust building blocks that deliver tremendous business value while satisfying cost and performance considerations.



ABOUT RED HAT

Red Hat is the world's leading provider of open source solutions, using a community-powered approach to provide reliable and high-performing cloud, virtualization, storage, Linux, and middleware technologies. Red Hat also offers award-winning support, training, and consulting services. Red Hat is an S&P company with more than 80 offices spanning the globe, empowering its customers' businesses.

facebook.com/redhatinc
@redhatnews
linkedin.com/company/red-hat

NORTH AMERICA
1 888 REDHAT1

EUROPE, MIDDLE EAST, AND AFRICA
00800 7334 2835
europe@redhat.com

ASIA PACIFIC
+65 6490 4200
apac@redhat.com

LATIN AMERICA
+54 11 4329 7300
info-latam@redhat.com

redhat.com
#INC0202675_1214