GIGASPACES

Real-Time Analytics for Big Market Data with XAP In-Memory Computing

March 2015

Real Time Analytics for Big Market Data

Table of Contents

GIGASPACES

Introduction

The financial services industry is undergoing a fundamental transformation from a product-centric to a customer-centric model leveraging cost effective virtualization and low-latency processing technologies.

To keep pace with this new business paradigm, large banks around the world are modernizing their IT capabilities by adopting next generation solutions that enable them to achieve cost efficiencies, contain risks, meet regulatory compliance requirements and gain a comprehensive view of their customers.

Analytics applications demand higher throughput and lower latency than ever to meet the challenge of processing terabytes of data from various sources, which must be available to multiple parties, making system management increasingly complex. The variety and volume of data is overwhelming traditional classic database solutions and legacy batch-oriented systems that were not designed to meet today's real time analytic requirements. Faster hardware alone will not meet the challenging requirements of real-time data.

This paper will introduce GigaSpaces' XAP In-Memory Computing Platform as the solution for banking firms to keep pace with the speed of real-time while maintaining transactional integrity. Running within modern virtualization environments such as private cloud.

GIGASPACES

Main Industry Challenges

Financial Services firms are increasingly under pressure to make faster, yet accurate decisions based on a variety of data to effectively mitigate and monitor operational and compliance risk, as well as leverage the power of analytics into more day-to-day decision making.

As a result, analytics applications demand higher throughput and lower latency than ever, to meet the challenge of processing terabytes of data, from various sources, which must be available to multiple processes, making system management increasingly complex. Moreover, the ability to mitigate risk and make fast decisions requires gaining meaningful information from these vast data streams in real-time.

The demand for real-time calculations on real-time data for real-time critical decisions has become an imperative to achieve the following common goals:

- Gain insight as competitive advantage
- Achieve real-time decision making (e.g. instant loan approval)
- Strengthen fraud analysis
- Accelerate time to market
- Reduce CapEx and OpEx
- Create consistent and rich customer experience
- Improve business agility and flexibility to speed product innovation
- More easily meet regulatory compliance requirements
- Enrich data models with external data sources

To realize new opportunities, banks will have to focus on 4 key areas:
1. **Risk Optimization:** Achieve compliance objectives while mitigating operational risk, fighting crime and optimizing returns
2. **Simplification of operational and technological complexity:** Consolidate disparate sources of customer, product and transactional data. Organizing data as master data, operational data, analytical data accommodating streaming, real-time data and integrating existing data stores
3. **De-duplicate processes, deleverage and pursue componentization of technology:** Separation of data from applications

GIGASPACES

4. **Adopt transformational technology solutions that reinforce a simplified, streamlined and agile enterprise**: to balance growth, efficiency and business resiliency in order to support new business innovation.

Achieving Real-Time Performance with In-Memory Computing

In-Memory Computing (IMC) is a technology and an architectural style that enables extensive analysis of massive data sets, as well as extreme transaction processing at lightning speeds. These types of architectures rely primarily on RAM as well as CPU power to process the data, which in turn negates the need for rotating disk-stored data access. As posed by Gartner, In-Memory Computing provides transformational opportunities by optimizing the execution of hours-long batch processes into minutes and even seconds. Millions of events can be scanned in a matter of few tens of milliseconds to detect correlations and patterns pointing at emerging opportunities and data insights "as things happen."

GIGASPACES
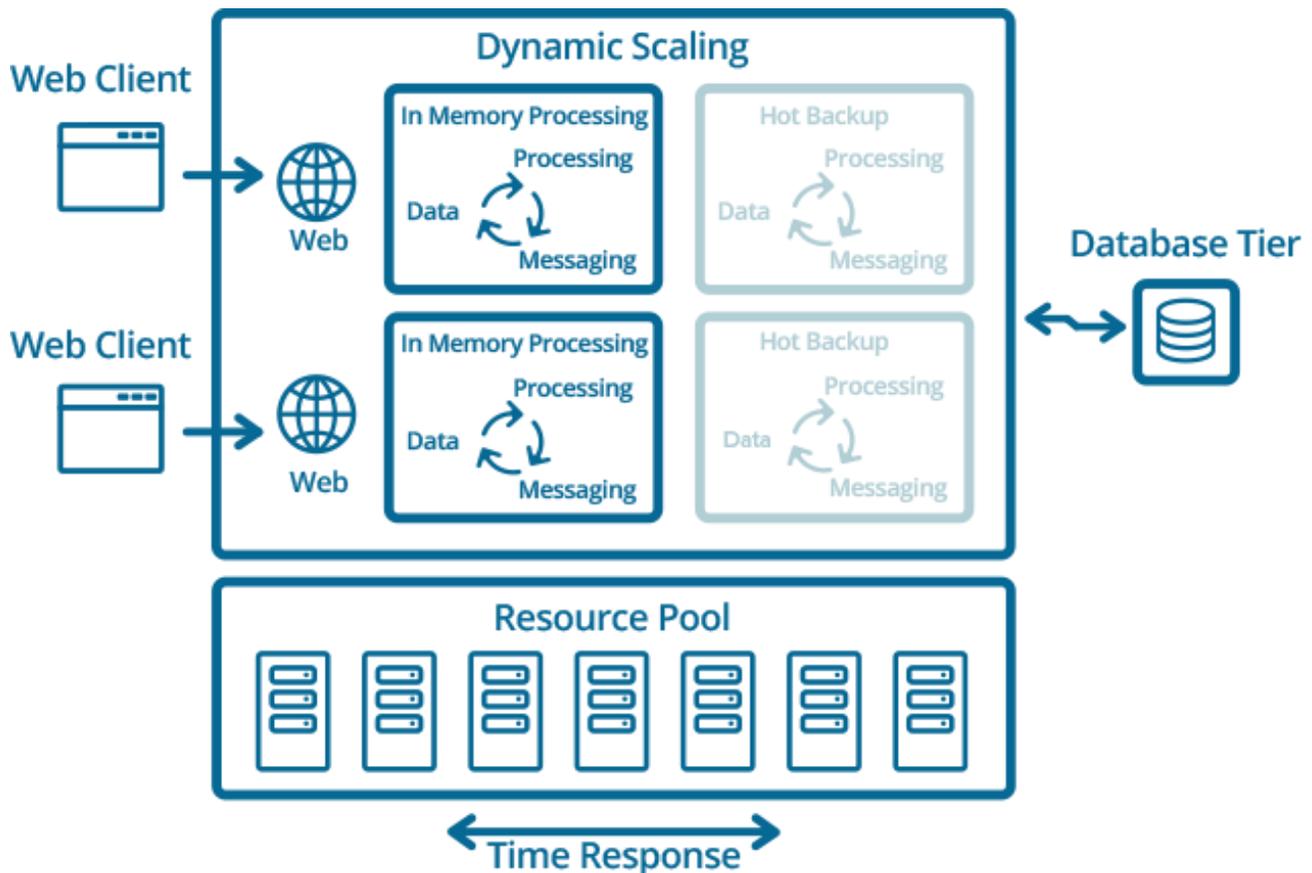
At the root of In-Memory Computing technology is an In-Memory Data Grid (IMDG).
This is a cluster of machines that work together to create a resilient shared data fabric for low-latency data access and extreme transaction processing. XAP IMDGs enable an application to run entirely on a single platform with all the tiers (processing, messaging and data) hosted within a single container. The platform provides fast data access by storing all data in-memory, ensuring high availability and scaling mission critical applications automatically and on demand.

The XAP Advantage

GIGASPACES

XAP In-Memory Computing is the only platform that enables the development, deployment and management of financial applications running their data, messaging and processing services within a single runtime environment.

XAP delivers true end-to-end scaling with:
- Dynamic provisioning to meet fluctuating loads.
- High performance for latency-sensitive environments.
- Continuous uptime/high availability through automatic failover and self-healing.
- Parallel code execution capability (map/reduce) for improved response time.
- Cost efficiencies through maximum resource utilization, standard development and deployment environments, and advanced management tools.
- Modern grid based architecture allows for infinite scaling leveraging the latest innovations in commodity hardware and flash storage technology.
- Controls and governance to protect the integrity of your data using an enterprise security model.

XAP is built to run real-time, event-driven analytics in memory, co-located with the data, messaging and processing making data immediately available to any process, no matter how complex, with no network hops, vastly speeding up results delivery for better risk management. With built-in multi-tenancy support, the platform enables multiple analytics processes to run on a shared infrastructure while maintaining strict isolation. The result – a more complete solution, providing faster results, at a lower cost, while preserving transactional integrity.

XAP In-Memory Data Grid and Open Source

XAP is a proven, end-to-end in-memory data grid (IMDG) platform ideally suited for analytics requirements. XAP offers a platform approach, which leverages best-of-breed open source tools, including STORM and SPARK for Stream and event processing, data integration tools from Mule and Talend, and Drools for rule execution and real-time decisions.

GigaSpaces is also the creator of Cloudify, an open source automation and orchestration framework integrated with XAP, which can be used with any cloud/virtual environment. While Cloudify is a general-purpose orchestration platform, it can serve many functions, including orchestrating XAP data fabric with other distributed applications such as WebSphere, JBoss, Tomcat and NoSQL databases, such as Mongo or Cassandra. This is accomplished on a bare-metal environment or on a private, public, or hybrid cloud environment.

GIGASPACES

Avoid Vendor Lock-in

XAP runs in the cloud or on premise, easily consolidating disparate sources of customer, product and transactional data from a variety of applications and data formats providing deeper insight in real-time and faster time to market. XAP runs on any OS that supports Java (Windows, UNIX or Linux), providing additional flexibility and portability that allows the decoupling of hardware from the software stack. As a result, firms are able to avoid vertically integrated stack lock-in. In addition, XAP leverages commodity servers and the latest Flash storage technology.

The XAP Price/Performance Edge

Analytics applications face an escalating need for performance, while data quantities grow exponentially and the need for real-time reporting and analytics.

An independent Total Cost of Ownership (TCO) study conducted by Financial Services Industry experts at Robert Frances Group (RFG) concluded that a **XAP Open Systems Solution is 6x more cost effective than a comparable proprietary Oracle Exadata solution when comparing transaction cost per minute** (Price per core per scan per minute model is derived, in part, from http://www.tpc.org/tpcc/)

XAP is Different

For financial analytics and transactions, XAP's main differentiator compared to other grid databases is XAP's maturity and deep support for reliable durable and consistent data processing. XAP can ingest a variety of data sources, streaming data, documents, and electronic correspondence (virtually any data source).

XAP overcomes relational database built-in performance limitations by acting as a single data source or repository, much like a system of record, by replicating and partitioning data while performing a "write-behind" to relational or NoSQL databases preserving entire system consistency.

XAP's Space-Based Architecture (SBA) means each processing unit instance that is acting as the deployment unit with the XAP runtime environment handles only the data preserved within the data grid instance it maintains. In an SBA application model, the data is partitioned in such a

GIGASPACES

way that the services that are triggered as a result will not have to read or write data from other partitions, thus achieving data affinity and in-memory read and write speeds that are far superior to traditional database latencies.

Other major XAP differentiators include:

- **Bidirectional integration** with different data sources, whether to consume data or to push data back out. Integration with real-time processing with Storm, Spark, Hadoop, relational and other databases
- **XAP MemoryXtend** allows XAP IMDG to manage up to 100TB using RAM and local SSD/central flash array devices to lower overall grid cost without sacrificing latency requirements.
- **XAP leverages Cloudify**, an open source automation and orchestration framework developed by GigaSpaces allowing XAP to be deployed seamlessly on any cloud/virtual environments.
- **Consistent, flexible transactional behavior and architecture** provides eventual consistency or fully ACID, synchronous or asynchronous transactions. Maintain data integrity with 100% transactional data handling through real-time primary with hot-backup replication of in-memory transactions.
- **XAP can be used as a Cache or the System of Record** interacting transparently with most big data and NoSQL solutions and standard SQL databases.
- **XAP simultaneously scales the cluster** which includes both the data and the business logic in the same grid.
- **No network traffic between servers in the grid** as XAP divides the data into partitions and the system keeps track of each allocated partition.

GIGASPACES
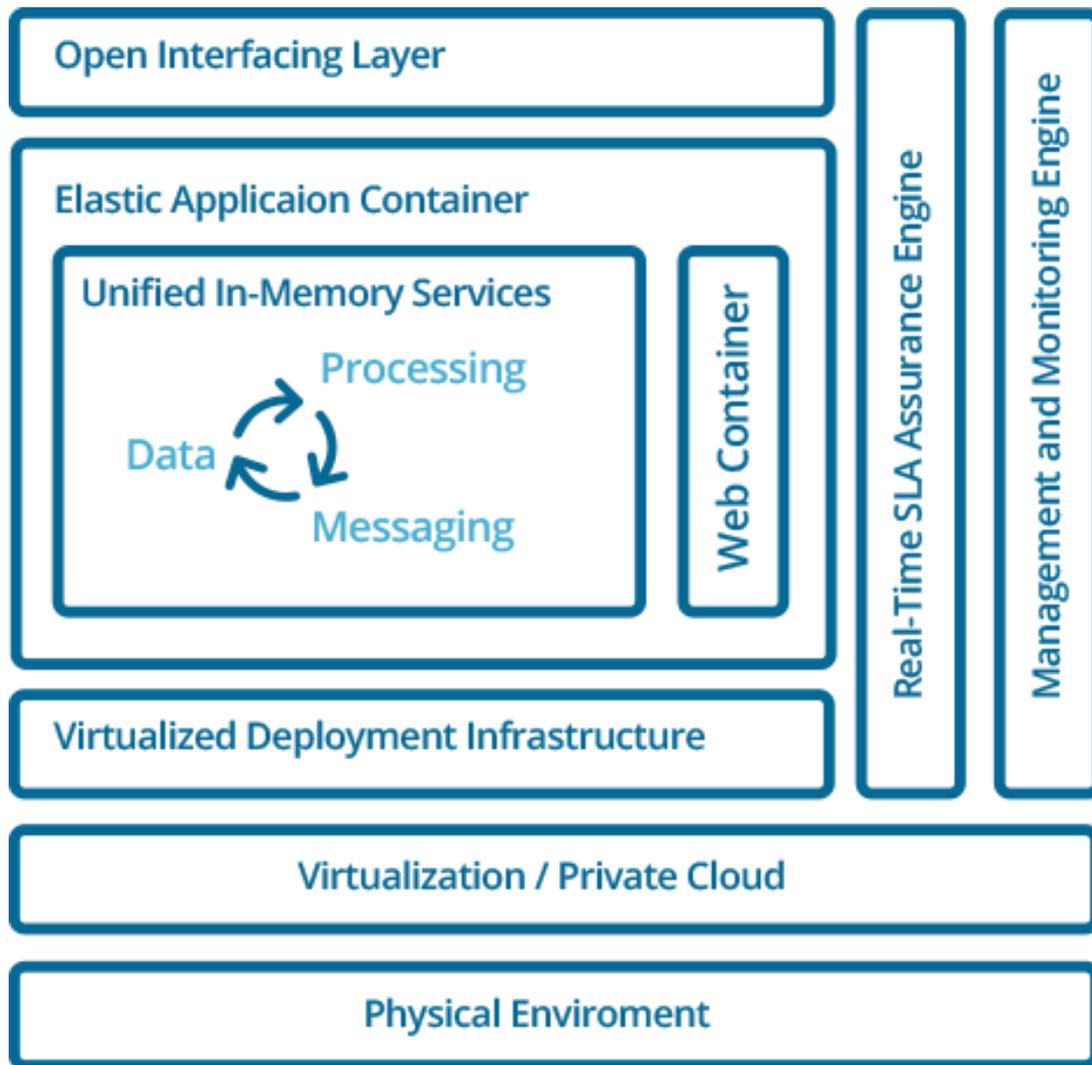
XAP Advanced Features

XAP has a number of advanced features that dramatically boost real-time analytic workloads including:

- **Data aggregation** – Similar to relational databases (Sum / Avg / Count / group by, etc.), but much faster execution in a distributed configuration within the data grid scanning millions of items in seconds. Ensures immediate consistency, rich query semantics & data indexing.
- **Stream processing** – Complex data event processing using XAP integration with Storm, Spark, Drools and other frameworks supporting massive data processing in a distributed elastic configuration.

GIGASPACES

- **Real-time Big Data** – Frontend Hadoop, Cassandra, MongoDB using built-in adaptors and Kafka to propagate real time processing to backend disk based big data systems for further offline analysis. In some cases, XAP can even eliminate the usage of these databases and is optimized for fast data

- **Replication over WAN** – Scaling across data centers with near real-time replication of massive data streams. Ideal for disaster recovery, cloud bursting and multi-site scenarios.

- **SSD as Data Grid storage medium** – Leveraging both RAM and SSD to manage large amounts of data at a fraction of the cost, reducing hardware footprint.

- **Customer 360** – Ability to consolidate data from various sources, both external and internal and structured or unstructured, into one unified model. XAP supports this via several fundamental features such as a Document / Schema-less model, transparent continuous data load from different databases and the Rest API.

- **Continuous High Availability** – Provided via backups and self-healing features and sub-seconds failover capabilities. XAP has no single point of failure.

- **Data Protection** - There is always at least one in-memory backup copy of your data, along with the additional database copy used in case of complete system shutdown to recover the entire IMDG state.

- **Scalability** – Share-nothing architecture, partitioning, automatic rebalancing with smart data routing and elastic provisioning enables XAP meet fluctuating and variable unexpected loads. Applications achieve linear scalability without significant re-architecting.

- **Front end mainframe** – Allows the XAP IMDG to pull data from mainframes, or any source for that matter, via public interfaces (SQL or web services for example). In the event of a cache misreading from the IMDG or once the IMDG is started, XAP accesses the mainframe to fetch relevant data seamlessly.

- **Interoperability** – By use of a schema-less data model through documents. Support for multiple languages / protocols (Java, .NET, REST) and Document Native support of popular programming languages (Java,.Net, C++), APIs (Object/SQL, Map/JCache, JMS, JDBC, JPA, etc.), and frameworks (Spring, JEE) enables smooth integration with existing environments & skill sets, essentially eliminating vendor lock-in.

- **Event Processing** – True in-memory computing, achieved by co-locating code and data. Supports multiple execution scenarios, such as distributed task executers and co-located event processors. Reduces moving parts, and supports advanced parallel processing patterns as well as full messaging semantics (p2p or pub-sub).

- **Configuration and Performance Monitoring** – Includes topology & availability monitoring and data grid queries and reports.

GIGASPACES

- **Easy Setup** – And installation via proven automation tools and an orchestration framework which includes a set of pre-packaged configuration (blueprints) designed to allow a user to automate the provisioning, configuration, setup and management of XAP clusters within your existing data center or in your private cloud.
- **Ease of Management** – Both underlying infrastructure and grid can be managed from a single point of control for fast and safe provisioning. Also includes desktop UI, web UI, admin API for management and monitoring of the cluster.

Real-Time Market Data Analytics Use Cases

Reliable Pricing and Risk Calculation

To perform reliable pricing and risk calculations (VaR, Basel) you must have reliable high quality market data. GigaSpaces XAP helps financial companies to meet business objectives and time to market, manage the fast growth of these types of instruments and data sources. A pricing system using GigaSpaces XAP provides flexibility and customization for growth locally, to provide intraday and end of day valuation, and market risk analysis with access to historical data to calculate the market risk stress testing.

Trading Strategies Development Made Easy

GigaSpaces XAP enables financial companies to develop sophisticated trading strategies quickly and easily. With today's economy where time to market is a key competitive advantage, it is an essential component of any successful trading environment. Flexibility to handle any new reference data item or pricing data item without any downtime is critical.

Unlimited Back-Testing with Low-Latency Data Access

Trading strategies fed into XAP are processed to execute back-testing while leveraging in-memory data processing technology – which supports fast data access to intra-day and historical data provided by tick feeds. Intelligent analysis identifies the optimized trading strategies leveraging historical data using GigaSpaces XAP as the storage medium both for the raw and processed market data.

Agile Trading Strategies Deployment

GIGASPACES

Optimized trading strategies created via back-testing cycles can be deployed immediately while the market is active. While these trading systems are running, users may update these in real-time to cope with market changes. Once transactions are executed, GigaSpaces XAP stores trading results and tracks the impact of the trading algorithms to compare with historical performance.

Personalized Customer Services

With access to real-time market data and both external and internal customer data, banks can offer competitive and personalized terms and conditions on interest rates, commissions and other services as a result of the immediate risk insight that can be gathered in real-time as soon as the customer applies for a new product or service. Milliseconds response time not only significantly improves users' experience and productivity, but with in-memory analytical processing, much deeper analysis becomes possible leading to new and richer insights that may change business processes altogether. Data on customers, orders, accounts, trades and the like is kept partitioned in replicated in-memory data grid instances. Each transaction updates this in-memory data while creating an entry in an in-memory log, which is then used to update external systems such as the enterprise database.

Bottom Line

XAP is the only in-memory data grid (IMDG) solution that allows you to **dramatically improve application performance without allocating expensive servers to handle peak loads and future growth.** XAP runs on low cost commodity hardware, is compatible with most open source solutions, can run on-premise or in the cloud and is highly available with no data loss concerns.

The need for fewer compute resources also means a "greener" data center footprint.
Moreover, XAP TCO is far superior to other solutions whether comparing upfront system investment, cost per transaction (scan) or cost to maintain the system over time. In addition, XAP from GigaSpaces is the only IMDG available today that supports virtually any internal and external data source necessary to provide your business with powerful insights to gain competitive advantage and drive next generation business models by leveraging real-time data.

GIGASPACES