

ThingSpan™: The Distributed Graph Platform for Intelligence Analytics

The Challenges and Opportunities

Objectivity's database platforms already play a key role in systems for telecom and process control equipment manufacturers, financial organizations, and government agencies to continually improve decision-making and operational efficiency.

While the value of big data is undeniable, many organizations struggle with adopting a solution that can help them rapidly process data to extract actionable intelligence. When they are not able to realize the value of their data quickly, billions of dollars or national security may be at stake.

The ever-increasing amount, speed and variety of information that has to be gathered, assimilated and acted upon makes it crucial to adopt a massively scalable distributed graph platform that can handle:

- Graph analytics and real-time relationship discovery
- Integration with the open source stack – Spark, Kafka, HDFS, YARN
- High-speed ingest with parallel querying
- Petabyte-scale to trillions of nodes and edges

“ Forrester Research estimates that graph databases will be the fastest-growing area in database management systems, with more than 25% of enterprises using graph by 2017¹. ”

The systems needed to deal with these sources and opportunities impose many significant technical challenges:

- They are computationally intensive.
- Path and pattern finding can provide new insights but requires massive parallelism.
- Stream processing must spot anomalies and trigger appropriate responses.

An enterprise-grade graph analytics platform can be used to augment the accumulation and analysis of all of the above sources of intelligence. Here are some examples:



Closed and Open Source Intelligence

Many organizations routinely collect and file masses of information but lack the means to interpret its real meaning, or to extract all of the facts relating to a known item or event. In many cases, the relationships between items only become clear as the users explore existing data to solve a problem. Users must also be able to define and record newly discovered types of data and relationships.

The dynamic schema and path finding capabilities of Objectivity ThingSpan™, coupled with the scalability, high availability and machine learning provided by Apache Spark, make it easy to rapidly deploy flexible intelligence analysis systems. Incoming data can trigger analytic algorithms while parallel processes perform pattern finding across the new and existing data.

Financial Streams

With so much money at stake, cybercrime and benefit fraud are a top priority for financial, retail and government organizations, but many security breaches are not identified until the damage has been done. A graph analytics solution based on ThingSpan can instantly spotlight any unusual patterns that may be indicative of fraud or a security breach by correlating data from security and network solutions with internal and semantic web applications. Unauthorized activity can then be shut down before theft or damage can occur.

IoT Sensor Streams

ThingSpan can be used to build a fusion framework that correlates vast amounts of discrete or streaming data from multiple devices

¹ Big Data Analytics News:

<http://bigdataanalyticsnews.com/4-predictions-for-nosql-technologies-in-2016/>

spread across the Internet of Things (IoT). ThingSpan supports parallel data loading and accelerated ingest so that data from many input streams can be continuously and quickly fed to applications for real-time analysis.

The ThingSpan Metadata Store is used to organize the data. A combination of Apache Spark Streaming filters and ThingSpan path finding is used to minimize the result sets that are presented to other systems and analysts. The fusion framework can be deployed at the edge and in immensely powerful ground-based Spark clusters.

The Technical Challenge

Many organizations rely on big data analytics solutions that involve a data ingestion layer that processes all data points about financial transactions and customer data—including market, news, and social media—before breaking them down into stored memory where they can be queried. However, while large volumes of data may be processed in this manner, institutions often rely on micro-batches and may need to wait days until the relevant data points can be surfaced. In many cases - particularly when dealing with profit opportunities and fraud - such delays are not an option.

In order to generate the high-performance, high-speed processing power and the sophisticated contextual analysis needed, an enterprise-class graph analytics solution is essential. Users need a highly scalable, real-time graph analytics platform to analyze massive volumes of data for pattern finding, predictive analytics and improving efficiency.

The challenge, however, is finding a distributed graph platform that can scale to petabytes of data and perform queries in parallel to data ingestion. While there are many platforms available, few of them offer real-time data analysis at scale, enabling organizations to capture streaming data and analyze it in relation to historical and contextual data to identify opportunities and risks across a broad array of use cases.

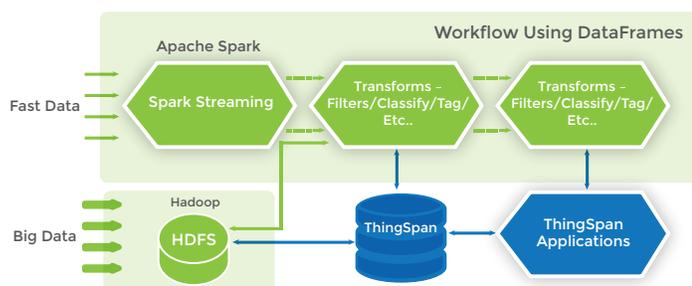
The ThingSpan™ Solution

ThingSpan, Objectivity's massively scalable graph analytics platform, is compatible with the distributed open-source data management framework, Hadoop, and utilizes Spark for high-speed streaming ingest and enriched data processing. It has the power to transform and analyze real-time streams and data lakes in context, offering a single logical view of all data—wherever located—to accelerate parallel processing.

ThingSpan leverages open-source tools by supporting the Hadoop and Spark ecosystem atop a high-performance, distributed graph database purpose-built for relationship and pattern discovery. It runs natively on top of POSIX or HDFS as a YARN application while using Spark for workflow and data transformation. It is also designed to support streaming systems based on Kafka, Flume and other distributed messaging tools for streaming data. Integration with Spark via DataFrames allows ThingSpan to ingest streaming data while maintaining and persisting relationships as first-class logical models.

This model allows for enriched and transformed data to simplify the support of complex, multi-dimensional queries associated with intelligence analytics. ThingSpan enables the capture of powerful insights based on data relationships to make better-informed decisions, increase value, and remain vigilant in order to counter threats.

ARCHITECTURE DIAGRAM



Objectivity, Inc. delivers massively scalable and highly performant distributed database platforms that are proven to power mission-critical applications for the most demanding and complex datasets in the enterprise. Objectivity helps organizations to rapidly build new Spark Streaming-enabled solutions for finding connections and patterns through graph analytics within petabytes of data, stored in HDFS, to achieve real-time relationship discovery. With a rich history serving Global 1000 customer and partners, Objectivity holds deep domain expertise in fusing vital information from massive data volumes to capture new revenue opportunities, drive competitive advantages and deliver better business value. Objectivity is privately held with headquarters in San Jose, California. Visit <http://www.objectivity.com> to learn more.



ThingSpan is a massively scalable distributed graph platform designed specifically for the complex issue of extracting actionable insights from Fast Data and Big Data sources to enable real-time relationship discovery. It is architected to integrate with major open source Big Data technologies, such as HDFS and Apache Spark. ThingSpan leverages Objectivity's core object and graph data-modeling technologies and the company's rich experience in building fusion solutions.