+ THINGSPAN™

Hadoop® and Objectivity's ThingSpan™

As the growth in deployment of massive sensor networks by enterprise and government organizations continues to redefine the sensor-to-insight data flow, it is easy to believe that existing Big Data software and tools can be used to build applications involving fast, streaming data. However, realizing value from larger amounts of Fast Data requires a different solution than those for deriving value from data-at-rest. This is because data stored using Apache Hadoop-based storage is not fast enough to act on real-time data.

Apache Hadoop is a scalable, open-source data storage platform that has emerged as the backbone of a growing number of enterprise organizations' Big Data systems. Hadoop consists of two major components: The Hadoop Distributed File System (HDFS) for storing data and a processing framework to enable parallel computation on the same computing nodes to minimize data movements.

In the initial version of Hadoop, MapReduce was the primary computing framework, but it has since been extended with Apache Spark and YARN to support a variety of processing models and leverage in-memory capabilities, respectively. Two of the key contributions of the Hadoop ecosystem were the support of a schema-on-read approach to facilitate the ingestion of a variety of data types and a software framework for leveraging large numbers of commodity servers for scale-out computing.

For example, sensor networks and internet of proprietary data streams generate massive amounts of fast, streaming data. Some have proposed piecing together Hadoop's capabilities to ingest and persist massive quantities of data together with high throughput messaging tools, such as Apache Kafka, to address the challenges of building Internet of Things (IoT) applications. Kafka is an open-source, fast, scalable, durable, and fault-tolerant publish-subscribe messaging system. Kafka is often used in situations requiring high throughput, reliability and replication.

THINGSPAN

• **ThingSpan, the enterprise graph platform for analytics at speed and scale,** is a powerful, distributed graph platform that enables faster processing and higher performance.

• **Native support of the Big Data ecosystem**
Architected to support industry-standard, open-source technologies, ThingSpan leverages these key Apache platforms: **Hadoop, Spark, Kafka,** and **Yarn.**

• **Rapid navigation and pathfinding queries**
ThingSpan uses Spark to collect and analyze real-time streaming data, instantly triggering actions when required.

• **Organize highly interconnected data by relationships**

ThingSpan enables organizations to map out the connections between data points in real-time, making it possible to immediately apply advanced graph analytics.

twitter.com/objectivitydb
facebook.com/ObjectivityInc
linkedin.com/company/objectivity

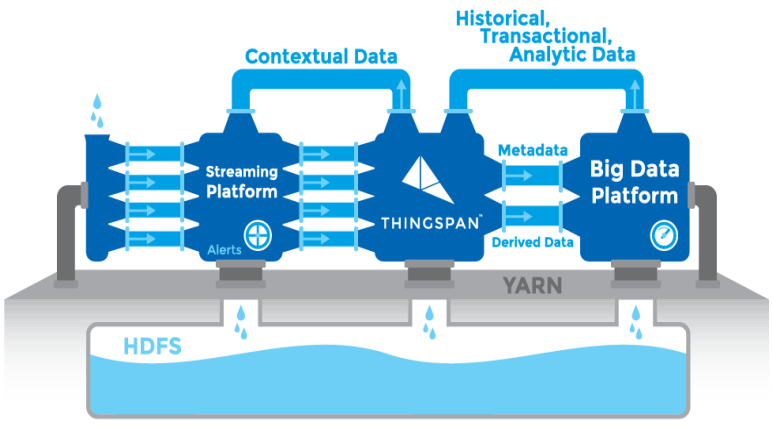
Challenge of Improving Sensor-to-Insight Data Flow

The idea of composing solutions from open-source components, such as Hadoop and Kafka, to support IoT applications is appealing on the surface, but the reality involves greater complexity. Tools like Kafka are designed to move Fast Data into HDFS, and they do not perform any processing during the data movement.

This approach results in "dumping" great volumes of raw sensor data in HDFS for analysis after the fact. This limits the system's ability to react to real-time events, as well as requires far larger computational resources to analyze the massive volumes of data after they are loaded into HDFS.

A more powerful ingest system involving real-time data Transformation, and integration leads to better data for HDFS because data can be filtered, aligned, aggregated, and enriched during the ingest process. Data can also be queried directly during ingest rather than "after the fact" to detect anomalies and other patterns of interest.

A more streamlined Fast Data pipeline also enables the combination of transactions with analytics to facilitate faster development of complex IoT applications. For example, the nature of sensor data can benefit from using machine-learning methods to automatically transform Fast Data. An advanced data ingest system can merge streaming data from sensors with static data from analytic tools based on machine-learning to automatically correct and transform sensor data for additional analysis.



INDUSTRY USE

ThingSpan builds on Hadoop to form a powerful data-storage and processing workflow for scenarios involving real-world objects, such as physical locations. In one use case for an oil and gas company, an Objectivity customer utilizes sensor-driven devices to capture petabytes of real-time data related to underground seismic activity from its ship and land-based seismic sensors together with well log and other data to dynamically build a common data model that supports geoscience applications.

Rather than managing data streams from different data sources into separate datasets, Objectivity's technology is used to create and persist a common data model that provides a single, common view of all relevant seismic data, including all associated metadata. When analyzing geospatial data, there is no lag time: It can be immediately visualized and acted upon data insights, leading to higher performance and better business intelligence.

twitter.com/objectivitydb
facebook.com/ObjectivityInc
linkedin.com/company/objectivity

Big Data in Real-Time

By leveraging Hadoop with Spark and supporting streaming messaging tools, such as Kafka, atop a platform purpose-built for relationship analytics, Objectivity's ThingSpan helps companies achieve business insights from Big Data and real-time streaming data with a high degree of efficiency at scale.

ThingSpan ensures superior performance by organizing data about people, locations, events, and devices into a logical model involving objects and the relationships between them. This allows for enriched and transformed data, as well as associated metadata to simplify the support of the complex relationship navigation and pathfinding queries associated with analytics.

ThingSpan's enterprise grade platform for graph analytics makes it easier for applications to manage and gain insight from data volumes well beyond the petabyte level. Now organizations can transform massive volumes of streaming, Fast Data from generic to relevant in real-time, thereby maximizing business value.

About Objectivity, Inc.

Objectivity, Inc. delivers massively scalable and highly performant graph analytics platforms that are proven to power mission-critical applications for the most demanding and complex datasets.

With a rich history serving Global 1000 customers and partners, Objectivity holds deep domain expertise in fusing vital information from massive volumes of data and sources to discover unknown connections at speed and scale. Objectivity's technology enables enterprises to make better decisions with precision, scale and efficiency. Objectivity is privately held with headquarters in San Jose, California. Please visit http://www.objectivity.com to learn more.

THINGSPAN

• ThingSpan for HDFS

ThingSpan adapts Hadoop's HDFS environment for simpler, high-speed data processing and analysis.

• ThingSpan for Apache Spark

Adapters for Spark enable users to manage Spark DataFrames and convert ThingSpan-collected data to Spark components, such as SQL and MLlib.

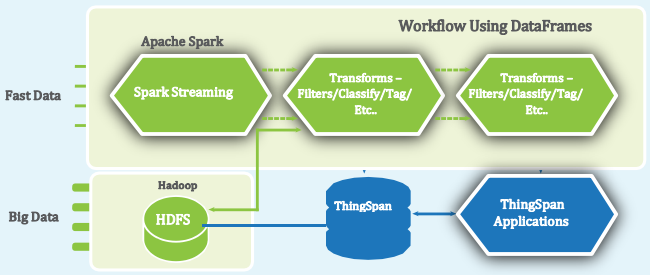
• ThingSpan Metadata Store

The metadata store enables users to pre-define metadata schemas to define relationships between data.

• ThingSpan Rest API

ThingSpan's Rest API provides a simple interface for defining and managing queries, and transforming data.

ARCHITECTURE



twitter.com/objectivitydb
facebook.com/ObjectivityInc
linkedin.com/company/objectivity