



## TIBCO StreamBase and the TIBCO Accelerator for Apache Spark

Real-time insights from vast quantities of closely sampled data

### STREAMBASE REAL-TIME ANALYTICS USE CASES INCLUDE:

- The Internet of Things
- Fraud detection
- Network operations
- Trading
- Web analytics

TIBCO StreamBase® event processing platform can be used to develop, host, and execute integration and predictive analytics against big data stores in real time. At the same time, open source big data software in the form of the Apache® Hadoop® distribution framework offers a wide variety of technology and business capabilities. Together, these solutions provide a powerful platform for gaining critical insights from increasingly vast quantities of data generated at increasingly frequent intervals for making informed decisions.

To streamline implementation of streaming analytics solutions, TIBCO has released a number of “accelerators,” collections of templates, rules, and integration patterns to help jumpstart the development effort. The TIBCO Accelerator for Apache Spark includes connectivity points and usage patterns that provide sophisticated, real-time analytics and data processing in conjunction with Hadoop distribution. Specifically, implementation provides an event processing engine that applies data transformation, application logic, and analytics in a sequential event-by-event processing model. A sample data model, rules, and functions are provided with the Accelerator.

## TIBCO SPARK ACCELERATOR ARCHITECTURE OVERVIEW

Hadoop itself is a large platform with a large number of components, each of which has a specialized capability. The following table highlights the main elements of Hadoop distribution and identifies how StreamBase interacts, leverages, and supports that architecture.

The Accelerator for Apache Spark offers more than 40 connection points to the Hadoop stack. These StreamBase adapters provide no-code, out-of-the-box connectivity that can be further enhanced and supported by either customer customization or TIBCO services using the StreamBase client API SDK.

COMPONENT	USAGE	STREAMBASE ADDED VALUE
Apache® Kafka®	Messaging	Kafka plays the critical roles of distribution and message delivery while StreamBase consumes data from Kafka as a processing node for real-time data transformation (STL), along with event-based processing logic and analytics. StreamBase can also publish events onto Kafka for delivery to another source.
Apache® Flume®	Streaming Data	Flume offers streaming distribution, with different semantics than Kafka. As with Kafka, StreamBase consumes data from Flume as a processing node for real-time data transformation, along with event-based processing logic and analytics. StreamBase can also publish events back onto Flume for delivery to another source.
Hadoop distributed file system	Data Storage	HDFS is a large-volume data store that serves as a repository for StreamBase, which can access data as individual events or as an entire packaged result.
Apache® HBase® & Apache® Cassandra®	Data Storage	HBase and Cassandra are two columnar store databases that serve as more specialized storage in Hadoop. StreamBase is able to access data either as individual data records or sets of records.
Spark	In-Memory Data Storage	Spark offers a high-performance, distributed in-memory data structure that can be accessed via MLib, the Machine Learning libraries of Spark.
H2O, JPMML, Drools, R, TIBCO® Enterprise Runtime for R (TERR), Matlab, and others	Algorithm Models	While not specifically part of Hadoop, these algorithmic models are all embeddable and executable within the StreamBase platform, which provides abstraction from deployment coupled with data marshaling before and after the algorithmic execution.
MQTT, OSI PI, WITSML, and others	Third-party Connectivity	While not specifically part of Hadoop, these no-code data sources configure and connect sources. They are often combined with a big data environment, with StreamBase providing the no-code, configuration-based connectivity.

The Spark Accelerator provides a number of StreamBase applications that address various access patterns and use cases for real-time processing with StreamBase on Hadoop. These include Kafka transaction binding, Flume binding, Apache® ZooKeeper® notifications, context binding, enrichment, transaction features, and model execution.

TIBCO documentation for the Spark Accelerator includes:

- *Functional Specification*: An overview of the components associated with the accelerator
- *Interface Specification*: A detailed examination of the data exchange interfaces of the accelerator
- *Quick Start Guide*: A guide to install and execute the accelerator demos

## WHY STREAMBASE FOR STREAM PROCESSING AND VISUAL ANALYTICS

The TIBCO StreamBase event processing platform is a high-performance system for rapidly building applications that analyze and act on real-time streaming data. With StreamBase, organizations can rapidly build real-time systems in record time and deploy them at a fraction of the cost and risk of alternatives.

StreamBase provides a proven platform for developing real-time analytics applications that have been deployed in the most demanding of production environments. Key capabilities of this platform include:

### SOPHISTICATED TOOLING AND DEVELOPMENT ECOSYSTEM

The TIBCO StreamBase Studio™ development environment leverages all of the modern development workflow and practices.

### EVENT-DRIVEN, DOMAIN-SPECIFIC VISUAL LANGUAGE

StreamBase includes a visual, domain-specific language that evaluates events on an event-by-event basis. The language has native support for temporal processing, analysis and event correlation, and a number of stream-processing language features.

### SCALABLE, DEPENDABLE, ENTERPRISE RUNTIME

The StreamBase runtime is in production in mission-critical applications across a large range of demanding use cases in industries including Airlines, Capital Markets Trading, Cybersecurity, Oil & Gas Production, and others. It has demonstrated extreme performance and platform resiliency.

### PERVASIVE CONNECTIVITY

StreamBase includes a large number of adapters to open source systems (the Hadoop ecosystem, MQTT, and others), commercial systems (OSI PI, WITSML, and others) and general infrastructure (messaging, database, files, sockets, and others). It also includes an SDK that generates Java, C/C++, Python, .NET, and Web Service APIs.

## TIBCO STREAMBASE PLATFORM

### TIBCO STREAMBASE LIVEVIEW™

Web and Desktop clients provide interaction with live, continuously updating data and queries on tables contained in the TIBCO Live Datamart Server. The clients provide a view into an in-memory data warehouse and provide push-based query results and alerts to end users. Business users can analyze, anticipate, and alert on important business events in real-time, and act on opportunities or threats as they occur.

### TIBCO STREAMBASE STUDIO™

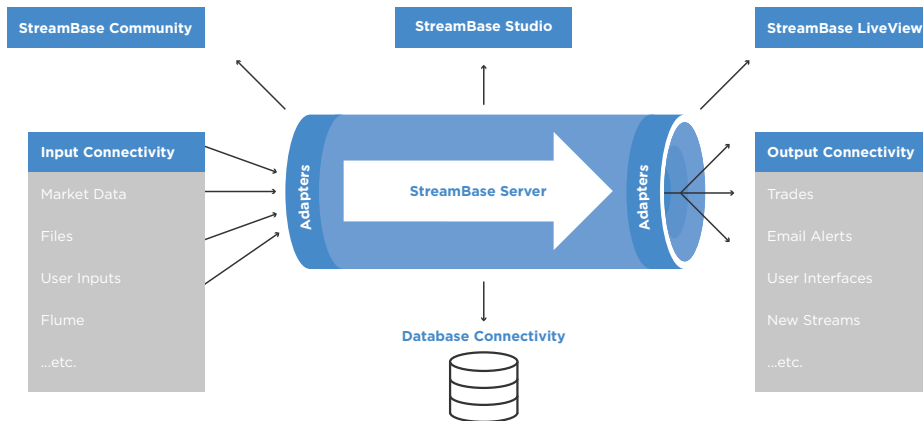
TIBCO StreamBase Studio™ is a powerful Eclipse-based integrated development environment for designing real-time applications using the standards-based StreamSQL text language or the first and only graphical event-flow language, StreamSQL EventFlow. Development is typically cut to a fraction of the time compared to custom-coding or other alternatives. StreamBase Studio provides a rich user interface and support for the full application lifecycle, spanning feed integration, application modeling, development, stream recording/playback, testing, debugging, and integration with downstream systems and dashboards.

### STREAMBASE SERVER

StreamBase Server run-time engine is an ultra low-latency application server optimized for processing real-time streaming event data at throughput rates of tens of thousands of messages per second up to 500,000 messages/second or more on multi-core CPU hardware.

### PRE-BUILT ADAPTERS

Pre-built Adapters connect to more than 150 real-time and historical data sources to easily and quickly increase data volumes for comprehensive input without custom development and integration. This robust connectivity improves your ability to detect relevant and urgent business events. It allows you to effectively hone in on the most critical business moments you need to act in real time.



## TIBCO STREAMBASE TOTAL COST OF OWNERSHIP

The TCO of software licensed from a commercial vendor is often measured against a single up-front cost that does not reflect the total investment over the lifespan of a project. There are a number of reasons why TIBCO StreamBase has a competitive or superior TCO to open source software for similar projects. Open source software has a critical role in today's modern technology stack. The intent of this paper is only to represent the TCO benefits of StreamBase over similar project types that involve sequential data processing, real-time analytics, and similar use cases.

The cost of a project can be broken down into several components:

- *Up-front Materials* - Any initial costs associated with software licensing or subscription, along with project hardware and other resources
- *Up-front Personnel* - Initial team required for the development cycle, expressed over time
- *Long-term Materials* - Ongoing costs associated with software subscription or maintenance
- *Long-term Personnel* - Ongoing team required to enhance and support the project
- *Time-to-Value* - The time from the initiation of the project to the delivery of functional business value

Based on customer implementations and feedback, expected relative project investment is as follows:

TECHNOLOGY	UP-FRONT MATERIALS	UP-FRONT PERSONNEL	LONG-TERM MATERIALS	LONG-TERM PERSONNEL	TIME-TO-VALUE
StreamBase	High	Low	Low	Low	Short (3+ months)
Open Source	Low	High	Low	Medium-high	Long (12+ months)

"There's been an explosion in real-time data led by in-running, or betting during the game. We're meeting the demand by collecting everything that's going on in an industry that's exploding. Our goals are to be an innovator and an enabler to companies by supplying the full scale of data."

—Einar Knobel, CEO, TXODDS

## TIBCO STREAMBASE INTANGIBLES

There are other benefits of StreamBase adoption beyond TCO, including:

### EVENT-PROCESSING SPECIALIZATION

StreamBase has been designed from the ground up to provide a domain-specific language and approach to sequential event-processing. This is a specialized capability not well represented in the Hadoop stack and nascent in its projects.

### PRODUCTIVITY

StreamBase leverages a visual decision-tree style representation for its high-performance domain-specific language, EventFlow. This language encapsulates real-time, event-driven concepts, and is extensible with custom code for a number of different languages.

### CONNECTIVITY AND ABSTRACTION

Out-of-the-box StreamBase provides support for over 200 data sources and destinations, ranging from Hadoop infrastructure (Kafka, Flume, HDFS, HBase/Cassandra, etc.), to generic technical infrastructure (JMS, MQTT, RDMS, Files, Sockets, etc.), to commercial and vendor-based systems (OSI PI, WITSML, etc.). Further, an API provides wizard-driven definition of new adapters for Java, C/C++, .NET, Python, and Web.

### ROBUST PERFORMANCE

StreamBase is a proven, high-performance platform that has been in production in demanding industries and mission-critical environments for more than 12 years. StreamBase provides TIBCO customers with a field-proven, reliable platform for delivering mission-critical applications.

### SIMPLICITY OF MODEL

StreamBase provides a simplicity that removes many of the decisions required for effective big data implementations via either abstraction or flexibility to change and evolve the implementation without affecting applications and logic.

### SINGLE, RELIABLE SOURCE OF SUPPORT

StreamBase provides full support for all of its products with an SLA response time, direct interaction with TIBCO Engineering, and a rich ecosystem of partners and services organizations ready to assist our customers.

### ACCELERATION TO BUSINESS BENEFIT

With all of these advantages, and with direct business involvement on specification and design, TIBCO Software customers see their projects yield rapid business benefits, as well as development and deployment completed within extremely aggressive timeframes.



**Global Headquarters**  
**3307 Hillview Avenue**  
**Palo Alto, CA 94304**  
**+1 650-846-1000 TEL**  
**+1 800-420-8450**  
**+1 650-846-1005 FAX**  
**www.tibco.com**

TIBCO fuels digital business by enabling better decisions and faster, smarter actions through the TIBCO Connected Intelligence Cloud. From APIs and systems to devices and people, we interconnect everything, capture data in real time wherever it is, and augment the intelligence of your business through analytical insights. Thousands of customers around the globe rely on us to build compelling experiences, energize operations, and propel innovation. Learn how TIBCO makes digital smarter at [www.tibco.com](http://www.tibco.com).

©2017, TIBCO Software Inc. All rights reserved. TIBCO, the TIBCO logo, LiveView, StreamBase, and StreamBase Studio are trademarks or registered trademarks of TIBCO Software Inc. or its subsidiaries in the United States and/or other countries. Apache Cassandra, Flume, Hadoop, HBase, Kafka, Spark, and ZooKeeper are trademarks of The Apache Software Foundation in the United States and/or other countries. All other product and company names and marks in this document are the property of their respective owners and mentioned for identification purposes only.

06/23/17