StreamAnalytix is a multi-engine, enterprise-grade, visual platform for unified streaming and batch data processing, and machine learning. Use compute engines like Apache Spark (and more) as the underlying technology to ingest, blend, and process high-velocity data streams as they arrive, run machine learning models, train and refresh models in real-time or in batch mode, visualize results on real-time dashboards, and raise corresponding real-time alerts and action triggers.

Build and run Apache Spark applications in the cloud 10x faster with an intuitive drag-and-drop interface, an exhaustive set of pre-built operators, full application lifecycle support, and one-click options for on-premise and cloud deployments.

**Why StreamAnalytix?**

- Build applications **5x to 10x** faster
- Boost performance by **4x** using the same infrastructure
- Have your existing teamwork at **3x** efficiency
- Process **1Mn+** events per second—both on-premise and in the cloud

**DATA 360**

- Data Integration
- Data Ingestion
- Data Preparation
- Data Quality
- Analytics & machine learning

**PLATFORM POWERED BY:**

- [Apache Spark](https://spark.apache.org)
- [Amazon EMR](https://aws.amazon.com/emr)
- [Databricks](https://www.databricks.com)
StreamAnalytix features

Unified batch and streaming data processing
Ingest and blend data at scale from any data source - batch or streaming.

With built-in support for Spark structured streaming, StreamAnalytix enables continuous applications by exposing a single API to write streaming as well as batch queries.

Use models trained and refreshed in batch workflows to make predictions on real-time data pipelines.

Click + code
Use an exhaustive set of pre-integrated drag-and-drop operators in an intuitive visual interface. Or introduce custom logic in the language of your choice (Java, Scala, or Python).

Figure 1: Build applications visually

End-to-end data processing
An end-to-end big data processing platform, StreamAnalytix supports data ingestion, data preparation and processing, analytics, machine learning, action triggers, data visualization, and data storage.

Data ingestion: Connect with any data source or storage system for both streaming and batch use cases on-demand. Use pre-built connectors or create your own using custom API. Ingest and output popular data formats like JSON, CSV, AVRO, and Parquet, or easily build your parsers for custom data formats.
**Data preparation and processing:** Perform data cleansing, data blending, and data enrichment at scale - on data as it arrives. Experience significantly faster processing with Spark-based structured streaming as the underlying technology.

**Analytics:** Apply built-in analytical operators for complex event processing, time window aggregation, geo-spatial analytics, correlation, and more.

**Data science and machine learning:** Leverage advance analytics and machine learning support for multiple technologies like Python, TensorFlow, R, Spark ML, and H2O. Use pre-built operators or easily prototype your custom algorithms.

**Complete machine learning life-cycle support:**
Leverage support for full ML life-cycle to perform data preparation, feature engineering, training, testing, hyper parameter tuning, model performance, versioning, deployment and hosting, and drift.

**Environment management:**
Use the powerful environment management feature to create virtual environments, work spaces and leverage support for multiple notebooks to get down.

**Actions and alerts:** Set thresholds for key metrics and corresponding real-time alerts and action triggers.

**Data visualization:** Use inbuilt or custom real-time dashboards to display the status of metrics and key performance indicators for a pipeline. Blend real-time and historical data with offline analytics and integrate everything you must keep track of, however disparate, onto a single screen.

---

**BUILT-IN OPERATORS AND CONNECTORS**

**Connect with any data source and sink−batch or streaming**

- **Message queues**
  - Kafka, RabbitMQ, MapR Streams, MQTT, and more
- **Distributed file systems**
  - HDFS, Hive, and more
- **Cloud sources**
  - Amazon S3, Redshift, Amazon Kinesis, AWS-IOT, Azure Event Hub, Google Pub/Sub, Google BigQuery, and more
- **No SQL**
  - HBase, MongoDB, Cassandra, Couchbase, and more
- **RDBMS**
  - Oracle, Hive, SQL Server, and more
- **Indexing stores**
  - Elasticsearch, Solr
- **Custom channel**

**Data processors**

- **Data transformation**
  - Masking, MapToPair, FlatMap, StreamCorrection, Deduplication, TransformByKey, and more
- **Data cleansing**
  - Filter, Imputation, Take
- **Data blending**
  - Join, Union, Intersection, Group, Intersection
- **Data enrichment**
  - Enricher, LookUps, Web-services, Expressions, SQL
Data-flow lifecycle management
Build end-to-end applications with support for the entire delivery lifecycle - design, build, test, debug, deploy, monitor, and manage.

Rapid application development: Rapidly build and operationalize applications using a powerful visual pipeline designer, and a drag-and-drop interface.

Interactive pipeline designing: Create schema automatically within pre-built operators. Data can be accessed from a data storage system, or configured from a source such as Kafka, JDBC, and more, and is automatically examined and assigned for each field and column.

Debug: Trace and debug messages at each step, from entry to exit, during both the development phase of your pipelines as well as during production.

One-click deployment – on-premise or in the cloud: Use one-click deployment options to deploy applications on-premise or on a public cloud.

Application diagnostic tools: Allows auto ‘Data Inspect’ before and after the use of every individual operator, for an end-to-end view of data transformation at every step. And use ‘Data Lineage’ for your production-deployed pipelines.

Statistical and temporal analytics
- Aggregation, Average, Count, CEP, Window, Standard Deviation, and more

Predictive analytics, machine learning, and deep learning
- SparkMLlib
- SparkML
- H2O
- TensorFlow
- PMML
- Custom processing (using Python or R)

Extensibility
Retain full control and flexibility to add new functionality and interfaces as the technology ecosystem evolves.
- Extensions API, custom data operators, custom processors (Spark and Storm operators), custom machine learning (Python and R)
Versioning: Create and save new versions of your data pipelines. Roll back changes conveniently by reverting to an older version.

Performance monitoring: Run streaming and batch pipelines regularly and consistently once they are in production. Get all performance metrics in real-time through interactive graphs.

Workflow orchestration: Enable integration of multiple pipelines to work in sync, and support parallel stitching for logic evaluation before it is put in production.

Powerful StreamAnalytix Applications:

1. Advanced Data Ingestion:
   Use the inbuilt ingestion application to load data from any source to your destination, in a few clicks. Just select your steps, make configurations and watch your data-flow being auto-created in real-time

2. Automated ETL Migration
   The auto-migration utility makes it easy to migrate all your legacy ETL workloads by preserving the structure, logic and execution rules in easy step wise process

3. Change Data Capture
   Use the in-built CDC application to detect and capture any changes and ensure high data synchronicity across the enterprise.
Open source flexibility
Work with the power and flexibility of best-of-breed open source technologies integrated into a high-performance, scalable, and reliable enterprise-grade platform.

Multi and hybrid cloud support
Leverage the platform’s ability to work across various platforms. Build your applications and pipelines and host them on any cloud provider of your choice or host over a hybrid environment.

Built-in extensibility
With the extensions API exposed by the platform, you can write your functionality in the language of your choice (Java, Scala, SQL, and Python), and make it available for all users across the platform.

Self-service
Use pre-built templates for frequently used application patterns and customize them to your needs. Easily access support and tutorials at every step.

Strong ecosystem integration
Compatible and integrated with leading big data technologies and platforms such as all key Hadoop distributions (Map R, Hortonworks, Cloudera), cloud platforms (AWS, Microsoft Azure, Google Cloud), and key third-party partnerships.

StreamAnalytix Lite—A Visual IDE for Apache Spark
StreamAnalytix Lite is a free, compact version of the StreamAnalytix platform. A lightweight visual integrated development environment (IDE), StreamAnalytix Lite offers you a full range of data processing and analytics functionality to build, test, and run Apache Spark applications on your desktop or any single node.

- Build and run enterprise-grade Apache Spark applications on your desktop
- Use a wide range of built-in operators, and an intuitive drag-and-drop interface to build Apache Spark pipelines within minutes, without writing a single line of code

STRONG INDUSTRY RECOGNITION

FORRESTER
The Forrester Wave™: Streaming Analytics, Q3’17
One of the 13 Most Significant Streaming Analytics Providers

Gartner.
Market Guide for Event Stream Processing, 2017- Gartner Research
One of the Key Event Stream Processing Platforms

Aragon Research
2017 HOT VENDOR
Hot Vendors in Streaming Analytics, 2017- Aragon Research
One of the 4 Hot Vendors in the Streaming Analytics Space in 2017

datanami
Datanami Editors’ Choice Award
Best Big Data Product or Technology: Real-time Analytics
StreamAnalytix architecture

StreamAnalytix integrates various key technologies, including support for multiple compute engines, a powerful array of pre-built connectors and operators to multiple systems, and functional extensibility for future-readiness.

- Use built-in advanced analytics and machine learning capabilities
- Use powerful multi-tenancy features
- Quick-start with a light-weight tool, downloadable onto your Windows, Mac, or Linux desktop, or a server node

Click here to download StreamAnalytix Lite for Free.

© 2020 Impetus Technologies, Inc.
All rights reserved. Product and company names mentioned herein may be trademarks of their respective companies. Sept 2020

StreamAnalytix is an enterprise grade, visual, big data analytics platform for unified streaming and batch data processing based on best-of-breed open source technologies. It supports the end-to-end functionality of data ingestion, enrichment, machine learning, action triggers, and visualization. StreamAnalytix offers an intuitive drag-and-drop visual interface to build and operationalize big data applications five to ten times faster, across industries, data formats, and use cases.

Visit www.streamanalytix.com or write to us at inquiry@streamanalytix.com