

# Getting Real Value from Real-time Streaming Analytics

*This white paper explores how real-time streaming analytics adds value to business. It also details how best to take advantage of real-time streaming analytics.*

## Introduction

Analysts believe that it has taken 15 years or so for companies to harness about 50% of the productivity potential of the Internet, and the next 50% of productivity gains likely requires connecting things. The Internet of Things is causing an interconnected world where “smart” devices allow things and people to be connected from anyplace, anytime, with anything and anyone. However, the IoT data loses its value if it is not detected and acted upon immediately. That is where streaming analytics platforms can help.

Many industry experts believe that, just as the database management opportunity gave birth to a wide range of database technologies and Big Data needed Hadoop, the real-time enterprise and IoT applications need development tools and processing capability to support real-time streaming analytics.

## What is Real-time Streaming Analytics?

Real-time streaming analytics is a technology that allows for the collection, integration, analysis, and visualization of data in real-time without disrupting the activity of existing sources, storage, and enterprise systems.

Real-time analytics is very different from the traditional analytics or batch processing that Big Data analysis has witnessed so far. In traditional analytics, analytic queries are run against historic data. Be it traditional analytics or the Hadoop-based batch processing, there is a lag time between the collection and storage of data sets, the processing for analysis, and the subsequent reporting of that analysis. In stream analytics, the system remembers the query, and every time the data changes, the answer changes based on the changed data set. It allows for high volumes of data processing in very little time.

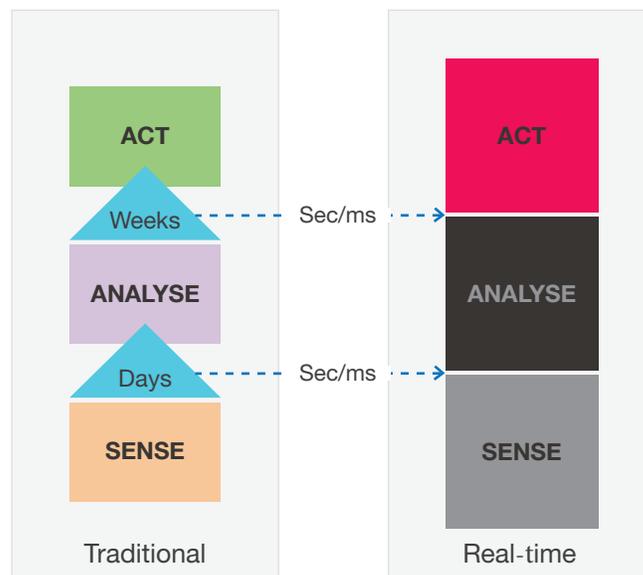


Figure 1: Difference between Traditional and Real-time Analytics

## Distinctive Characteristics of Real-time Streaming Analytics

Real-time streaming analytics distinguishes itself from other forms of analytics in the following ways:

- **Analysis of data in motion as it arrives:** The data processing could be of two types:
  - 1) Routine operations such as pre-processing followed by monitoring, reporting, statistics, and alerts
  - 2) Complex decision-making with real-time scoring based on predictive analytics models (and this could also include real-time learning)
- Immediate action after processing

In addition, a well-designed real-time streaming analytics platform allows businesses to:

- Capture all incoming events, including transactions, data outputs, and machine-to-machine communications to effective reception, inspection, analysis, storage, and distribution
- Store events, either later or in parallel to real-time analysis

## Business Value

Real-time sensing and analysis of data allows businesses to act with certainty, confident that subsequent actions are rooted in a relevant, timely understanding of the unfolding events. It makes complex decision-making easier. With zero data waiting time, nothing gets lost, overseen or outdated, because the velocity and volume of data is not an issue. The results of the analytics are translated and fed back into the local systems in real-time, which means the time lag between the incoming data and the outgoing data is as low as a few milliseconds.

Real-time streaming analytics also helps businesses by:

- **Visualizing the business in real-time:** Streaming analytics platforms come with dashboards and other visualization software to help people visualize, monitor, and analyze a huge amount of data from multiple sources in real-time. Many platforms provide tools to feed other dashboard/ visualization tools and/ or custom monitoring applications.
- **Cutting preventable losses:** Real-time streaming analytics helps to avoid preventable losses through early detection of at-risk situations.
- **Detecting urgent situations:** Application developers and business analysts can use the tools provided by streaming platforms to define analytical patterns of real-time business events.
- **Automating immediate actions:** Streaming analytics platforms work in the background until it detects an immediate risk or opportunity. Developers can design applications to issue alerts through email, text, push notifications, message queues, or service calls.

## How Real-time Analytics Can Add Value?

Forrester defines streaming analytics platform as a “software that can filter, aggregate, enrich, and analyze a high throughput of data from multiple disparate live data sources and in any data format to identify simple and complex patterns to visualize business in real-time, detect urgent situations, and automate immediate actions.”

Real-time streaming analytics must be sensitive to business concerns such as costs, resource demand, and time to market. What does it take to make real-time streaming analytics real?

At the highest level, it translates to an "always on" infrastructure that senses all business critical data, events, and transactions accurately, analyzes them, and is linked to appropriate and immediate actions.

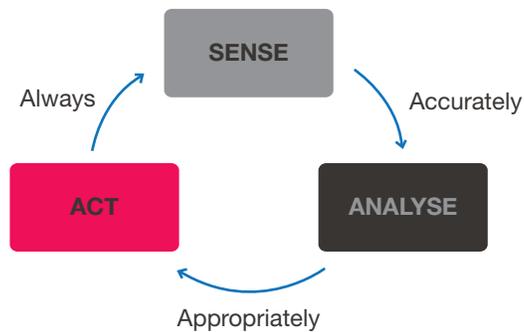


Figure 2: Systems must be designed as a feedback loop for continuous improvement.

In execution, this conceptual loop of sense-analyze-act can be achieved through a design built around three core components or “gears”:

- **Stream processing**, which ingests multiple high-volume streams at a high speed, routing received data in parallel for real-time analytics and batch processing
- **Real-time analytics**, which runs computations and/ or predictive analytic models, and triggers appropriate and immediate actions
- **Batch processing**, which stores and processes data-feeds offline and could build or update the predictive models used for real-time scoring

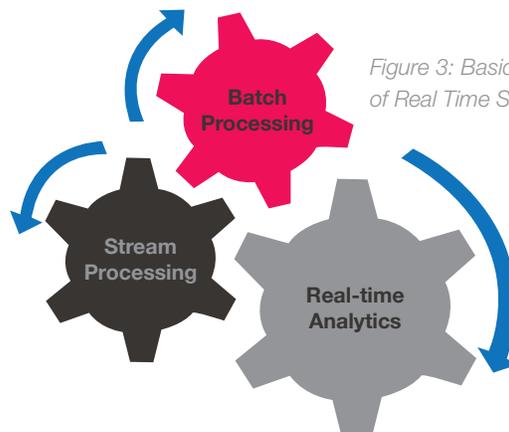


Figure 3: Basic “Gears” of Real Time Streaming Analytics

## Recommendations

At the architecture level, many approaches could work. But maximizing business value requires designs that can integrate with current systems and also provide a future-proof solution, which de-risks the company from technology flux. Optimal designs should allow for the following:

- **Loose Coupling or Abstract Layers** for businesses to choose the best components to suit their needs, without being locked-in to any one vendor. As technologies improve, individual elements can be replaced, upgraded, or improved without the need for a total system overhaul. Asynchronous streaming, facilitated by loose coupling, allows events to be processed without delays that are imposed by other elements in the system.
- **Open Source Platform** to avoid vendor lock-in and benefit from the community intelligence that accrues over time.
- **Event and Micro Batch Support** to enable both types of use-cases:
  - a) Low latency response to discrete events
  - b) Ability to create micro-batches where some delay is tolerable
- **Publish-Subscribe Model** that automatically distributes the right feeds to the relevant departments.
- **Visualization** that supports any standard interface, and can be integrated with any BI/ dash-boarding tool.
- **Reliability** that ensures no loss of data, supports both "exactly-one" and "at-least once" processing as required by the application and preserves the right sequence for time-series data. High availability should be preserved through automatic recovery sequences that reduce downtime to an absolute minimum.
- **Scalability** to stretch or compress elastically with demand by adding or removing incremental infrastructure.
- **Extensibility** to allow for quick and easy custom additions in functionality or modules.
- **Proven Platform** with pre-built analytical models to facilitate rapid operationalization, to accelerate time to market.

## How StreamAnalytix Can Help?

StreamAnalytix, an enterprise class real-time streaming analytics platform based on a best-of-breed Open Source stack can help organizations across industry verticals to quickly and reliably take into production a wide range of streaming data applications. It enables use cases in areas such as the Internet-of-Things (IoT), sensor data analytics, e-commerce and Internet advertising, security, fraud, insurance claim validation, credit-line-management, call center analytics and log analytics. It also enables enterprise IT and business transformation with horizontal capabilities like Streaming ETL to speed up slow batch processes to near-real-time.

To know more about the product and its architecture, visit [www.streamanalytix.com](http://www.streamanalytix.com)

### About StreamAnalytix

StreamAnalytix, a product of Impetus Technologies, enables enterprises to analyze and respond to events in real-time at Big Data scale. Based on Apache Storm, StreamAnalytix is designed to rapidly build and deploy streaming analytics applications for any industry vertical, any data format, and any use case.