gathr.ai

A Future-proof Architecture for Streaming Data Analytics

Are You Ready For Real-time?

Are you ready to start reaping the benefits of streaming analytics, but are worried about investing in new technology that could rapidly be surpassed or become outdated?

Technology is, and always will be, a moving target. So how can you stay ahead of the curve, take advantage of the latest technologies, while also ensuring that you don't make a mistake in your selection of a technology platform?

The answer lies in a strategy called "future proofing." The term "future proof" refers to the ability for something to continue to be of value, well into the future. This gives you the confidence that it will not quickly become obsolete. Obviously, future proofing provides a significant advantage in any overall technology strategy.

In this paper, we explore the demand for streaming analytics, discuss some of the available open source engines, the challenges of adopting them, and introduce a future-proof platform that uses an integrated development environment (IDE) (abstracted at a level above the streaming analytics engine) — enabling you to ride the future innovation wave.

The Value of Now

Streaming analytics is fast becoming a must-have technology for enterprises seeking to transform their analytics to take advantage of "fast data" sources and build real-time or near real-time applications.

Why? It's all about acting on "perishable insight," which is information that firms can only detect and act upon at a moment's notice. Streaming analytics is the key to harnessing real-time flow of perishable insights originating from the Internet-of-Things (IoT), mobile phones, market data, sensors, Web clickstream, and transactions. Real-time streaming is essential for Big Data analytics because it delivers insights in real-time when they are most valuable.

> **ACCORDING TO GARTNER**
>
> By the end of 2024, 75% of enterprises will shi from piloting to operationalizing AI, driving a 5X increase in streaming data and analytics infrastructures.

With the Big Data environment evolving daily and flooding companies with data, stream-processing capabilities are becoming increasingly important. Streaming allows organizations to take action in real-time and make critical decisions based on live data.

How Can You Benefit From RTSA?

The availability of real-time information sources and the IoT represent a new source of revenue and profit, product innovation, and investment opportunity. However, the real value comes from harnessing the high volume, velocity, and range of variety of streaming data being captured to process and apply analytics in real-time.

Real-time streaming analytics enable enterprises to collect, integrate, analyze, and visualize data as it is generated. It processes all the activities of the data as the data is being produced, without disrupting the activity of existing sources, storage, and enterprise systems.

Here are a few examples of how leading enterprises are using real-time streaming analytics to gain deeper insights and respond more quickly to changing conditions and opportunities:

Cut Prevantable Losses

Enterprises can cut preventable losses with real-time streaming analytics. Examples include:

- **Financial Services**

  The signs or signals of risky and/or fraudulent transactions can be imbedded in predictive intelligence models to alert firms to suspicious activities.

- **Preventive Maintenance**

  Sensors can detect and communicate wear-and-tear before they become machine failures that interrupt or cripple operations

- **Medical**

  In urgent healthcare situations, patient monitoring systems require real-time responses to alert professionals to urgent issues that demand immediate attention or intervention.

- **Brand Reputation**

  Given the wide adoption and influence of social media, brands are being discussed anywhere at any time. Automated monitoring can alert brand managers of the activity, and give them time to contain, control, or contribute to the brand conversation.

- **Disaster Warning Systems**

  Earthquakes, tsunamis, tornadoes, and other potentially devastating events can often be detected through a variety of monitoring and global sensing systems. Real-time streaming analytics can bridge the time lag between detection and communication, allowing authorities greater time to mobilize emergency responses.

Gain Operational Insights

Allows businesses to take swift actions and ensure uninterrupted operations.

Examples include:

- **IT Systems and Network Monitoring**

  Traffic, load, and threat vulnerability - these can be and controlled dynamically as conditions change.

- **Financial Transaction Processing**

  Authentications, validations, fraud detection must be processed in real-time.

- **Manufacturing**

  Closed loop control systems can monitor processes for variations and make adjustments to compensate for out-of-limit conditions in real-time.

The Complexity of an RTSA Platform

Streaming data is flowing by, and if it's not captured, analyzed and put to immediate use, the opportunity slips away. Consequently, streaming analytics requires blazing fast ingestion, analysis, and actions on multiple sources of fast data.
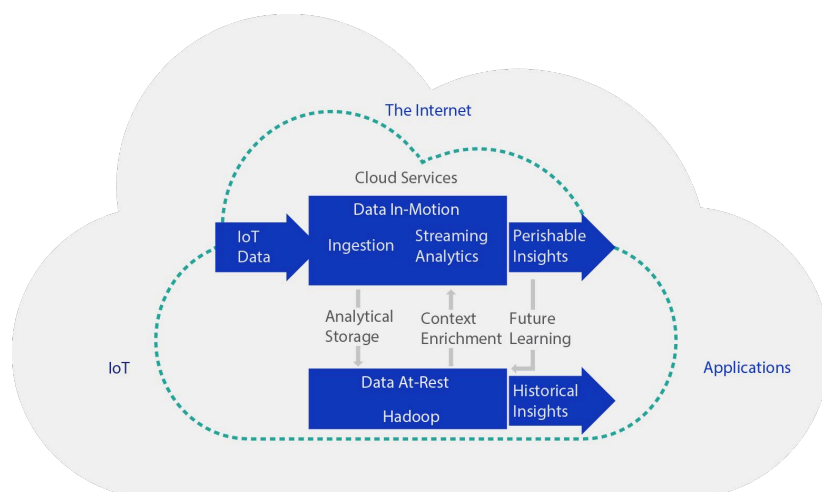
The streaming analytics platforms must be able to ingest, filter, aggregate, enrich, and analyze a high throughput of data from disparate live data sources to identify patterns, detect urgent situations, and automate immediate actions. That's the first complexity — everything the platform must do, and how rapidly it must do it.

> **THE STREAMING APPLICATION MODEL**
>
> In the streaming application programming model, the incoming data controls the code, unlike normal programming.

The second major challenge is that real-time streaming platforms don't resemble the types of analytic applications that enterprises are used to.

According to Forrester analyst Mike Gualtieri in the Impetus webinar, Future Proofing Your Streaming Analytics Platform, the streaming application programming model is unfamiliar to most application developers. It's a different paradigm from normal programming where code execution controls data. in streaming applications, the incoming data controls the code.

The Complex and Increasingly Expanding Data Map

The Allure of Open Source and the Dynamic Nature of the Market

Several offerings on the market seek to establish themselves as Big Data-processing platforms for large-scale data processing.

One of the best-known streaming technologies is Apache Storm. Storm has been battle-tested by many of the largest enterprises in the world, who routinely deal with large scale, real-time streaming data sources, including:

customer intelligence applications can be a business's greatest tool. With the ability to track individuals and their actions, businesses can harness this technology to create better customer experiences that are relevant and targeted. This is completely possible if data is analyzed in real-time. For instance, mobile coupon offering organizations are considering current location of their users collected via mobile GPS to instantly release geo-specific deals. The selection of the relevant deals is typically carried out in real-time by predictive matching of local merchants offers to the customers likely to accept them based on their past transaction history and other available CRM data.

- Cerner (Healthcare)

- Groupon/Yahoo!/Yelp (Social media)

- Aeris Communications (Telematics/IoT)

- Verisign (Security)Cerner (Healthcare)

Apache Strom is available as an open sources software (OSS); as is Spark Streaming - a comparatively new entrant for the Apache Spark parallel-processing framework.

Both Storm and Spark Streaming focus primarily on working directly at the level of the stream processing platform and prove difficult to integrate, often requiring additional tools. Despite the hype around them, enterprises struggle with both offerings' narrow focus and seem to rarely get beyond the proof-of-concept (PoC) stage.

Why then, given the difficulty of incorporating these tow open source technologies, are they so popular? Enterprises typically lean toward Apache Storm and Spark Streaming in the preliminary stages for several reasons.

Both projects are well known in the open source ecosystem around Hadoop. They are readily identifiable as alternative processing engines, and therefore as MapReduce alternatives. As they are available in open source, they can be easily acquired and installed to support research and development (R&D), prototyping and PoC projects.

Spark Streaming and Storm's well-known capabilities have forced companies to rethink their data-processing use cases and to consider the possibilities of stream processing to reduce time-to-action. As companies recognize the need for streaming analytics, the demand for an enterprise-grade streaming platform in the market increases.

The Allure of Open Source and the Dynamic Nature of the Market

With streaming analytics, remote human and machine activity can be sensed and responded to instantly.

Seize New Opportunities

Real-time analytics can stream data processed for one particular purpose to new opportunities, delivering value and generating new revenue streams. Examples include:

- **New Products**

  A major running shoe manufacturer has added sensors to its shoes to track exercise routines, offering consumers valuable new health insights.

- **Behavior Insights**

  Media companies and operators can analyze audience behavior to provide insights to drive new services and business models.

- **Soil Insights**

  When outfitted with sensors, tractors can add value by providing insights into soil conditions, such as humidity, temperature, and composition to farmers.

The Big Questions

Today, technology is finally building end-to-end components that make streaming analytics a reality. Tools that can process and analyze high-volume data streams are maturing rapidly, and new entrants — both commercial and open source — are arriving on the scene with surprising regularity. As a result, enterprise executives are asking themselves the following relevant questions:

What platform design will best address the needs of the market and drive rapid, large-scale adoption?

With so many new innovations continuously flowing into this domain, how do I choose a platform?

Is it smart to wait for the space to settle and then see who the clear "winner" is? Or, is there a way to get started now?

However, Storm and Spark Streaming are neither data-scientist nor application-developer friendly. The most significant feature that they lack is a graphical user interface (GUI) for building topologies. Without this, only developers with an understanding of Java or Clojure can build applications using these engines.

Data scientists and developers need intuitive visual tools to create streaming apps. Programming is tedious, and with Spark Streaming and Storm, developers are forced to manually account for scalability, handle input data skews, hand-code fault tolerance for the application and attempt to force event ordering/re-ordering.

Forrester's Wave report concedes that Strom and other open source software technologies have generated a great deal of buzz, but argues that they lack support, functions, and amenities of commercially-supported third-party offerings.

What should be done in this scenario?

The Need to Future-proof

One of the major obstacles that stands in the way of enterprises deciding to move forward — to select a vendor and to jump into the rapidly moving river of real-time streaming analytics — is the absence of a future-proof guarantee. This is especially true right now as the impact of innovation in this space is expected to continue, with several new open source projects already underway.

The ideal real-time streaming analytics platform would have an architecture with a common abstraction layer below the user interface that allows the selection of one or more streaming engines and change in engines — as your goals, requirements and strategies change. This abstraction layer would also streamline upgrades and embed new technologies into the existing system, based on relevance and credibility, with time.

But can one access such state-of-the-art technology? YES. And there is only one like it available on the market today - Gathr..ai

> ADVANCES IN REAL-TIME STREAMING ANALYTICS
>
> The technological advances related to real-time streaming analytics are moving and changing as rapidly as data itself.

The Ideal Platform

A real-time streaming platform must meet the needs of data scientists, developers and data center operations teams without requiring extensive custom code or brittle integration of many third-party components.

Maximizing the potential of data means analyzing data in motion as well as data at rest — from the moment it is created and after it is stored. Even better is to combine the two in an integrated lambda architecture. Achieving such a scenario may be simpler than you might think if you've got the right integration architecture in place, and the right tools for building analytics and integrations.

As we stated earlier, the question companies are trying to answer is this: How can we take advantage of new, innovative technologies while protecting our investment?

We know that the trend in today's Big Data world is to look for ways to leverage open source. This is because open source drives the potential of a vibrant stream of innovation coming from the global community. As a result, breakthrough technology is available to software developers at a low cost. This is not expected to slow down anytime soon; in fact, it is picking up speed, and investment is flowing into software companies that are pursuing this strategy.

However, two main factors hinder adoption:

- Firstly, unless software development is the industry you are in, you may not have the extensive team of highly-experienced developers required to code an application using open source, especially not without an IDE that offers a GUI developers interface

- And secondly, even if you did, you may fear that the speed with which technology advances may leave you behind no matter what

> **The gathr.ai platform**
>
> Gathr.ai offers the reliability, manageability, vendor support and professional service of an enterprise-class solution, while also availing you of the community innovation, low-cost, lack of vendor lock-in, and future proofing benefits that open source offers.

The answer is three-fold. You need a platform that:

- Provides an integrated development environment
- Utilizes open source
- Offers a level of abstraction above the open source engine that insulates you from the changing technologies that cloud cause disruptive change at the infrastructure level

By providing a hybrid option, Impetus' Gathr leverages the best of both worlds from enterprise class solutions and open source while mitigating the disadvantages of both. In other words, Gathr offers the reliability, manageability, vendor support and professional services of an enterprise-class solution, while also availing you of the community innovation, low-cost, lack of vendor lock-in, and future proofing benefits that open source offers.

The Gathr Advantage

In an effort to protet you from the dilemma, risk, and expense of building your own platform while not limiting or locking you into up-and-coming future innovations, Gathr has abstracted the graphical developer user interface above the level of the stream processing engine.

Gathr.ai is a state-of-the-art RTSA platform combining enterprise class with open source. Built on an engineering framework that articulates the essential requirements for modern businesses, Gathr delivers:

- Reliability

- Large-scale efficiency

- Robust intuitiveness

- A future-proof and context-efficient abstraction architecture

Open, flexible, extensible, and easy-to-use, Gathr rides right on the top of any standard Hadoop stack and is integrated with Apache Storm.

It also works with other NoSQL options like Apache Cassandra and Oracle NoSQL DB. Other choices for persistence and indexing can be easily integrated.

Enterprise grade - open source based - streaming analytics platform

Unified Business Interfaces | Common Utilities | Smart Workflows
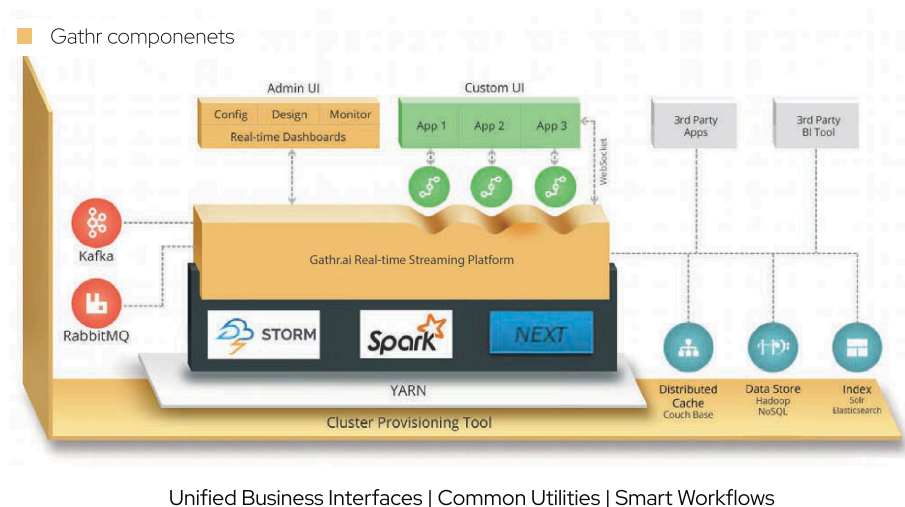
About Gathr.ai

Gathr.ai, 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 IoT, sensor data analytics, e-commerce and Internet advertising, security, fraud, insurance claim validation, credit-line-management, call centre analytics, and log analytics.

UNMATCHED, FUTUREPROOF TECHNOLOGY

Gathr.ai is equipped to support multiple streaming engines and customized extensions in order to deliver an unmatched level of technology future-proof investment protection

It also enables enterprise IT and business transformation with horizontal capabilities like Streaming ETL to speed up slow batch processes to near-real-time

Additionally, it is equipped to support multiple streaming engines and customized extensions in order to deliver an unmatched level of technology future proof investment protection.

Currently powered by Storm, Gathr is soon to be extended to include Spark Streaming as an added underlying stream-processing engine option. Additionally, Impetus Technologies has announced a statement of direction that Gathr will not only continue to support both Storm and Spark Streaming but will also evaluate other technologies as they become available, thereby keeping you ahead of the curve now and in the future.

Easily build fast and reliable data pipelines using Gathr.ai

Scan and
start free 14-day trial

Scan to
schedule a demo