A Complete Guide to Real-Time Analytics:

Use Cases, Best Practices, and Reference Architectures

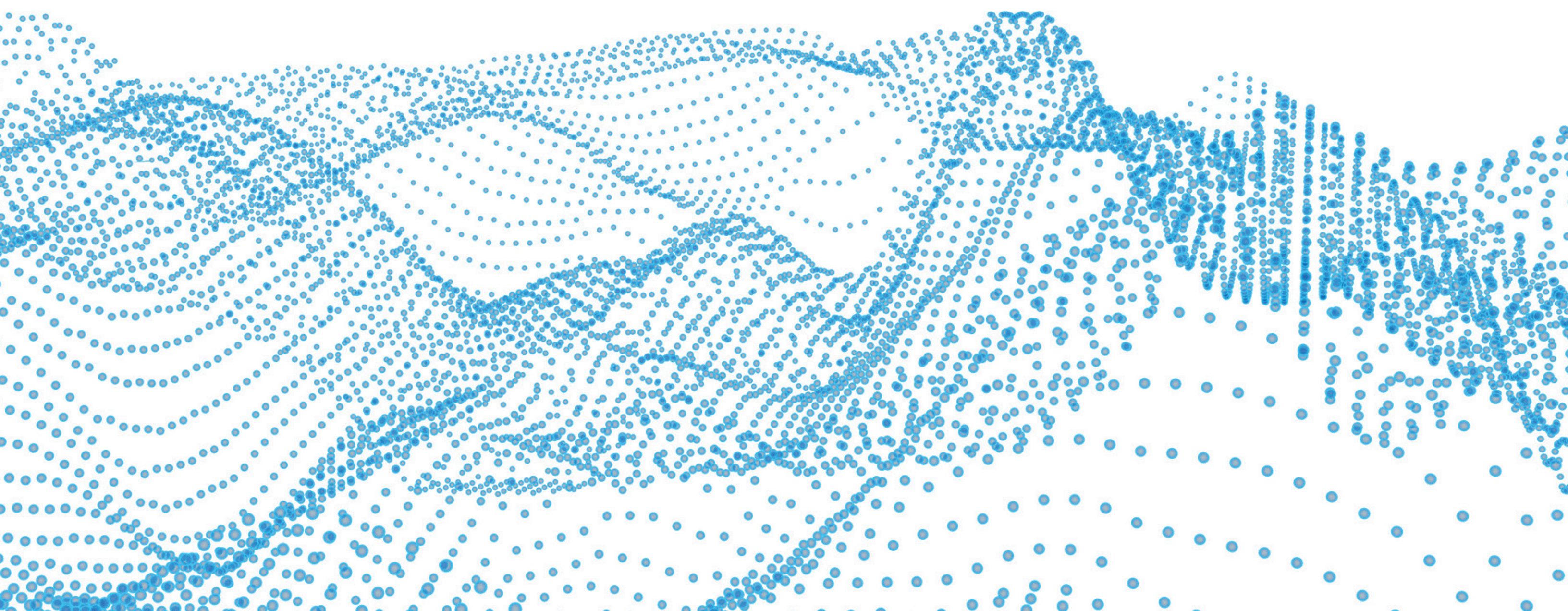


Table of Contents

Introduction to real-time analytics

As more companies rely on data to power business decisions and improve their competitive advantage, businesses realize that data has a shelf life. The sooner they can turn data into actionable insights, the faster it can provide value for the business. Therefore, companies are looking for ways to derive actionable insights from information faster than ever. Previously there was a general understanding that data processing took time. Business leaders got used to looking at yesterday's or last week's analytics results, and customers were tolerant of some latency in application data. However, we have become a culture of instant information, and businesses and users expect answers now, in real time.

In today's business landscape, companies that can make real-time data-driven decisions in response to business operations benefit by delivering higher levels of customer satisfaction, improving customer retention, and ultimately obtaining a larger competitive advantage.

Why real-time analytics is important

Real-time analytics allows organizations to capture live data streams, process them quickly, and extract insights or perform operations on the data in real time or near real time. It's based on stream processing technology that can handle a very high throughput of event data. Real-time analytics is already used in many of the applications we use today. For example:

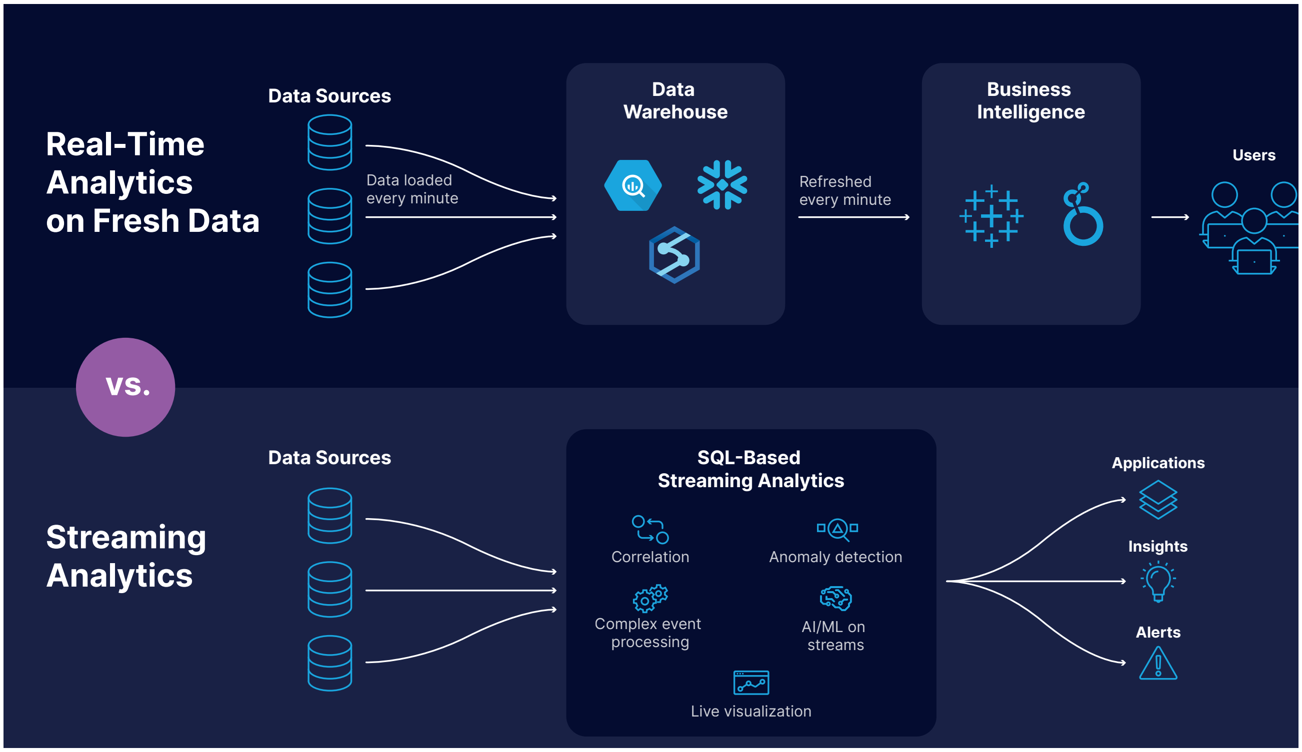
- Banks use real-time analytics to detect fraudulent transactions as they occur.
- Streaming services and e-commerce use real-time analytics to build recommendation engines to provide customers with better product recommendations and services.
- Real-time analytics can be used to personalize app experiences, like in the case of customized onboarding, where each page in the app changes to suit the user being onboarded. Consider a SaaS product: if a customer identifies as a "sales/ marketing" user, all the app onboarding content and features are tailored and set up for that target user.
- In sports like the NFL, real-time analytics provides teams with a per-second analysis of player performance and fitness to improve players and team output and monitor their health at the moment.
- Real-time analytics is also used for automation, like in self-driving cars, where the different car sensors and metrics are continuously evaluated and analyzed in real time to improve accuracy.

Without real-time analytics, all these different use cases would simply not have been possible.

Evolving from batch to real-time analytics

The way companies handle analytics has significantly changed over the years. Traditionally, most businesses' analytics platforms were designed around batch processing. Data generated from different sources were extracted, transformed, and loaded into a big data store, like a data warehouse, in pre-defined batches. Business users could then query these big data stores to retrieve information for reports and other business processes. Although this approach has proved successful, data volumes and use cases are on the rise, and companies and users demand insights from data in seconds, not hours or days.

To meet the demand for faster data processing to support faster analysis, companies have often looked to optimize their existing architectures using techniques such as manually scaling out data transformation processes or reducing batch processing intervals. This technique has proven successful. However, data availability for analytics was still generally limited to minutes, at best. Furthermore, businesses seeking to extract insights about their customers, products, or applications in real time found it hard to do so.



Real-time analytics two ways: on fresh data at rest vs data in motion. Fresh data is essential to business intelligence teams that want to build live reports from data housed in a cloud data warehouse or data lake. Streaming analytics enables correlation, anomaly detection, complex event processing, and machine learning on data streams.

Real-time analytics provides a solution to a number of the challenges with traditional batch processing. In a real-time analytics approach, business events are collected and processed continuously. As shown in the above diagram, real-time analytics encompasses both streaming analytics and analytics on fresh data at rest. In streaming analytics, or analytics on data in motion, data is continuously queried, giving companies the ability to respond to business events as they happen. Companies can also continuously supply their business intelligence and analytics teams with fresh data for analytics in cloud targets such as data warehouses and data lakes. This way, they can ensure they have up-to-date reports in their business intelligence tools (e.g. Tableau or Looker).

While enabling analysis over fresh data or data in motion should be considered for a variety of existing business challenges, it’s not meant to be a replacement for batch-oriented analytics architectures. Many cases still exist where batch analytics is preferred, for example, in a payroll and billing system where the information is processed weekly or monthly. Companies enjoy the best of both worlds when batch analytics is combined with real-time streaming to cater to multiple business use cases.

Architectural requirements for real-time analytics

Implementing real-time analytics requires a different architecture and approach than traditional batch-based data analytics. This is in addition to a different set of tools and technologies that can support the streaming and processing of large volumes of data. When it comes to real-time analytics, raw source data rarely is what you want to be delivered to your target systems. More often than not, you need a data pipeline that starts with data integration and then enables you to do several things to the data in-flight before delivery to the target.

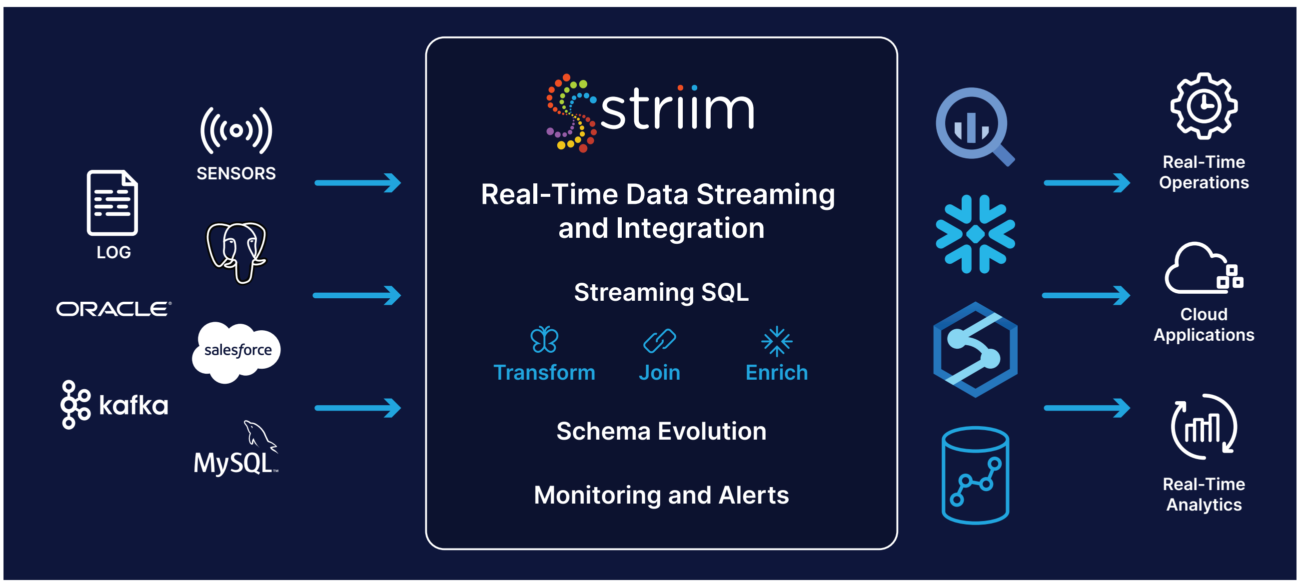
Data integration

The data integration layer in the real-time analytics architecture provides capabilities for continuously ingesting data of varying formats and velocity from either external sources or existing cloud storage. The data integration layer is the backbone of any analytics architecture, as downstream reporting and analytics systems rely on consistent and accessible data. Therefore, the integration channel must be able to handle large volumes of data from a variety of sources with minimal impact on source systems and sub-second latency. This layer leverages data integration tools such as Striim to connect to various data sources to ingest streaming (data in motion) data and deliver it to various targets.

For example, the Striim platform enables the continuous movement of unstructured, semi-structured, and structured data — extracting it from a wide variety of sources such as databases, log files, sensors, and message queues, and delivering it in real-time to targets such as Big Data, Cloud, Transactional Databases, Files, and Messaging Systems for immediate processing and usage.

Event/stream processing

The event processing layer provides tools and components for processing data as it is ingested. Data coming into the system in real-time are often referred to as streams or events because each data point describes something that has occurred in a given period. Often, these events need to be cleaned, enriched, processed, and transformed in flight before they can be stored or used to provide value. Therefore another essential component for real-time data analytics is the infrastructure to handle real-time event processing.



Event/stream processing with Striim

Some data integration platforms like Striim perform in-flight data processing such as filtering, transformations, aggregations, masking, and enrichment of streaming data before delivering with sub-second latency to diverse environments in the cloud or on-premises. In addition, Striim can also deliver data for advanced processing to stream processing platforms like Apache Spark and Apache Flink, which can manage and process large data volumes with advanced business logic processing.

In a stream processing architecture, analytics and data consumption can also be performed in real time. Instead of reading finite data sets in batches from tables, streaming queries or applications can ingest real-time event streams and continuously produce and update results to data consumers as more events are consumed.

Data storage

A real-time analytics infrastructure also requires a scalable, durable, and highly available storage service to support the massive volumes of data required for various analytics use cases. **Data warehouses** and data lakes are the most commonly used storage architectures for big data. Companies looking for a mature, structured data solution that focuses on business intelligence and data analytics use cases may consider a data warehouse. However, data lakes are suitable for enterprises looking for a flexible, low-cost big-data solution to power machine learning and data science workloads on unstructured data.

All data required for real-time analytics is rarely contained within the incoming stream. Applications deployed to devices or sensors are generally built to be very lightweight and intentionally designed to produce minimal network traffic. Therefore the data store should be able to support data aggregations and joins for different data sources and must be able to cater to a variety of data formats.

Presentation/consumption

The core of the real-time analytics solution is a presentation layer to showcase the processed data in the data pipeline. When designing a real time architecture, this step must be at the forefront because it's ultimately the end goal of the real-time analytics pipeline.

The data consumption layer provides analytics across the business for all users through purpose-built analytics tools that support analysis methodologies such as SQL, batch analytics, reporting dashboards, and machine learning. This layer is essentially responsible for:

- Providing visualization of large volumes of data in real time.
- Directly querying data from big stores, like data lakes and warehouses.
- Turning data into actionable insights using machine learning models that help businesses deliver quality brand experiences.

Best practices for real-time analytics

Real-time analytics requires a high degree of proficiency and understanding before companies can use it to meet critical real-world business needs. This section outlines five ways an organization can adopt best practices for real-time analytics and drive the technology toward its maximum potential.

1. Focus on the right use cases for real-time analytics

While real-time data process/analytics can be game changing, it's not practical for all situations. In some situations, delivering real-time analytics is a waste of money and counterproductive. For example, you probably don't need to analyze your company's monthly financial report in real time. Setting up real-time analytics data pipelines for daily or monthly reports can be overkill.

If you're a high transaction volume business and want to maximize revenues (e.g., hotel reservations, airline reservations, online e-commerce), real-time analytics can provide insights that allow you to respond instantly to market opportunities with promotions to the most likely customers. For example, a hotel reservations company can create personalized customer booking experiences based on events (music events, sporting events, award shows) happening in a particular location at a particular time, ultimately improving customer experience, loyalty, and company revenue.

Before starting on a real-time analytics project, it's important to prioritize to determine the benefit or outcome the pipeline will provide to your firm or product and where real-time analytics will benefit the company. List top areas where you need to know the information right now.

2. Build a reliable and robust infrastructure

To ensure you maximize the benefits of real-time analytics, the underlying technological infrastructure has to be strong, flexible, and robust to support real-time analytics requirements. Disappointment might occur when a company expects accurate real-time insights but fails to achieve them simply because its infrastructure does not function as expected. A good real-time analytics infrastructure usually includes:

- The infrastructure to ingest large volumes of data into the data pipeline.
- A component to process and analyze the live data stream in real time with minimal latency.
- A component to store large volumes of real-time data.
- Support for various analytics methods, including real-time alerting, dashboards, and data science use cases.

Additionally, think of where you'll manage your data and analytics: in the cloud or on-premise. Finally, ensure you choose a solution that enables your company to scale quickly. For example, the real-time analytics infrastructure to process 50,000 transactions will differ vastly from that needed to process 1 million transactions. Building a reliable and scalable infrastructure that can serve you as your business grows is critical.

3. Combine real-time data with historical data

Real-time data's value increases exponentially when merged with historical data, enabling end users to combine and compare insights "in the moment." Real-time data allows businesses to make decisions quickly, and historical data provides more context for real-time data. For example, a credit card transaction scoring model could be created once using historical data and then used to evaluate real-time credit card transactions for days or weeks. When historical data is mapped over time, decision-makers can gain a richer understanding of how a particular scoring model is performing.

Companies need to adopt a hybrid approach of combining historical and real-time data to meet customer expectations and fulfill business demands for smarter decision-making.

4. Develop a monitoring and governance framework

Develop a proper monitoring and governance framework with administrators and business leaders who constantly monitor real-time analytics performance from an aggregate top level. Any deviation from the governance strategy needs to be corrected, and systems guardrails should be in place to stop processes when the need arises, for example, ensuring that sensitive customer data is not being used for experimental analytics purposes. These "checks" help refine the analytics models and lead to flexible business rules.

It's also important to track the results to make sure the models are working correctly and tweak rules and analytics on a regular basis to get the right results for quick decision-making. For example, a fraud detection system that does quality control using real-time analytics can keep flagging particular transactions as fraudulent because the system hasn't encountered similar transaction parameters before. However, it might not mean the transaction is fraudulent, and this could ultimately lead to a terrible customer experience and loss of revenue for the business. So, you need a governance and observability framework that monitors what goes on beneath the surface as machines make hard decisions.

5. Provide an extensible data access layer to cater to multiple use cases

An easy and flexible data access layer is critical to the success of real-time analytics. It enables the business to quickly utilize the value from data for a plethora of use cases. Different company areas require data for analysis for disparate purposes, for instance, in an e-commerce company. Inventory managers require product data to analyze purchasing decisions, sales and marketing teams require customer purchase data to discover trends and make better advertising campaigns, finance teams require data to create budget estimations, and data science and application teams require data to build better product recommendation engines and improve the overall customer experience. Each of these teams requires different types of data and interacts with these data in separate ways, and the real-time analytics framework needs to serve all of these needs at the same time.

In order to realize the benefits of real-time analytics and serve these different teams efficiently, there needs to be a focus on the data delivery mechanism, ensuring that data is easily accessible to the various teams who need it. As a result, more companies these days are investing in concepts like the data mesh architecture that focuses on democratizing data access and accelerating the time in which companies can get the data and convert it into insights for the business.

Real-time analytics practical use cases using Striim

Companies can apply real-time event streaming analytics to a variety of industry scenarios. This section provides examples using real-life systems and reference architectures from our partners to provide information on how to implement real-time analytics and how it can add value for businesses.

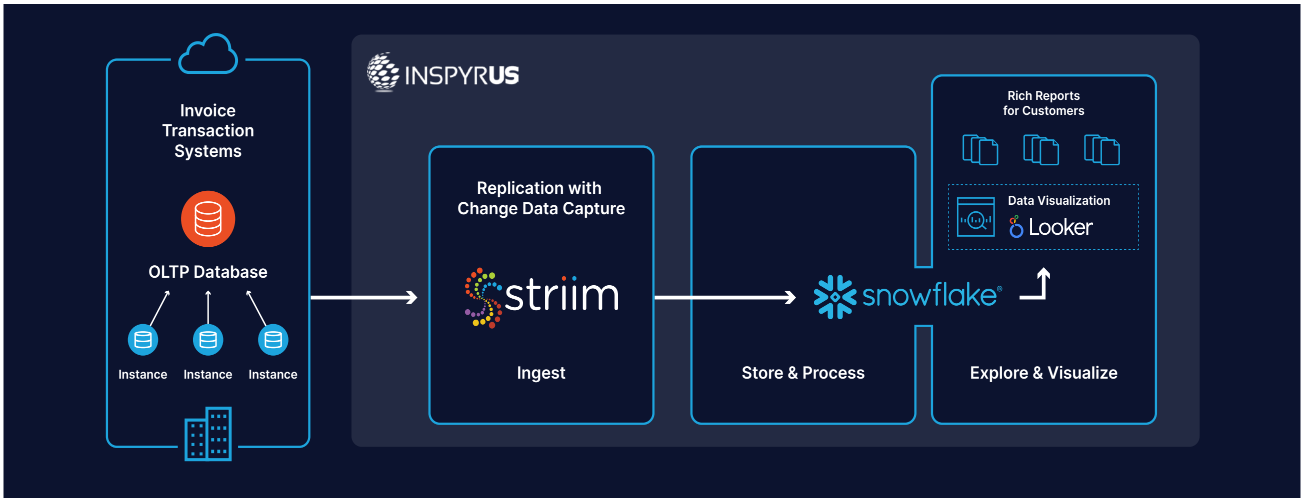
Inspyrus: Enhancing accounts payable automation services

Inspyrus, a MineralTree Company, is a FinTech company that helps businesses reduce costs and unlock accounts payable (AP) efficiencies. The company's offerings include a comprehensive software-as-as-service (SaaS) solution that automates up to 90% of invoice and payment processes.

Use case:

To support the SaaS solution, Inspyrus relies on an online transaction processing (OLTP) database. While this delivered the performance and reliability the company needed for operational usage, extracting data for reporting purposes required an hour-long refresh cycle. Inspyrus looked for a new approach that would enable them to:

- Provide rich reports for its SaaS customers based on real-time operational data
- Store and analyze all data in a scalable cloud platform for high availability and global reach
- Make improvements to data flows and applications in hours instead of weeks



Inspyrus' real-time analytics architecture diagram

Inspyrus chose Striim for their superior database **change data capture** capabilities and integration with Snowflake. Today, **Inspyrus** uses Striim to capture data from its OLTP databases, hosted in the company's private cloud. The solution extracts, transforms and loads changes to **Snowflake in real time**, while respecting the order and transactionality of the changes. By integrating Snowflake with the dbt transformation workflow and Looker Business Intelligence Platform, Inspyrus enables customers to explore their data in real-time through rich, intuitive reports. Striim has also allowed Inspyrus to slash time to market when launching new capabilities in its data stack.

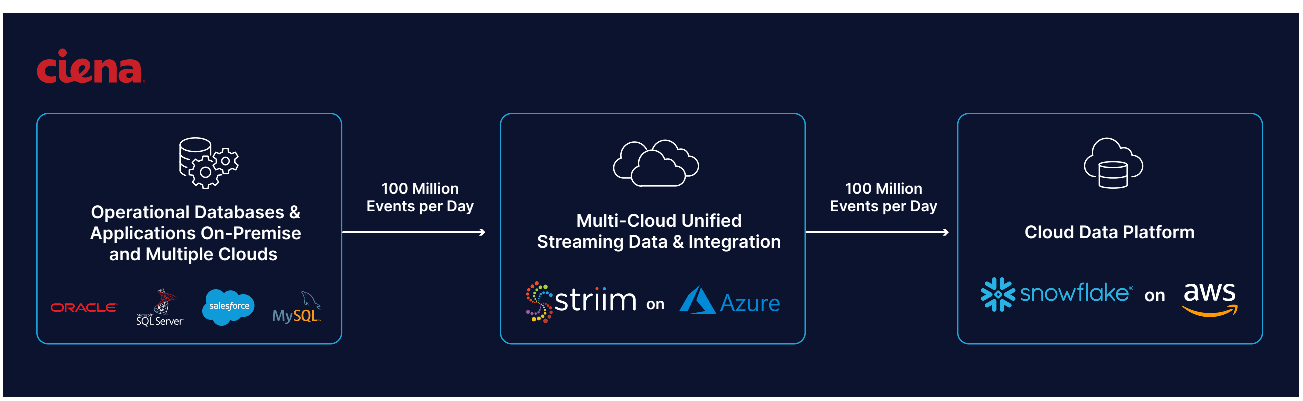
Ciena: Enabling quicker data-driven decisions with real-time analytics

Ciena is an American telecommunications networking equipment and software services supplier that provides best-in-class networking solutions to support 85 percent of the world's largest communications service providers and submarine network operators, data and cloud operators, and large enterprises. For over 20 years, Ciena has focused on its mission to deliver richer, more connected experiences for businesses and users, enabling customers to adapt to the ever-changing technology sector.

Use case:

Ciena's data team wanted to build a modern, self-serve data and analytics ecosystem that:

- Improves customer experience by enabling real-time insights and intelligent automation to network changes as they occur.
- Improves data access across the enterprise by removing silos and empowering every team to make data-driven decisions quickly.



Ciena's architectural diagram

To meet its goals, Ciena chose Snowflake as its data warehousing platform for operational reporting and analytics and **Striim as its data integration and streaming solution** to replicate changes from its Oracle database to Snowflake. Striim was responsible for collecting, filtering, aggregating, and updating (in real time) 40-90 million business events to Snowflake every day across systems that manage sales, manufacturing, accounting, and dozens of other crucial business functions to enable advanced real-time analytics.

By leveraging its real-time analytics platform, Ciena has offered customers up-to-date insights as changes occurred in their network, thus improving the customer experience. For example, real-time analytics help Ciena customers better understand and optimize their network resources by balancing and utilizing them in the most optimal way possible. Additionally, operators can begin experimenting with machine learning by using real-time analytics to proactively identify network events that could impact performance.

Real-time analytics is transforming how businesses operate

As data volumes continue to grow, real-time analytics will become increasingly relevant to more companies and industries

by accelerating the time to insight, enabling faster data-driven decision making, improving the customer experience, and offering a competitive advantage.

By complementing traditional batch-based analytics implementations with real-time analytics, companies can explore a new level of decision intelligence that was impossible a few years ago. From hyper-personalizing the online shopping experience to optimizing telecommunication networks in real-time, businesses in every industry are leveraging real-time analytics to help their customers and companies navigate this data-driven world.

About Striim

Striim was founded with a simple goal of helping companies make data useful the instant it's born.

Striim's unified, real-time data streaming and integration platform for analytics and operations collects data in real time from enterprise databases (using non-intrusive change data capture), log files, messaging systems, and sensors, and delivers it to virtually any target on-premises or in the cloud with sub-second latency enabling real-time operations and analytics.

Try it now at **go2.striim.com/free-trial**

Contact us at:

Tel: +1 650 241 0680
Email: **sales@striim.com**
Web: **www.striim.com**

striim