Stream Processing:

Instant Insight Into Data As It Flows

Catalyze Business Opportunities
with a Powerful New Perspective on Data

RTInsights.com

hazelcast

CHAPTER 1

From Data Exhaust to Real-Time Insight

As the digital economy matures, the way data is processed is changing rapidly. Mobile applications, sensors, and other always-on technologies are constantly sending seemingly infinite streams of data. And enterprises need to process that data instantly—or as close to real time as possible.

That's where stream processing can help. Stream processing lets you sense, query, analyze, and process data as it flows. Depending on the application, the technology can detect conditions within as little as milliseconds—making it perfect for time-sensitive transactions such as payment processing, alerts, problem detection, and real-time analytics.

Yet many business and technology leaders believe they aren't ready or able to adopt stream processing. They question whether their enterprises and IT infrastructures are prepared for the new paradigm, despite the superior processing power, speed, and results the technology enables. However, the introduction of new stream processing capabilities and tools that simplify deployment and reduce time to value is changing minds.

Think of it this way: Every day, customer expectations are rising. Stream processing can help you give customers what they want and expect, faster and more accurately than ever before.

How Stream Processing Delivers Value

Get faster, real-time insights.

Get faster, real-time insights. By analyzing data as soon as it is created and acting on that insight, you can respond faster to time-sensitive events.

Identify important data trends and patterns.

Increasing your understanding of data trends helps you develop actionable insights—in areas such as stock trade analysis, predictive maintenance, or manufacturing quality control—that would be masked by state-based, batch-oriented analysis.

Simplify communications.

Using stream processing as a messaging layer between systems enables data enrichment, helping you create a more complete picture of your data.

Stream processing offers a new perspective on your data (see Figure 1). When you use stream processing, you can see the order of data processing events or the time allocated for events within each application. The technology offers insight into those changes and what they mean to your business. Stock prices are a classic example, where a downward trending intraday stock price could predict a significant price drop. In addition, stream processing increases the granularity and resolution of data, to identify critical events. For instance, data changes may represent an urgent situation, like an intermittently overheating machine, which might otherwise be missed by periodic polling of the temperature.

Unlike a traditional database view, which offers a static pool of data or a snapshot view, stream processing supports data that changes over time and the ability to understand those changes. It's the difference between measuring how fast a vehicle is traveling and how quickly it is accelerating. In the first case, you get a snapshot view—the vehicle's speed in a single moment. In the second, you gain insight into what is happening over time—the rate of change as the vehicle speeds up or slows down. That's insight you can't get with traditional data processing technologies.

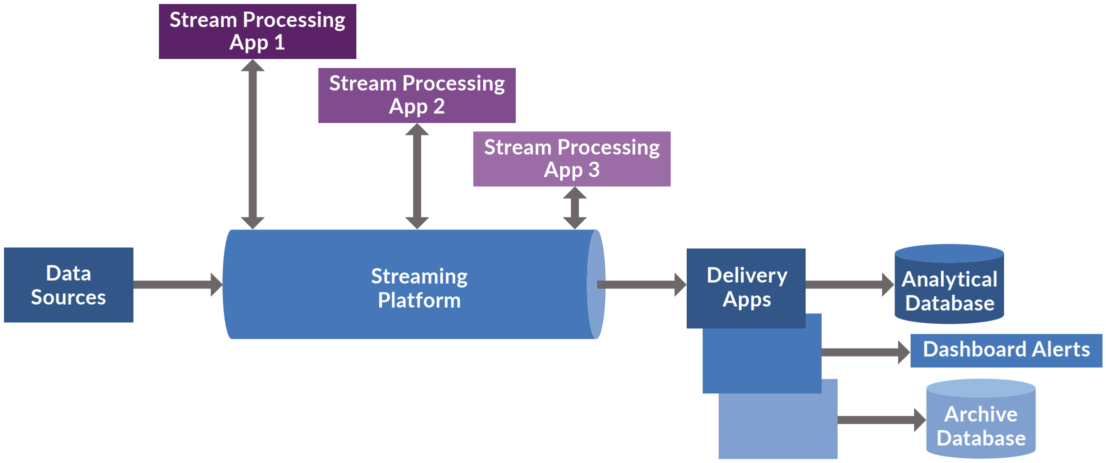

Figure 1. Simple Stream Processing Architecture

What's more, stream processing is applicable to more than just real-time data to handle traditional, large-scale data processing jobs. It easily lets you create many different outputs for various types of analysis. For example, you can aggregate and transform raw data in different ways—such as using sales data to determine total revenue per region, total sales by state, or number of customers per country. This capability is represented by the Kappa Architecture (see Figure 2), in which a stored stream of data is transformed in multiple ways into an analytics database. The Kappa Architecture is a great way to model an analytics environment where many different queries need to be run by a range of end users.

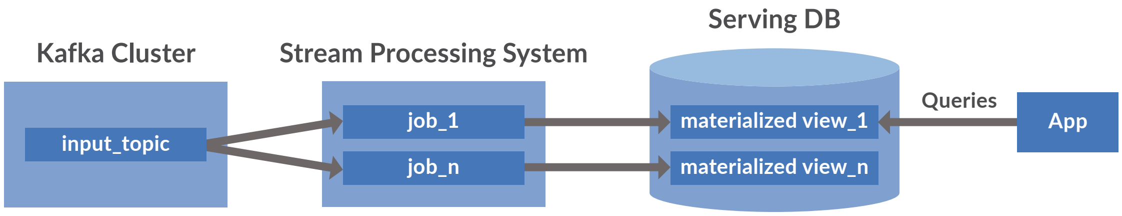

Figure 2. The Kappa Architecture for Analyzing Streaming Data

By allowing you to see your data over time, stream processing also enables deeper understanding by seeing previously hidden and non-obvious trends and patterns, which in turn catalyzes new opportunities and innovations for your business. The real-time aspect of streaming also lets you respond faster to changing conditions. Let's explore how real-time insight enabled by stream processing can benefit your business.

CHAPTER 2

Insight-Empowered Business Optimization

Stream processing technology can make existing data processing systems better, faster, and more reliable. By using all of your data output to enhance your business opportunities, expand existing opportunities, and react faster or more effectively to problems, the technology helps you run your business run better. Following are several use cases that illustrate the promise of stream processing technology.

Predictive Analytics

To better anticipate and avoid machine downtime, one leading manufacturer streams sensor-based processing data from the shop floor with production data from its machine cells. Bringing these data sources together in real time and adding machine learning capabilities helps the company view current activity—which machines are producing as scheduled and which manufacturing cells are reporting quality issues—to understand when production should be temporarily halted to address issues that will impact downstream quality and yield. This proactive approach can save on costs that otherwise would be wasted on defective work in process. Stream processing also helps the company know the optimal times to perform equipment maintenance by analyzing sensor measurements to predict when parts face impending failure. This insight helps the manufacturer avoid catastrophic machinery failure, reduce unplanned downtime, and enhance product quality. This approach is more efficient than scheduled maintenance since it reduces planned inspections. This is significant considering that inspections can be costly, so minimizing the number of inspections enable more savings.

Fraud Detection

Financial services companies want to protect their customers from fraud, but they also need to process transactions quickly. They also need to minimize the number of false positives, which may lead customers to a different payment method, resulting in lost transaction fees for the company. To balance these conflicting requirements, innovative banks are using stream processing to run multiple fraud detection algorithms in just milliseconds. Using in-memory storage, artificial intelligence, and machine learning, the solutions ingest, validate, and score vast amounts of payment data with ultra-low latency. By running more fraud algorithms within a millisecond-level SLA, banks achieve greater accuracy in fraud detection—protecting against fraudulent transactions without delaying consumer purchase transactions or losing out on transaction fees due to false positives.

Payment Processing

The exponential growth of mobile devices is complicating the payments infrastructure for leading financial services companies. Facing a dramatic increase in mobile payments and rising consumer expectations for real-time processing, one bank decided to deploy stream processing to handle data processed at the network edge. By integrating processing with an in-memory data store, the firm can process data when it is generated or ingested. The solution classifies workloads in real time, enabling instant responses to live transactions.

360-Degree Customer View

Imagine a consumer who attempts to place an order using the website, sends an email, and then talks to a chatbot. Previous data processing technology could not recognize that the customer might be having a problem. Using stream processing technology, one retailer began collecting and analyzing customer data from a variety of sources. By correlating these events in real time, the retailer is able to recognize when customers need assistance. A service representative can contact the customer and try to solve the problem before the consumer defects to the competition.

Financial Trading

Traditional technologies tend to place data into artificial buckets, such as transactions per day or a month, for example. But one trading company deployed stream processing technology to break that mold. Instead of analyzing single sources of data—such as completed trades—the firm streams data from multiple sources, such as stock market surveillance sites, weather services, and customer websites. By streaming and analyzing a more heterogeneous collection of data, the company was able to generate new insights that help traders take advantage of the best opportunities.

Fleet and Operations Management

The operations of large retail franchises are notoriously multifaceted and complex—store operations, supply chain issues, and logistics tend to remain isolated in their own data silos. But one national chain restaurant deployed stream processing to bring these worlds together. By streaming various types of data about sales transactions, customer feedback, deliveries, and supplies, the company gets comprehensive insight in an instant—whether a truck breakdown prevents a supplier from delivering goods on time, or a franchise experiences quality problems.

CHAPTER 3

Essential Technology Capabilities for Stream Processing

Streaming represents a different way of thinking in how data is processed. It may not always be obvious, but the use of data streams is often a more practical way to handle, process, and analyze your data. And modern streaming technologies offer newer capabilities that are not found in earlier generations of streaming tools. Consider the following technology capabilities that might benefit your architecture to get more value from your data.

Multiple Data Streams

Stream processing is not only about handling one data stream at a time. Most enterprises rely on various sources of data—all with different frequencies and formats—in multiple streams. Most raw stream data is often very sparse, lacking details about sensor locations or time in operation, for example. By bringing the data together and blending different sources into a single stream, you can transform and enrich the data, which provides more context. Analysis of this enriched data can yield new relationships, patterns, and trends that were not obvious before.

Time Windows

Unlike batch processing, stream processing has no end, which means you need to decide how data is grouped together. This is done by specifying the windows of time in which you analyze data, which will affect your results. Depending on the data, you may want to analyze anywhere from 30 seconds to 24 hours or more. Beyond the capabilities of previous technologies, today's time windows can include the following:

- **Tumbling windows:** Where the time window size is fixed and each window is non-overlapping (see Figure 3)
- **Sliding windows:** Where the time window size is fixed and each window advances incrementally, creating overlap
- **Session windows:** Where the time window size varies based on the grouping of data and on the time elapsed

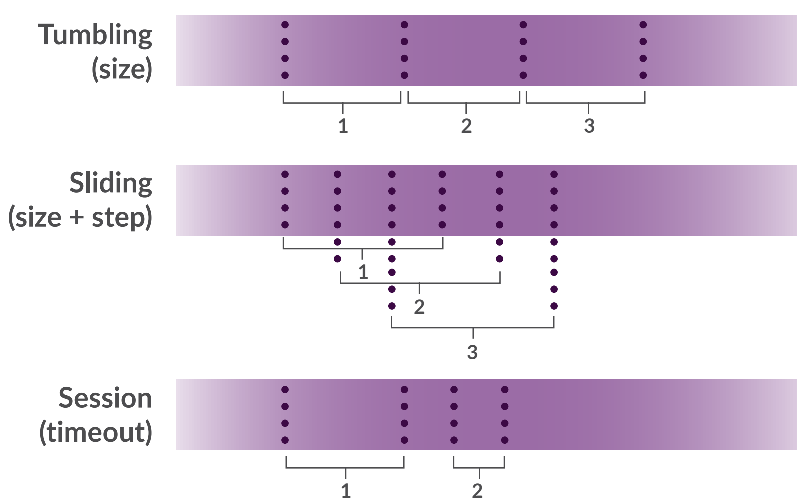

Figure 3. Window Types

Data Enrichment

Data enrichment is an important technique for creating new insights from stream processing. It offers an easy way to connect different, well-known data sources, especially databases, into a single repository of information. When you join or correlate data from two separate streams into one, or you join a stream with batch data, the result is enriched data that supports enhanced analysis. You can then create aggregations such as calculating average and mean as well as standard deviations in the specified time windows.

Microservices Statefulness and Communications

Microservices give you the ability to add new capabilities or applications to your architecture without overhauling existing applications. In a microservices architecture, you can use data streams as a lightweight communications protocol for passing data and messages from one microservice to another (see Figure 4). Each microservice is a stream processing application that takes the incoming stream of data, processes it, then passes it to another stream for eventual consumption by another microservice.

For example, assume that your e-commerce engine takes orders, processes them, and then archives the order data for internal purposes. What if you wanted to be able to alert customers looking at online product pages that they have already purchased a product? You wouldn't be able to retrieve that information quickly from your internal archive database.

However, you could modify your e-commerce application to deliver order information to a stream. That stream could then be used to create a fast, lightweight order lookup database that can insert a "previously purchased" notation into the relevant product pages—making the pages appear to be custom-built for each customer.

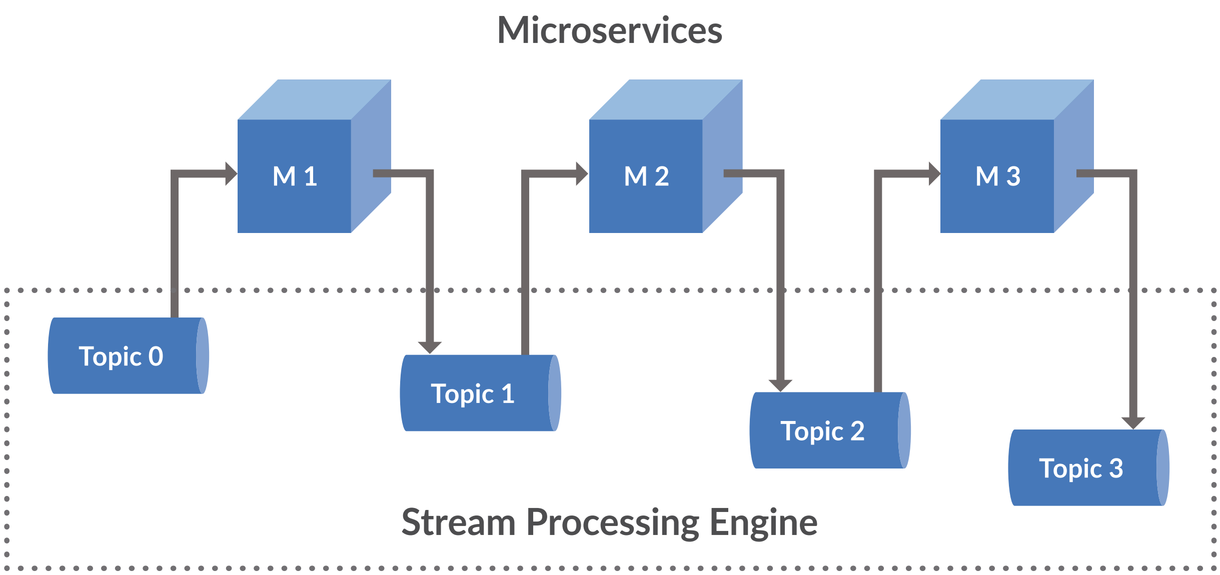

Figure 4. Microservices Architecture

Machine Learning and Artificial Intelligence

Machine learning (ML) and artificial intelligence (AI) systems are often used with streaming data. A stream processing architecture provides the infrastructure for feeding live data into a deployed ML model or an AI application.

Combining stream processing with ML/AI enhances processing speed and increases analytical power. For example, you could decide to split a data stream to feed separate machine learning algorithms in parallel. You could use it to support completely different models or different versions of the same model, which would be useful for A-B testing (see Figure 5).

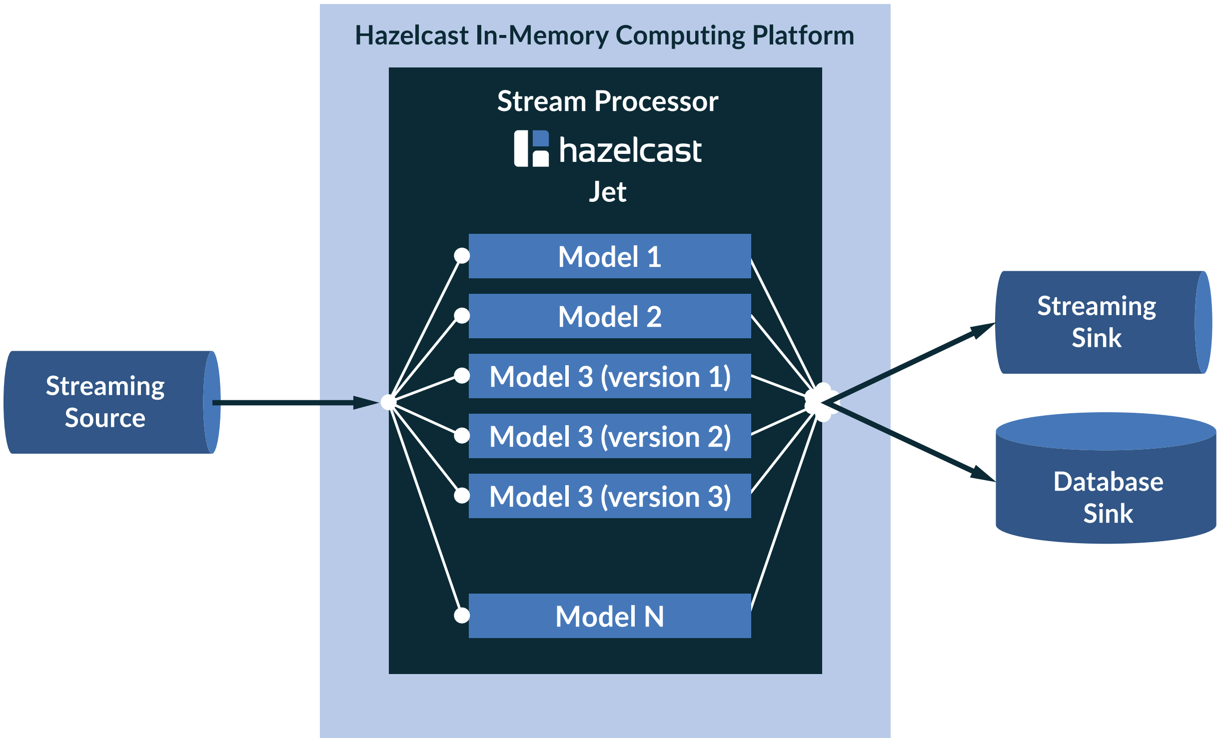

Figure 5. Machine Learning Models in Parallel

Another option would be to join the outputs from different machine learning models to summarize the outputs of both models in one stream. Let's say model 1 produces an output of score A and model 2 creates an output of score B. If you combined score A with score B to create a composite score, you could compare that score with other scores to assess the performance of each model.

CHAPTER 4

Planning a Stream Processing Architecture

The complexity of stream processing requires a powerful yet well-constructed technology architecture that addresses the following needs.

Low Latency

To keep up with the rapid pace of data ingestion and enable single-digit millisecond or microsecond responses needed for real-time analytics and processing, your architecture must support low latency. Low-latency architectures are required, for example, when you are supporting millions of concurrent users or simultaneous transactions.

Several technological characteristics can enable low latency. As a start, an architecture that can scale horizontally—by adding more instances to the grid, to spread the workload across more cores—is critical for meeting new, extremely distributed, and parallel use cases. Scaling up is also important, to be able to take advantage of more cores and more memory within each instance. Along with a scale-out and scale-up architecture, the system needs to be able to optimize work across all available resources to achieve the lowest latency. Low-level optimizations such as the use of green threads help to reduce the overhead of OS-level threads, further reducing latency. Integrated components like an in-memory store also help to reduce overall processing latency.

In-Memory Processing

Stream processing requires a distributed computing platform that manages data using in-memory storage. Stream processing typically requires enriching the stream with data lookups, and having an in-memory store enables the fastest processing. Architectures that rely on in-memory processing can run applications quickly, delivering faster insights and engaging new data as it is ingested.

Modern Hardware

The cost of random-access memory (RAM) continues to fall, but RAM-based products are still at a cost disadvantage when compared with solid-state drive (SSD) technology. When choosing between higher performance that helps them do more with data and spending less money on storage and doing the same things as before, enterprises often choose the lower-cost option. That's about to change. Newer innovations that reduce the total cost of ownership for in-memory technologies are creating options for companies striving to do more by implementing faster systems.

Stream Processing at the Network Edge

Enterprises are increasingly generating more data from devices at the network edge. Yet edge devices are often ill-equipped to process large volumes of data. With space restrictions, bandwidth limitations, scaled-down hardware, and even security concerns, edge processing requires software technologies that are designed for high performance despite such constraints. An architecture designed to process or aggregate data at the edge can reduce workloads at the central data center. Some solutions can even perform certain analysis tasks at the edge, reducing processing time and speeding the delivery of results.

Cloud Processing

Companies are beginning to use public cloud offerings as an all-purpose data center, instead of relying on traditional on-premises, batch-oriented processing. Architectures designed to engage public cloud services for stream processing help enterprises gain new insights by using diverse data and combining data sets. By building cloud services into your architecture, you can also process data closer to where it is generated, decreasing latency and speeding response times. You also gain the agility to deploy new instances quickly without the delays associated with procuring and provisioning on-premises hardware.

Scalable Platform for Growth

No one knows what technology innovations will emerge or which business imperatives will place additional demands on your architecture. Developing a platform that can accommodate growth—of data volumes, processing power, or applications—makes sense. Build an architecture that allows you to add more instances, handle larger data loads, or dynamically change the platform to address evolving business needs. Ensure it can support new technologies such as Docker, Kubernetes, and any cloud solutions that can address your most pressing requirements.

Heterogeneous Data

Data comes from a wide and growing variety of sources. A stream processing architecture must be able to ingest multiple types of data, no matter the origin. It should also be flexible enough to match the ingestion approach with each data type.

No Need for DIY

Some business needs are so unusual that you need to build your own tools or capabilities. With stream processing, leading solutions already include nearly all of the features you need, reducing time to value. As you build an architecture for stream processing, assume you can rely on built-in capabilities to handle key tasks, such as counting and aggregation.

Foundational Capabilities

To protect the enterprise, your stream processing architecture must address the fundamental need for business continuity, security, management, and monitoring. Failover capabilities in a distributed stream processing system are critical for ensuring the system continues to run despite node failure. Built-in security is required for any environment with sensitive data that must be protected from unauthorized access. Management tools help to run the system smoothly and efficiently without unnecessary administrative overhead. And monitoring tools will help you understand the operational characteristics of your system to make adjustments when necessary to make sure you get the most efficiency from your system.

CHAPTER 5

The Hazelcast In-Memory Computing Platform

Hazelcast delivers the System of Now™, an industry-leading in-memory computing platform that provides Global 2000 enterprises, including the world's top financial institutions, with ultra-high performance for time-sensitive, cloud-native applications.

The Hazelcast In-Memory Computing Platform comprises Hazelcast IMDG, the most widely deployed in-memory data grid, and Hazelcast Jet, the industry's most advanced in-memory stream processing solution. This technology is uniquely designed to allow you to gain computing insights faster, enable actions within shorter durations, and engage new data at the speed with which it is arriving. In addition, a distributed caching architecture allows you to scale up to hundreds of terabytes and scale out for maximum efficiency when dealing with remote data or edge processing.

As the core of a stream processing architecture, Jet provides the high throughput and low latency engine that handles your more stringent streaming requirements. Its speed optimizations let you handle extreme loads. Its lightweight architecture lets it run in limited hardware environments such as in edge-optimized servers. It is designed for enterprise deployments with built-in business continuity and security capabilities. Management and monitoring is done with the Management Center user interface. It operates in conjunction with IMDG to quickly lookup related data, much faster than can be done with disk-based stores, to enrich data streams for more detailed analytics.

Built for ultra-fast processing at extreme scale, Hazelcast's cloud-native in-memory data grid and event stream processing technologies are trusted by leading companies such as JPMorgan Chase, Charter Communications, Ellie Mae, UBS, and National Australia Bank to accelerate business-critical applications. The world's largest e-commerce sites rely on Hazelcast for sub-millisecond response times to support massive volume spikes associated with Black Friday, Cyber Monday, or Singles' Day. Leading banks rely on the Hazelcast Platform to drive their fraud detection to new levels of performance, resulting in millions of dollars in fraud avoidance each year.

Intel® Optane™ DC Persistent Memory

Since stream processing often requires fast data lookups to enrich streaming data, storing the related data in random-access memory (RAM) is the ideal way of dramatically improving throughput and reducing latency. However, the one big hurdle that emerges in such a configuration is the relatively high cost of RAM. In many cases, the investment in more RAM-heavy hardware servers is justifiable, and as RAM prices continue to decrease, the use of in-memory processing in a streaming environment becomes more accessible.

Recent innovations make the adoption of in-memory processing even more practical. The Intel Optane DC Persistent Memory technology offers two ways in which in-memory processing can be more cost-effective. The first way is in volatile memory mode, in which Optane chips act as an alternative to RAM, and run at nearly the same speed but at a much lower cost and much higher capacities. This lets businesses more easily justify in-memory technologies and thus take advantage of the performance benefits that in-memory processing offers.

The second way in which Optane supports in-memory technologies is in the persistent mode. In this mode, Optane can be used as a faster alternative to solid state drives (SSDs) for restoring a system back to full operational status. For example, Hazelcast provides a hot restart capability in which in-memory data is persisted in non-volatile memory so that if a cluster goes down temporarily (or if an individual node requires offline maintenance), the in-memory contents can be restored quickly by reading data from the hot restart store. If the hot restart data is stored in Optane in persistence mode, recovery of that node can be up to 3.5x faster than using SSDs.

Instant Insight Spurs Action and Innovation

As the use of streaming processing expands, the right solution represents a critical global business opportunity for insight and innovation. Companies should look for the following core enablers:

- ✓ **Speed** at a rate that can only be delivered by in-memory systems
- ✓ **Scalability** using cloud-native solutions that allow businesses to scale up or down instantly in response to changing conditions
- ✓ **Stability** with a distributed architecture that protects against downtime and related business impacts
- ✓ **Security** so companies can protect their sensitive data from unauthorized users

When generating and processing data at the edge, the speed delivered by a modern stream processing solution helps you innovate and stay ahead of the competition. To learn how in-memory processing can optimize stream processing, visit us at www.hazelcast.com.

LEARN MORE

To understand how Hazelcast can support your broader data processing requirements, read the white paper:

Advancements in High-Speed In-Memory Systems

RTInsights.com

Accelerate Your Business with Real-Time Insights

RTInsights is an independent, expert-driven web resource for senior business and IT enterprise professionals in vertical industries. We help our readers understand how they can transform their businesses to higher-value outcomes and new business models with AI, real-time analytics, and IoT. We provide clarity and direction amid the often confusing array of approaches and vendor solutions. We provide our partners with a unique combination of services and deep domain expertise to improve their product marketing, lead generation, and thought leadership activity.

hazelcast

Hazelcast delivers the System of Now™, an in-memory computing platform that empowers Global 2000 enterprises to deliver innovative, low-latency, data-centric applications. Built for ultra-fast processing at extreme scale, Hazelcast's cloud-native in-memory data grid and event stream processing technologies are trusted by leading companies such as J.P. Morgan Chase, Charter Communications, Ellie Mae, UBS and National Australia Bank to accelerate business-critical applications. Hazelcast is headquartered in San Mateo, CA, with offices across the globe. To learn more about Hazelcast, visit https://hazelcast.com.