Hitachi Streaming Data Platform: Gain Active Insight Into Your Data Streams

The Quest: Process, Analyze and Gain Insight From Your Data In Motion

Businesses and government alike are realizing that data is becoming a perishable commodity, just like produce: If it sits idle or is stored too long, it goes to waste. Analysts have projected that by 2020 there will be more than 20 billion connected devices globally, contributing to the generation of massive amounts of data that’s now measured in the order of zetabytes\(^1\). While the volume of data being tapped and collected is staggering, for many organizations, the quest to extract real-time value from this data and translate it into meaningful and productive outcomes for both business and society continues. Streaming analytics helps companies move toward a viable solution by providing the ability to gain active insight from “data in motion,” as opposed to conventional analytics, which focuses on analyzing “data at rest,” in reference to data that is stored or archived.

\(^1\)1ZB = 10\(^{21}\)bytes
Streaming analytics provides the power to act immediately on data, and it is having far-reaching benefits across both government and a variety of industries. For example, real-time insight into live video, sensor and social media data is now enabling public safety services to respond more rapidly to emergencies. Meanwhile, energy companies are tapping into real-time usage data to efficiently balance supply and demand dynamically, resulting in reduced costs and improved service delivery. Likewise, industrial systems in transportation, manufacturing and telecommunications are now being actively adjusted based on combining streaming and stored data. This approach helps to glean predictive and prescriptive insight to improve performance, reduce costs and even prevent outages and failures.

The Evolution of Stream Processing for IoT

While streaming analytics is not new, the internet of things (IoT) places incremental demands on streaming analytics solutions that weren’t previously demanded. The sheer volume, velocity and variety of data collected from IoT requires an enterprise-class solution that is highly reliable and massively scalable. IoT is enabling a digital transformation across many industries. As a result, data is no longer just in the IT domain, but it is also fundamental to the success of all aspects of a business, ranging from operations, manufacturing, production and service, to sales and marketing. Each group can benefit from tapping into a different segment and perspective of real-time data to improve efficiency or service. This cross-domain impact and dependency requires a highly flexible, scalable and reliable streaming analytics platform.

IoT data is also collected and analyzed from a variety of human and machine sources that are now more widely dispersed than ever before. This trend will perpetuate, with ongoing advances in low-cost sensors and wireless connectivity, such as 5G, ZigBee and LoRa, which allow data sources to be even more remote and mobile. This development is resulting in an increasing need for analyzing remote data closer to the actual source since transporting massive amounts of data to centralized big data storage and analytics systems increases data backhauling, storage and compute costs. Furthermore, it can result in delays in gaining critical insights from the data and/or taking needed action before it’s too late to have a meaningful impact. An autonomous vehicle, for example, performs better if it can make real-time decisions based on active streaming sensor data correlated to live road traffic, weather and other data within the vehicle. It is much less efficient to gather, process and relay back a change in direction from a remote location while the vehicle is in motion.

The need for gaining real-time insights and making decisions as quickly as possible at the “edge of the IoT” network will continue to increase across many industries. Next-generation advanced streaming data platforms must provide the reliability, scalability and flexibility to deliver on the growing demands of IoT.

Introducing Hitachi Streaming Data Platform

Hitachi Streaming Data Platform (HSDP) is an advanced commercial, enterprise-class, high-performance software platform. HSDP continuously processes, analyzes and extracts insights from data streams anywhere data originates: on the edge, at the center or anywhere in between. HSDP is optimized for processing high volumes and velocities of data that originate from a wide variety of sources, including sensors, devices, networks and machines. With HSDP, streams of data can be continuously analyzed and correlated while the computation is performed in memory for high efficiency and rapid analysis. This powerful capability analyzes fine-grain data dynamically in a sliding window of time so that you can visually see what is happening in real time, not after it has occurred. This approach helps to improve active monitoring for critical network, IT, traffic, manufacturing or other such applications.

A smart-ingest analytic software platform, HSDP is designed to capture and analyze complex queries while correlating them to stored historical data, all in real time. The benefit of this active correlation is: Not only can you see what is happening “in the now,” but you can also project and predict what may happen based on past trends from stored data by combining machine learning. This approach is advantageous for applications like predictive monitoring and maintenance so operations and service teams can be alerted of possible threshold violations prior to the actual event, to take preemptive action. The benefits for predictive maintenance analytics alone are projected to generate close to US$25 billion by 2019, according to ABI Research2.

HSDP is a full-function stream analytics platform that provides the four primary functions as described in Figure 1: data ingestion from various sources, including IoT, network and logs; data preprocessing and transformation; data analytics and computation; and data outcome and actions.

HSDP is also a fundamental building block within Hitachi’s IoT core platform, Lumada, and complements other big data analytics solutions, including those from Pentaho, a Hitachi Group Company.

Key Advantages of HSDP

Hitachi Streaming Data Platform is an advanced, enterprise-class, streaming analytics platform with several key benefits, including:

Flexible Distributed Architecture

HSDP simplifies developing solutions that require multistage, geo-distributed analysis of data, which includes analysis on the edge and cascading the insights to a central location. HSDP software can run on a virtual machine (VM) and even on a docker container. It provides insights anywhere to support your current requirements and with the flexibility for future applications yet to be defined.

Open and Customizable

HSDP is designed for simple integration and customization into other systems and solutions. Software development kits (SDKs) are available for building input and output adapters to support a variety of data ingest sources and targets, and the HSDP developer network community is available to facilitate this. HSDP supports Java APIs and Continuous Query Language (CQL) for customization and can be added with other commercial products and open source software and technologies.

High Performance and Scalability

High performance and scalability are critical to address the needs of large-scale commercial and industrial applications, and HSDP is optimized to support both low-latency and high-throughput applications. For example, HSDP has been deployed for applications requiring millisecond-level visibility into real-time delay-sensitive transactions and can support throughput of up to 1 million tuples per second, per core, for high-performance applications.

Industry-Leading Technology

HSDP is a truly industry-leading offering when it comes to innovation, and it is unique in having over 50 patents for its use across Japan and the U.S. In fact, HSDP has more patents combined than any single vendor has for critical stream analytics technologies, including high-speed processing technology, stream data processing technology, reduced latency and improved throughput. HSDP is truly unrivaled when it comes to streaming data analytics.

HSDP Application Examples

HSDP is currently deployed in commercial and IoT solutions addressing use cases for finance, energy, manufacturing, IT operations, security, transportation and more. A few of these examples and their benefits are shown in Table 1.

HSDP Key Features

HSDP is distinguished from other streaming analytics platforms and general-purpose open source solutions because of the following key features (as identified in Table 2).

Learn More

Hitachi Streaming Data Platform can help you process, analyze and gain insights from your data streams. Visit HDS.com or contact your local Hitachi Data Systems representative to learn more.

### TABLE 1. HITACHI STREAMING DATA PLATFORM USE CASES

<table>
<thead>
<tr>
<th>Use Case</th>
<th>Industry Vertical</th>
<th>Description</th>
<th>Business Outcomes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Real-Time Power Distribution With Predictive</td>
<td>Energy</td>
<td>Monitors electricity consumption and generation in real time; actively balances the “demand” and “supply”; proactively detects abnormal signs via sensor data; prevents large-scale blackouts through proactive detection of system failure signs; improves measurement-based decision support.</td>
<td>Company achieved better operating efficiencies, improved decision-making, and better customer satisfaction.</td>
</tr>
<tr>
<td>Maintenance</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Real-Time Heating, Ventilation and Air</td>
<td>Industry</td>
<td>Achieves real-time optimization of high-energy-consuming air conditioning systems in data centers via nonintrusive and continuous (24/7/365) monitoring and visualization of key HVAC sensor data metrics.</td>
<td>Company reduced electricity consumption and improved operating efficiencies.</td>
</tr>
<tr>
<td>Conditioning (HVAC) Energy Consumption</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Optimization</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Real-Time Production Line Monitoring and</td>
<td>Manufacturing</td>
<td>Real-time manufacturing production line monitoring maintains operability of each process and traces the origins of defective products made on the line more easily.</td>
<td>Administrators were able to isolate, predict and reduce downtime in complex production line systems.</td>
</tr>
<tr>
<td>Optimization</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>High-Speed Financial Exchange Index Service</td>
<td>Finance</td>
<td>Allows major stock exchange to provide one of the world’s most advanced offerings: “High-Speed Index Service.” It enables investors to trade on detailed market movements using “best quote indexes” as indicators, with fewer tracking errors.</td>
<td>Company improved service delivery and financial gains for investors.</td>
</tr>
<tr>
<td>Real-Time Network Analytics and Optimization</td>
<td>Telecom</td>
<td>Alleviates mobile network congestion through real-time prescriptive analytics to detect congested mobile cell sites; enables optimization and feedback loop by recommending actions for traffic shaping or compression.</td>
<td>Large Tier 1 operator was able to realize 15% capital expenditure (capex) savings by adding fewer new cell sites.</td>
</tr>
</tbody>
</table>
TABLE 2. HITACHI STREAMING DATA PLATFORM FEATURES

<table>
<thead>
<tr>
<th>Feature</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>CQL (DDL and DML)</td>
<td>Continuous Query Language (CQL) is an extension of traditional SQL. CQL executes in memory and is designed for high throughput and low latency environments. It has a “windowing” concept that allows the system to treat each stream, packet and flow individually and allows for “stateful” analysis unlike open source technologies where this capability has to be custom coded. It’s easy to use and adopt programing interface based on SQL. It also includes built-in support for temporal analysis, statistical, mathematical, string and data conversion.</td>
</tr>
<tr>
<td>Software Development Kit (SDK) to Build Custom Extensions and Adapters</td>
<td>SDK simplifies the ability for plugging in custom processing and analytics written in Java or C and building input and output adapters. It also simplifies ingesting data from a variety of sources and publishing insights to a variety of targets.</td>
</tr>
<tr>
<td>Developer Customization</td>
<td>Developing custom processing and analytics using extensions in JAVA, HSDP provides APIs for plugging in custom extensions. These APIs allow developers to integrate their machine learning algorithms that were developed in R or Python. The API can be either in C or Java, and invoked by CQL on the data stream.</td>
</tr>
<tr>
<td>In-Memory Processing</td>
<td>Data is processed while it is still in memory, thus eliminating unnecessary disk access. By avoiding excess disk I/O, HSDP enables data to be processed faster.</td>
</tr>
<tr>
<td>High-Availability Architecture</td>
<td>The architecture includes multiple levels of fault-tolerant support for achieving desired levels of high availability, including support for auto-restart.</td>
</tr>
<tr>
<td>Combined Real-Time and Historical Insights</td>
<td>Supports capability for combining past insights and reference data with live streaming insights. This is useful for use cases such as anomaly detection and data enrichment.</td>
</tr>
<tr>
<td>Incremental Computational Processing</td>
<td>A preloaded query is processed iteratively when triggered by the input data, and the processing results are available for the next iteration. This approach improves efficiency as HSDP processes only those elements that have changed.</td>
</tr>
<tr>
<td>Temporal Analysis</td>
<td>Time is an important aspect for analysis with HSDP. HSDP natively supports temporal analysis of data. The data tuples can be analyzed based on arrival time or created time.</td>
</tr>
<tr>
<td>Scale-Up and Scale-Out Options</td>
<td>In a scale-up scenario, query groups can analyze in parallel multiple threads on the same HSDP instance. In a scale-out scenario, query groups can analyze in parallel on multiple HSDP instances.</td>
</tr>
</tbody>
</table>