Confluent Unveils New Capabilities to Apache Flink Offering to Simplify AI and Bring Stream Processing to Workloads Everywhere

May 2, 2024

Confluent's new AI Model Inference seamlessly integrates AI and ML capabilities into data pipelines

Confluent Platform for Apache Flink® enables stream processing for on-prem and private cloud workloads with support from Confluent's leading Flink experts

Confluent's new Freight clusters offer cost savings for high-throughput use cases with relaxed latency requirements

BENGALURU, India--(BUSINESS WIRE)--May 2, 2024-- Confluent, Inc. (NASDAQ:CFLT), the data streaming pioneer, announced AI Model Inference, an upcoming feature on Confluent Cloud for Apache Flink®, to enable teams to easily incorporate machine learning into data pipelines. Confluent introduced Confluent Platform for Apache Flink®, a Flink distribution that enables stream processing in on-premises or hybrid environments with support from the company's Flink experts. Confluent also unveiled Freight clusters, a new cluster type for Confluent Cloud that provides a cost-effective way to handle large-volume use cases that aren't time-sensitive, such as logging or telemetry data.

AI Model Inference simplifies building and launching AI and ML applications

Generative AI helps organizations innovate faster and deliver more tailored customer experiences. AI workloads need fresh, context-rich data to ensure the underlying models generate accurate output and results for businesses to make informed decisions based on the most current information available.

However, developers often have to use several tools and languages to work with AI models and data processing pipelines, leading to complex and fragmented workloads. This can make it challenging to leverage the most current and relevant data for decision-making, leading to errors or inconsistencies and compromising the accuracy and reliability of AI-driven insights. These issues can cause increased development time and difficulties in maintaining and scaling AI applications.

With AI Model Inference in Confluent Cloud for Apache Flink®, organizations can use simple SQL statements from within Apache Flink to make calls to AI engines, including OpenAI, Amazon SageMaker, GCP Vertex, and Microsoft Azure. Now enterprises can orchestrate data cleaning and processing tasks on a single platform.

"Apache Kafka and Flink are the critical links to fuel machine learning and artificial intelligence applications with the most timely and accurate data," said Shaun Clowes, Chief Product Officer at Confluent. "Confluent's AI Model Inference removes the complexity involved when using streaming data for AI development by enabling organizations to innovate faster and deliver powerful customer experiences."

AI Model Inference enables companies to:

- **Simplify AI development by using familiar SQL syntax to work directly with AI/ML models**, reducing the need for specialized tools and languages.
- **Establish seamless coordination between data processing and AI workflows** to improve efficiency and reduce operational complexity.
- **Enable accurate, real-time AI-driven decision-making** by leveraging fresh, contextual streaming data.

"Leveraging fresh, contextual data is paramount for training and refining AI models, and for use at time of inference to improve the accuracy and relevancy of outcomes," said Stewart Bond, Vice President, Data Intelligence and Integration Software, IDC. "Organizations need to improve efficiencies of AI processing by unifying data integration and processing pipelines with AI models. Flink can now treat foundational models as first class resources, enabling the unification of real-time data processing with AI tasks to streamline workflows, enhance efficiency, and reduce operational complexity. These capabilities empower organizations to make accurate, real-time AI-driven decisions based on the most current and relevant streaming data while enhancing performance and value."

Support for AI Model Inference is currently available in early access to select customers. Customers can sign up for early access to learn more about this offering.

Confluent Platform for Apache Flink® enables stream processing in private clouds and on-premises environments

Many organizations are looking for hybrid solutions to protect more sensitive workloads. With Confluent Platform for Apache Flink®, a Flink distribution fully supported by Confluent, customers can easily leverage stream processing for on-prem or private cloud workloads with long-term expert support. Apache Flink can be used in tandem with Confluent Platform with minimal changes to existing Flink jobs and architecture.

Confluent Platform for Apache Flink® can help organizations:

- **Minimize risk with unified Flink and Kafka support and expert guidance** from the foremost experts in the data streaming industry.
- **Receive timely assistance in troubleshooting and resolving issues**, reducing the impact of any operational disruptions to mission critical applications.
- **Ensure that stream processing applications are secure and up to date** with off-cycle bug and vulnerability fixes.

With Kafka and Flink available in Confluent's complete data streaming platform, organizations can ensure better integration and compatibility between technologies, and receive comprehensive support for streaming workloads across all environments. Unlike open source Apache Flink which only maintains the two most recent releases, Confluent offers three years of support for every Confluent Platform for Apache Flink® release from launch, guaranteeing uninterrupted operations and peace of mind.

Confluent Platform for Apache Flink® will be available to Confluent customers later this year.

New auto-scaling Freight clusters offer more cost-efficiency at scale

Many organizations use Confluent Cloud to process logging and telemetry data. These use cases involve large amounts of business-critical data but are often less latency-sensitive since they typically feed into indexing or batch aggregation engines. To make these common use cases more cost-efficient for customers, Confluent is introducing Freight clusters —a new serverless cluster type with up to 90% lower cost for high-throughput use cases with relaxed latency requirements. Powered by Elastic CKUs, Freight clusters seamlessly auto-scale based on demand with no manual sizing or capacity planning required, enabling organizations to minimize operational overhead and optimize costs by paying only for the resources that they use when they need it.

Freight clusters are available in early access in select AWS regions. Customers can sign up for early access to learn more about this offering.

Additional resources

- Start your free trial of Confluent Cloud — No credit card required
- Read here for more information on data streaming for real-time AI
- Learn more about Confluent's Apache Flink offerings

As our roadmap may change in the future, the features referred to herein may change, may not be delivered on time, or may not be delivered at all. This information is not a commitment to deliver any functionality and customers should make their purchasing decisions based upon features that are currently available.

About Confluent

Confluent is the data streaming platform that is pioneering a fundamentally new category of data infrastructure that sets data in motion. Confluent's cloud-native offering is the foundational platform for data in motion—designed to be the intelligent connective tissue enabling real-time data, from multiple sources, to constantly stream across the organization. With Confluent, organizations can meet the new business imperative of delivering rich, digital front-end customer experiences and transitioning to sophisticated, real-time, software-driven back-end operations. To learn more, please visit www.confluent.io.

View source version on businesswire.com: https://www.businesswire.com/news/home/20240501428486/en/

Natalie Mangan
pr@confluent.io

Source: Confluent, Inc.