Stream Processing with Apache Flink

Fundamentals, Implementation, and Operation of Streaming Applications

Fabian Hueske & Vasiliki Kalavri

Get started with Apache Flink, the open source framework that powers some of the world's largest stream processing applications. With this practical book, you'll explore the fundamental concepts of parallel stream processing and discover how this technology differs from traditional batch data processing.

Longtime Apache Flink committers Fabian Hueske and Vasia Kalavri show you how to implement scalable streaming applications with Flink’s DataStream API and continuously run and maintain these applications in operational environments. Stream processing is ideal for many use cases, including low-latency ETL, streaming analytics, and real-time dashboards as well as fraud detection, anomaly detection, and alerting. You can process continuous data of any kind, including user interactions, financial transactions, and IoT data, as soon as you generate them.

* Learn concepts and challenges of distributed stateful stream processing
* Explore Flink’s system architecture, including its event-time processing mode and fault-tolerance model
* Understand the fundamentals and building blocks of the DataStream API, including its time-based and stateful operators
* Read data from and write data to external systems with exactly-once consistency
* Deploy and configure Flink clusters
* Operate continuously running streaming applications

“Stream Processing with Apache Flink is a great book for everyone from old-timers in the streaming world to beginner software and data engineers writing their first stream processing jobs. As the book reviews Flink, it also teaches core streaming fundamentals that with help readers level up their technical thought process. Total recommended read.”

* Ted Malaska

Director of Enterprise Architecture at Capital One

Fabian Hueske is a PMC member of the Apache Flink project and has been contributing to Flink since day one. Fabian is cofounder of data Artisans (now Ververica) and holds a PhD in computer science from TU Berlin.

Vasiliki (Vasia) Kalavri is a postdoctoral fellow in the Systems Group at ETH Zurich. Vasia is a PMC member of the Apache Flink project, an early contributor to Flink. She has worked on its graph processing library, Gelly, and on early versions of the Table API and streaming SQL.