Apache Flink™: Stream and Batch processing at Scale

Marton Balassi   (ELTE/SZTAKI, Hungary)

Paris Carbone   (KTH, Stockholm, Sweden)

Gyula For a   (SICS, Stockholm, Sweden)

Vasia Kalavri   (KTH, Stockholm, Sweden)

Asterios Katsifodimos (TU Berlin, Germany)

What is Flink?

Applications

Hive | Cascading | Giraph

Mahout | Pig | Crunch

Data processing engines

MapReduce | Flink

Spark | Storm | Tez

App and resource management
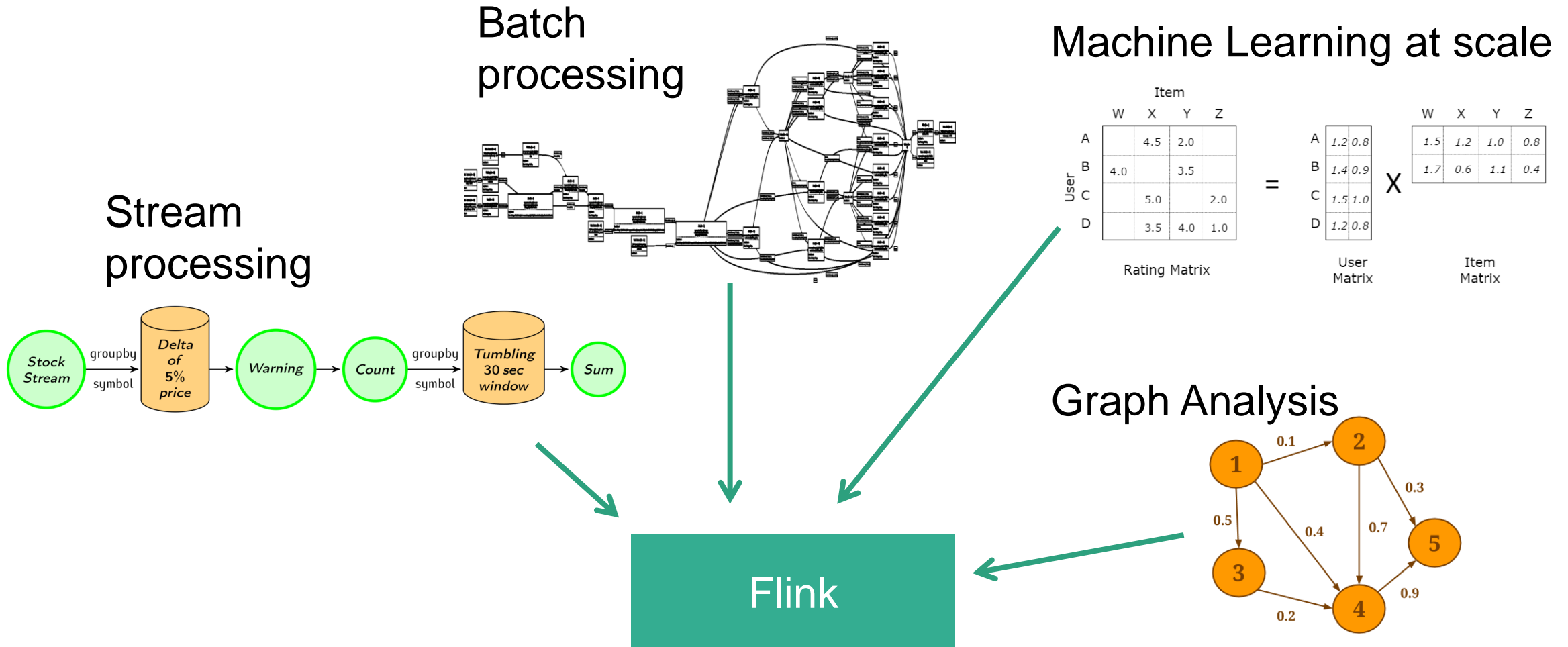
Yarn | Mesos

Storage, streams

HDFS | HBase | Kafka | ...

What can I do with Flink?

Batch processing

Machine Learning at scale

Stream processing

Graph Analysis

Flink

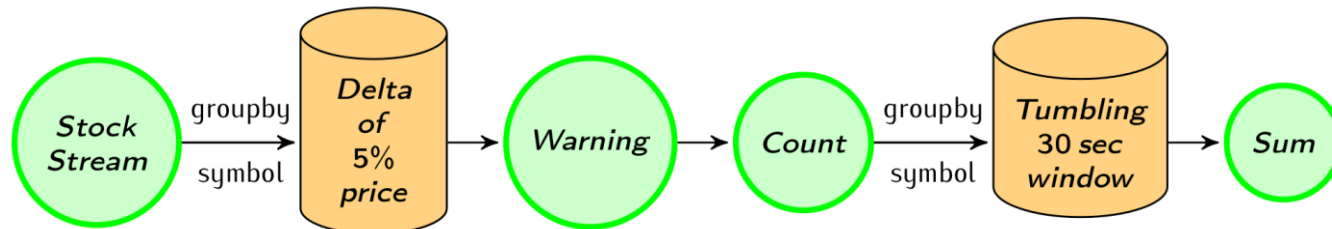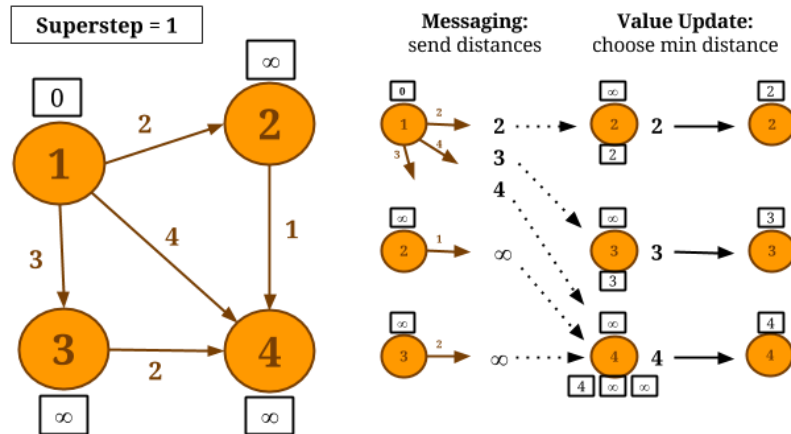An engine that can **natively** support all these workloads.

But what will I do with Flink today?

- **Graph processing**
  - ETL on Datasets
  - Graph creation & analysis

- **Stream Processing**
  - Rolling Aggregates
  - Windows & Alerts

Agenda

- **Introduction**
  - 15' Overview
  - 15' Gelly (Graph) API
- **30' Break**
- **Graph Processing**
  - 20' DataSet/Gelly Hands-on
- **Stream processing with Flink**
  - 10' DataStream API
  - 15' Fault Tolerance Demo
  - 45' Streaming Hands-on