

Apache Flink 简介

Apache Flink 是一个框架和分布式处理引擎，它同时支持了批处理和流处理，也能用来做一些基于事件的应用。Flink 可以搭建廉价机群，快速处理任意规模的数据。使用官网的一句话来介绍 Flink 就是 **"Stateful Computations Over Streams"**。

Flink 最适合的应用场景是低时延的数据处理（Data Processing）场景：高并发 pipeline 处理数据，时延毫秒级，且兼具可靠性。

Flink 是一个纯流式的计算引擎，它的基本数据模型是数据流。流可以是无边界的无限流，即一般意义上的流处理。也可以是有边界的有限流，这样就是批处理。因此 Flink 用一套架构同时支持了流处理和批处理。Flink 的一个优势是支持**持有状态的计算**。如果处理一个事件（或一条数据）的结果只跟事件本身的内容有关，称为无状态处理；反之结果**还和之前处理过的事件有关**，称为**有状态处理**。稍微复杂一点的数据处理，比如说基本的聚合，数据流之间的关联都是有状态处理。

Flink 的主要特点如下：

Flink 对于提供的 ProcessFunction API 可以非常细粒度灵活的控制 Event 处理。

Flink 提供了对于 SQL 的支持，对 Event 进行数据聚合后，可以使用 SQL 进行数据分析。

Flink 是来一个 Event 处理一次，对比按照时间间隔微批次处理，所以保证了数据实时处理。

Flink 对计算状态持久化提供了非常多的特性。