

一文了解 Apache Flink 核心技术

作者：伍翀（云邪）

整理：李泽聚

Apache Flink 介绍

Apache Flink（以下简称Flink）是近年来越来越流行的一款开源大数据计算引擎，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。使用官网的语句来介绍 Flink 就是“Stateful Computations Over Streams”。

首先 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。其次，Flink 的一个优势是支持有状态的计算。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。

<https://ververica.cn/developers/apache-flink-core-technology/>