

Apache Flink 管理大型状态之增量 Checkpoint 详解

作者 | Stefan Richter & Chris Ward

翻译 | 邱从贤（山智）

via | <https://flink.apache.org/features/2018/01/30/incremental-checkpointing.html>

Apache Flink 是一个有状态的流计算框架，状态是作业算子中已经处理过的内存状态，供后续处理时使用。状态在流计算很多复杂场景中非常重要，比如：

- 保存所有历史记录，用来寻找某种记录模式
- 保存最近一分钟的所有记录，用于对每分钟的记录进行聚合统计
- 保存当前的模型参数，用于进行模型训练

有状态的流计算框架必须有很好的容错性，才能在生产环境中发挥用处。这里的容错性是指，不管是发生硬件故障，还是程序异常，最终的结果不丢也不重。

<https://ververica.cn/developers/manage-large-state-incremental-checkpoint/>