



Flink介绍

部门：数据智能部
日期：2021-12-9
分享人：杨成凯

A teal geometric shape, resembling a stylized 'L' or a corner, is located in the bottom-left corner of the slide.

CONTENTS

目 录

- 01 目前flink作业一览
- 02 flink相关概念
- 03 flink文档走读
- 04 flink作业的源码走读

01 目前flink作业一览



- <http://115.159.51.208:2081/#/job/running>



"Don't try to understand it. Feel it."



02 flink相关概念



- 自身架构
 - 有界流和无界流
 - 运行环境
 - 后端存储
- 实时应用角度来看
 - 有状态
 - 时间语义
 - 三层API
- 从运维角度层面来看
 - 检查点
 - 最多一次 vs 至多一次
 - 多种监控方式
- <https://flink.apache.org/flink-architecture.html>

03 Flink文档走读



Flink的概括总结

- Apache Flink is a framework and distributed processing engine for stateful computations over unbounded and bounded data streams.



- QuickStart
- Application Development
- Connectors
- Operations
 - CheckPoints
 - State Backends
 - Task Failure Recovery
- Concepts
 - Stateful Stream Processing
 - Timely Stream Processing

04 Flink作业的代码走读



目前标准的Flink作业代码模板

核心API

- source: kafka/mq/nifi/自定义source测试 etc
- 无状态算子 map/filter
- 有状态算子 window开时间窗后
- sink: kafka sink/mysql sink/FileSystem sink/自定义sink

新代帐指标值实时接入



目前标准的Flink作业代码模板

TableAPI

- source: kafka
- FlinkSQL+hive维度表
- sink: kafka/hive

新代帐日志追踪

nginx日志清洗

反爬行为分析



目前标准的Flink作业代码模板

底层API

- source: kafka
- KeyedProcessFunction: State描述符
- sink: kafka/hive

nginx日志清洗