# 第一章Flink基础介绍

## 1.1简介

图形用户界面, 应用程序

描述已自动生成

图1-1 Flink框架处理流程

Flink 的具体定位是:Apache Flink 是一个框架和分布式处理引擎，如图 1-1 所示，用于对无界和有界数据流进行有状态计算。Flink 被设计在所有常见的集群环境中运行，以内存执行 速度和任意规模来执行计算

## 1.2 批处理和流处理

批处理：对有界数据流的处理通常被称为批处理。批处理不需要有序地获取数据。在批处理模式下，首先将数据流持久化到存储系统（文件系统或对象存储）中，然后对整个数据集的数据进行读取、排序、统计或汇总计算，最后输出结果。

实时流处理：对于无界数据流，通常在数据生成时进行实时处理。因为无界数据流的数据输入是无限的，所以必须持续地处理。数据被获取后需要立刻处理，不可能等到所有数据都到达后再进行处理。处理无界数据流通常要求以特定顺序（如事件发生的顺序）获取事件，以便能够保证推断结果的完整性。

# 第二章Flink环境部署

## 2.0 WordCount样例

1. 创建项目

图形用户界面, 文本, 应用程序

描述已自动生成图形用户界面, 文本, 应用程序, 电子邮件

描述已自动生成

2. 添加项目依赖

表格

中度可信度描述已自动生成

3.编写代码

WordCount: 单词频次统计的基本思路是:先逐行读入文件数据，然后将每一行文字拆分成单 词;接着按照单词分组，统计每组数据的个数，就是对应单词的频次。

## 2.1 部署模式

## 2.2 独立模式

## 2.3 Yarn模式

# 第三章Flink运行时架构

## 3.1系统架构

## 3.2 作业提交流程

## 3.3 重要概念

# 第四章Flink时间和窗口

# 第五章Flink DataStream API

# 第六章Flink 处理函数

# 第七章Flink 状态机制（State）

# 第八章Flink 容错机制(checkpoint)

# 第九章Fink 窗口机制（windows）

# 第十章Flink Table API与SQL

# 第十一章Flink实时数仓搭建规划

## 基础架构

## 整库同步

## 实时作业的监控

## 元数据管理

## 表结构变更

## Hive SQL平迁Flink SQL