

Chapter 11 Where to Go from Here? 第11章 发展方向

前言:

这是一段漫长的旅程,您已经完成了本书的结尾!但是您的Flink旅程才刚刚开始,本章指出您可以从此处采取的可能途径。我们将为您简要介绍本书中未包含的其他Flink功能,并为您提供一些指向其他Flink资源的指导。在Flink周围有一个充满活力的社区,我们鼓励您与其他用户建立联系,开始贡献力量,或者找出公司与Flink建立的合作关系,以帮助激发您自己的工作

The Rest of the Flink Ecosystem flink生态的其余部分

前言:

尽管本书特别侧重于流处理,但Flink实际上是一个通用的分布式数据处理框架,也可以用于其他类型的数据分析。此外,Flink为关系查询,复杂事件处理(CEP)和图处理提供了指定lib和api

A Welcoming Community

Apache Flink拥有一个不断发展的欢迎社区,其贡献者和用户遍布全球。这里有一些资源,您可以用来提问,参加Flink相关活动以及了解人们使用Flink的目的

Mailing lists

user@flink.apache.org : user support and questions

dev@flink.apache.org : development, release, and community discussions

community@flink.apache.org : community news and meetups

blog

<https://flink.apache.org/blog>

<https://www.ververica.com/blog>

Meetups and conferences

<https://flink-forward.org>

<https://www.meetup.com/topics/apache-flink>

The DataSet API for Batch Processing

1. dataset介绍

DataSet程序被指定为一系列转换,就像DataStream程序一样,区别在于DataSet是有界数据集,

a. 提供了常规的filtering, mapping, selection, joins, and grouping等以及读写外部系统的connectors

b. 也可以使用dataset迭代程序直到满足收敛

c. 截止到flink1.7 batch job在作业内部表现为dataflow programs和stream job运行在相同的runtime,社区正在努力将batch和stream合并

Table API and SQL for Relational Analysis 表api和sql分析

1. Table API介绍

即使底层的DataStream API和DataSet API是分开的,但是可以使用更上层的api实现流批的统一,Table API是用于Scala和Java的语言集成查询(LINQ) API,可以执行查询以进行批处理或流分析,提供了常见的运算符来编写包括selection, projection, aggregations, and joins的关系查询,并且还具有对自动完成和语法验证的IDE支持。

2. sqlapi及sql cli的使用

a. flink sql遵循ANSI SQL标准,并利用Apache Calcite进行查询解析和优化

b. 可以将sql查询嵌入到dataset或者datastream api中,或者直接使用SQL CLI客户端向flink cluster提交查询

c. 此外,您可以使用CLI客户端提交分离的查询,这些查询将其结果直接写入外部存储系统

FlinkCEP for Complex Event Processing and Pattern Matching --flink cep处理复杂事件和模式匹配

1. FlinkCEP是用于复杂事件模式检测的高级API和库。它在DataStream API的顶部实现,并允许您指定要在流中检测的模式。常见的CEP用例包括财务应用程序,欺诈检测,复杂系统中的监视和警报以及检测网络入侵或可疑用户行为

Gelly for Graph Processing --图处理之gelly

Gelly是Flink的图形处理API和库。它建立在DataSet API和Flink对高效批处理迭代的支持之上。Gelly在Java和Scala中都提供了高级编程抽象,以执行图转换,聚合和迭代处理(例如以顶点为中心和应用汇总和)。它还包括一组可供使用的常见图形算法

注意:

Flink的高级API和接口相互之间以及与DataStream和DataSet API的良好集成,可以轻松地将它们混合并在同一程序中的库和API之间进行切换。例如,您可以使用CEP库从DataStream中提取模式,然后使用SQL分析提取的模式,或者可以使用Table API将表过滤并投影到图形中,然后再使用Gelly库中的图形算法对其进行分析