

知 Sqoop1和Sqoop2架构有什么不同?

厉梦如 2019-05-19 发表

问题描述

Q: Sqoop1和Sqoop2架构有什么不同?

解决方法

A:

I Sqoop1只是提交一个Map作业(负责数据的传输和转化)实现Hadoop (HDFS/HBase/Hive) 与关系数据库进行数据的导入导出, 只有一个简单的client, 通过命令行进行交互;

I Sqoop2是一个完整的mapreduce过程, (Sqoop2的Map作业负责数据源的传输, 而reducers负责数据源的转化)。引入了Sqoop server (具体服务器为tomcat), 对connector实现了集中的管理。Connector将控制Sqoop2中整个数据移动过程。每一个特定的数据存储可能有自己特定的Connector (一个Oracle专用, 一个MySQL专用)。一些数据存储可能有多个Connector来完成不同的目的。Metadata中包含数据的信息如数据的所在地, 大小等等。Sqoop2客户端可以是命令行或者浏览器方式交互。