

大数据技术-第九章：Sqoop数据迁移
Sqoop相关知识



CONTENTS

01. Sqoop起源

02. Sqoop定义

03. Sqoop优势



01

Sqoop起源



多数使用Hadoop技术处理大数据业务的企业，有大量的数据存储在关系型数据中。由于没有工具支持，Hadoop和关系型数据库之间的数据传输是很困难的事情。传统的应用程序管理系统，即应用程序与使用RDBMS的关系数据库的交互，是产生大数据的来源之一。由RDBMS生成的这种大数据存储在关系数据库结构中的关系数据库服务器中。

当大数据存储和Hadoop生态系统的MapReduce，Hive，HBase等分析器出现时，他们需要一种工具来与关系数据库服务器进行交互，以导入和导出驻留在其中的大数据。Sqoop在Hadoop生态系统中占据一席之地，为关系数据库服务器和Hadoop的HDFS之间提供了可行的交互。

02

Sqoop定义



Sqoop是Apache旗下一款“Hadoop和关系数据库服务器之间传送数据”的工具。主要用于在Hadoop(Hive)与传统的数据库(MySQL、Oracle、Postgres等)之间进行数据的传递，可以将一个关系型数据库中的数据导进到Hadoop的HDFS中，也可以将HDFS的数据导进到关系型数据库中。它是连接传统关系型数据库和 Hadoop 的桥梁，把关系型数据库的数据导入到Hadoop 系统, 如HDFS/HBase/Hive；把数据从Hadoop 系统里抽取并导出到关系型数据库里。利用MapReduce加快数据传输速度。批处理方式进行数据传输。

03

Sqoop优势



- 高效、可控地利用资源
 - 任务并行度，超时时间等
- 数据类型映射与转换
 - 可自动进行，用户也可自定义
- 支持多种数据库
 - MySQL
 - Oracle
 - PostgreSQL

Turing AI 万维图灵 | 大数据系列课程

大数据

BIG
DATA

智 / 能 / 科 / 技 放 / 眼 / 未 / 来

