## Sqoop简介

SQL to Hadoop And Hadoop to SQL

Sqoop是Hadoop和关系数据库服务器之间传送数据的一种工具。它是用来从关系数据库如：MYSQL Oracle到Hadoop的HDFS，并从Hadoop的文件系统导出数据到关系数据库。

传统的应用管理系统，也就是与关系型数据库的使用RDBMS应用程序的交互，是产生大数据的来源之一。这样大的数据，由关系数据库生成的，存储在关系数据库结构关系数据库服务器。

当大数据存储器和分析器，如MapReduce, Hive, HBase, Cassandra, Pig等，Hadoop的生态系统等应运而生图片，它们需要一个工具来用的导入和导出的大数据驻留在其中的关系型数据库服务器进行交互。在这里，Sqoop占据着Hadoop生态系统提供关系数据库服务器和Hadoop HDFS之间的可行的互动。

Sqoop：“SQL 到 Hadoop 和 Hadoop 到SQL”

Sqoop是Hadoop和关系数据库服务器之间传送数据的一种工具。它是用来从关系数据库如MySQL，Oracle到Hadoop的HDFS从Hadoop文件系统导出数据到关系数据库。它是由Apache软件基金会提供。

## Sqoop版本

文档使用的版本：sqoop-1.99.7-bin-hadoop200、hdoop-2.7.33

## Sqoop是怎样工作的？

|  |
| --- |
| http://www.yiibai.com/uploads/allimg/201501/1-15011R22R32N.jpg |

## 传统模式与Sqoop模式区别

|  |
| --- |
|  |

## Sqoop导入

导入工具从RDBMS到HDFS导入单个表。表中的每一行被视为HDFS的记录。所有记录被存储在文本文件的文本数据或者在Avro和序列文件的二进制数据。

## Sqoop导出

导出工具从HDFS导出一组文件到一个RDBMS。作为输入到Sqoop文件包含记录，这被称为在表中的行。那些被读取并解析成一组记录和分隔使用用户指定的分隔符。