

# Sqoop理论与代码文件

## Sqoop概述

功能

主要用于在Hadoop(Hive)与传统的数据库(mysql、postgresql...)间进行数据的传递,可以将一个关系型数据库(例如:MySQL,Oracle,Postgres等)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。

说明

Sqoop1与2的区别

Sqoop1.x

Sqoop2.x

图示

原理

将导入或导出命令翻译成mapreduce程序来实现

在翻译出的mapreduce中主要是对inputformat和outputformat进行定制

## Sqoop使用案例

导入数据

RDBMS到HDFS

在mysql中建表

导入数据

查询导入

使用关键字查询导入数据

RDBMS到Hive

RDBMS到HBase

导出数据

脚本打包

mkdir opt

vim opt/job\_HDFS2RDBMS.opt

脚本使用

bin/sqoop --options-file opt/job\_HDFS2RDBMS.opt