tutorialspoint

SIMPLYEASYLEARNING

www.tutorialspoint.com

https://www.facebook.com/tutorialspointindia

https://twitter.com/tutorialspoint

About the Tutorial

Sqoop is a tool designed to transfer data between Hadoop and relational database servers. It is used to import data from relational databases such as MySQL, Oracle to Hadoop HDFS, and export from Hadoop file system to relational databases.

This is a brief tutorial that explains how to make use of Sqoop in Hadoop ecosystem.

Audience

This tutorial is prepared for professionals aspiring to make a career in Big Data Analytics using Hadoop Framework with Sqoop. ETL developers and professionals who are into analytics in general may as well use this tutorial to good effect.

Prerequisites

Before proceeding with this tutorial, you need a basic knowledge of Core Java, Database concepts of SQL, Hadoop File system, and any of Linux operating system flavors.

Copyright & Disclaimer

© Copyright 2015 by Tutorials Point (I) Pvt. Ltd.

All the content and graphics published in this e-book are the property of Tutorials Point (I) Pvt. Ltd. The user of this e-book is prohibited to reuse, retain, copy, distribute or republish any contents or a part of contents of this e-book in any manner without written consent of the publisher.

We strive to update the contents of our website and tutorials as timely and as precisely as possible, however, the contents may contain inaccuracies or errors. Tutorials Point (I) Pvt. Ltd. provides no guarantee regarding the accuracy, timeliness or completeness of our website or its contents including this tutorial. If you discover any errors on our website or in this tutorial, please notify us at contact@tutorialspoint.com

Table of Contents

About the Tutorial i

Audience i

Prerequisites i

Copyright & Disclaimer i

Table of Contents ii

1. SQOOP – INTRODUCTION 1

How Sqoop Works? 1

Sqoop Import 2

Sqoop Export 2

2. SQOOP – INSTALLATION 3

Step 1: Verifying JAVA Installation 3

Step 2: Verifying Hadoop Installation 4

Step 3: Downloading Sqoop 11

Step 4: Installing Sqoop 11

Step 5: Configuring bashrc 11

Step 6: Configuring Sqoop 11

Step 7: Download and Configure mysql-connector-java 12

Step 8: Verifying Sqoop 12

3. SQOOP – IMPORT 13

Syntax 13

Importing a Table 14

Importing into Target Directory 16

Import Subset of Table Data 16

Incremental Import 17

tutorialspoint
SIMPLYEASYLEARNING

4. SQOOP – IMPORT-ALL-TABLES 19

Syntax **19**

5. SQOOP – EXPORT 21

Syntax **21**

6. SQOOP – JOB 23

Syntax **23**

Create Job (--create) **23**

Verify Job (--list) **23**

Inspect Job (--show) **24**

Execute Job (--exec) **24**

7. SQOOP – CODEGEN 25

Syntax **25**

8. SQOOP – EVAL 27

Syntax **27**

Select Query Evaluation **27**

Insert Query Evaluation **27**

9. SQOOP – LIST-DATABASES 29

Syntax **29**

Sample Query **29**

10. SQOOP – LIST-TABLES 30

Syntax **30**

Sample Query **30**

tutorialspoint
SIMPLYEASYLEARNING

1. SQOOP – INTRODUCTION

The traditional application management system, that is, the interaction of applications with relational database using RDBMS, is one of the sources that generate Big Data. Such Big Data, generated by RDBMS, is stored in **Relational Database Servers** in the relational database structure.

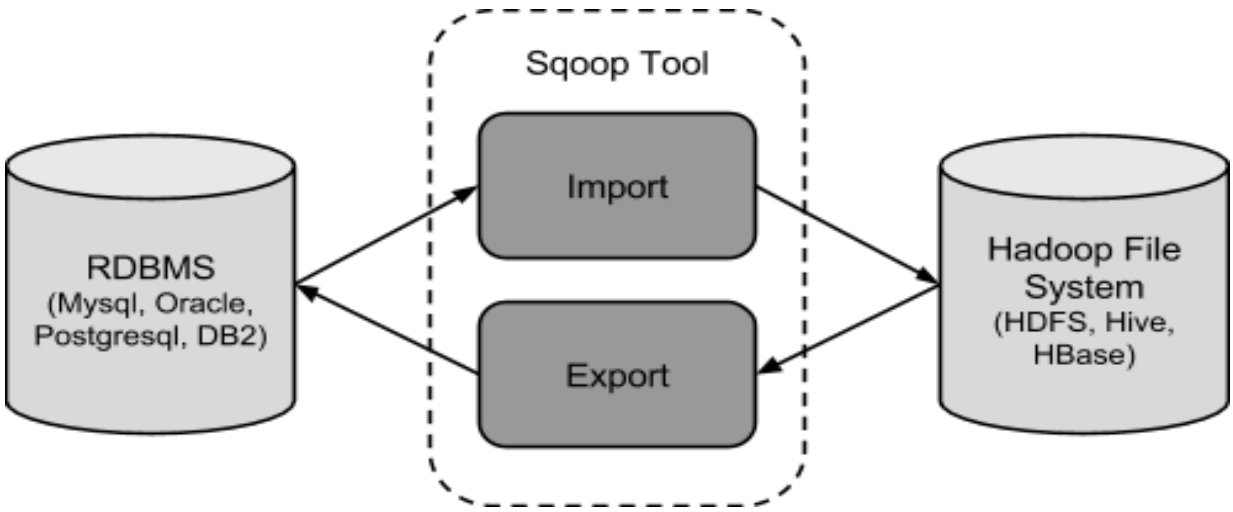
When Big Data storages and analyzers such as MapReduce, Hive, HBase, Cassandra, Pig, etc. of the Hadoop ecosystem came into picture, they required a tool to interact with the relational database servers for importing and exporting the Big Data residing in them. Here, Sqoop occupies a place in the Hadoop ecosystem to provide feasible interaction between relational database server and Hadoop's HDFS.

Sqoop: "SQL to Hadoop and Hadoop to SQL"

Sqoop is a tool designed to transfer data between Hadoop and relational database servers. It is used to import data from relational databases such as MySQL, Oracle to Hadoop HDFS, and export from Hadoop file system to relational databases. It is provided by the Apache Software Foundation.

How Sqoop Works?

The following image describes the workflow of Sqoop.



tutorialspoint
SIMPLYEASYLEARNING

Sqoop Import

The import tool imports individual tables from RDBMS to HDFS. Each row in a table is treated as a record in HDFS. All records are stored as text data in text files or as binary data in Avro and Sequence files.

Sqoop Export

The export tool exports a set of files from HDFS back to an RDBMS. The files given as input to Sqoop contain records, which are called as rows in table. Those are read and parsed into a set of records and delimited with user-specified delimiter.

tutorialspoint
SIMPLYEASYLEARNING

End of ebook preview

If you liked what you saw…

Buy it from our store @ **https://store.tutorialspoint.com**

tutorialspoint
SIMPLYEASYLEARNING