

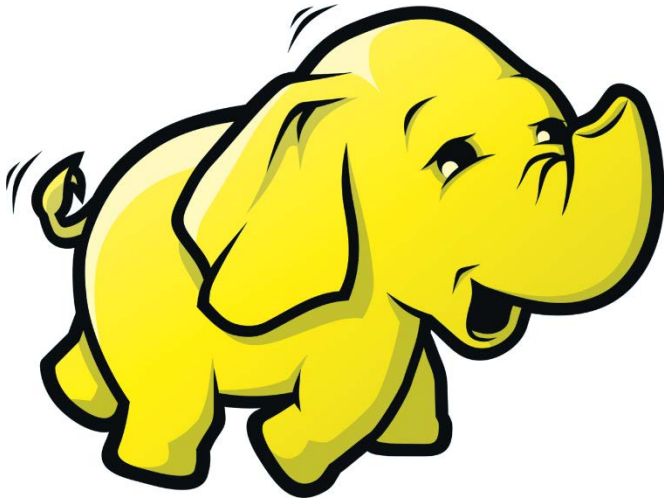
# Transferring Relational Data to HDFS Using Sqoop



Thomas M. Henson

@henson\_tm | [www.thomashenson.com](http://www.thomashenson.com)

# Overview



What is Sqoop?

Benefits of Sqoop

Quick peek at documentation

Demo transferring data from MySQL

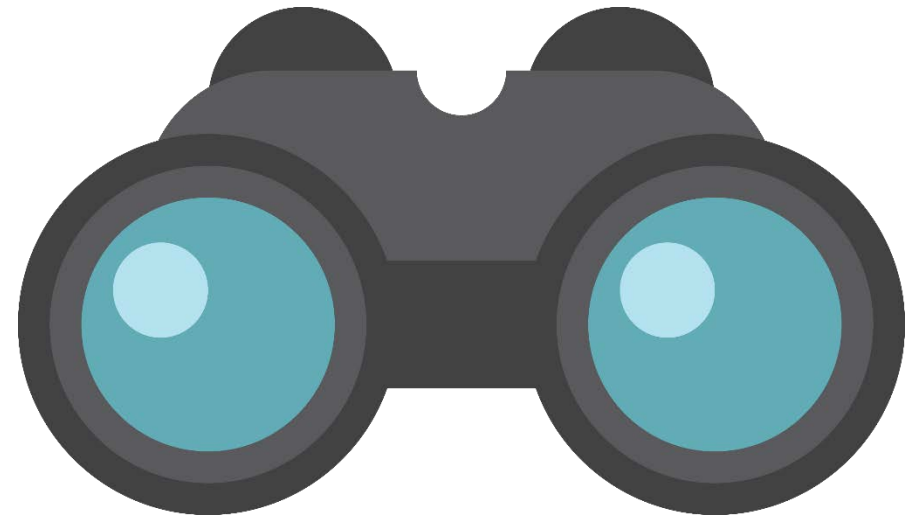
Find out where you can learn more

# Sqoop

Tool developed to do bulk transfer of data between relational databases and Hadoop.

# How Sqoop Works

- Written in Java
- Relational database
  - MySQL
  - Oracle
  - SQL Server
  - PostgreSQL
- Third party NoSQL

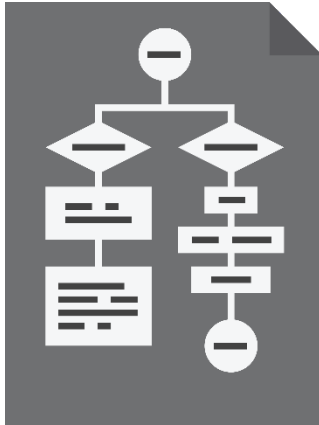


# Benefits of Sqoop

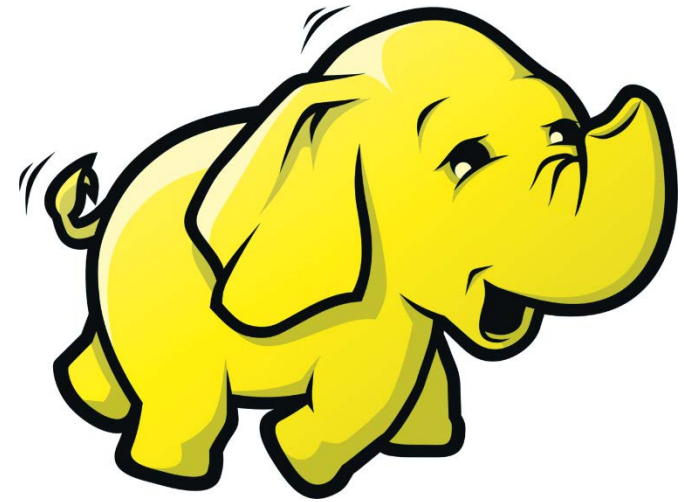
Open Source

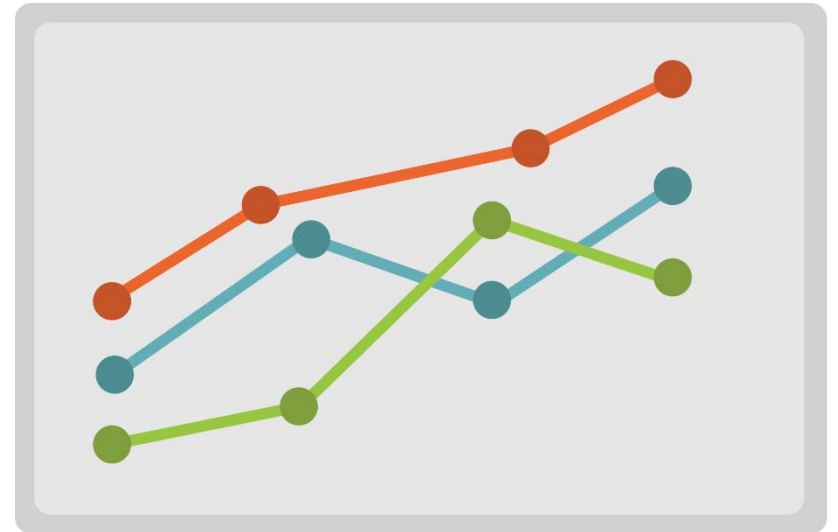
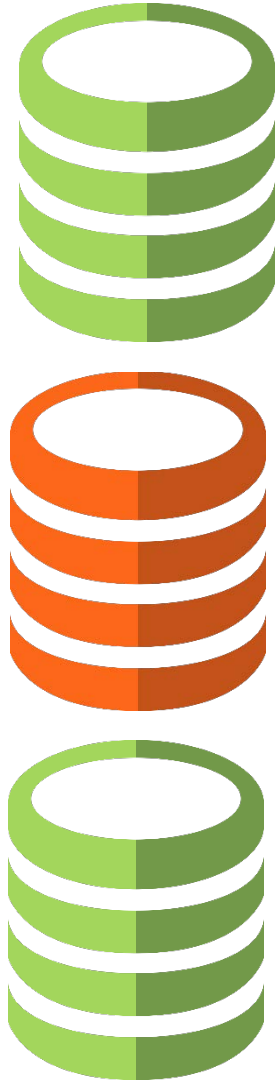
Automation

Two-way

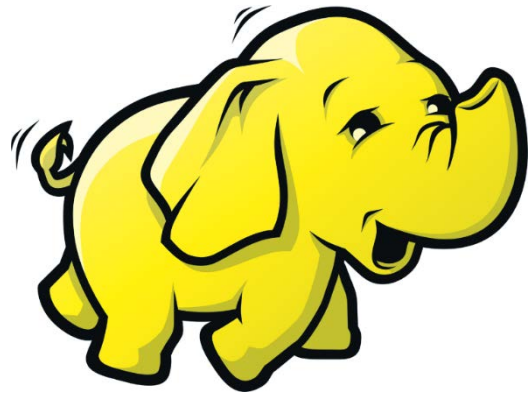
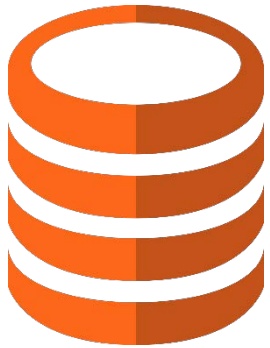


## Sqoop Use Cases



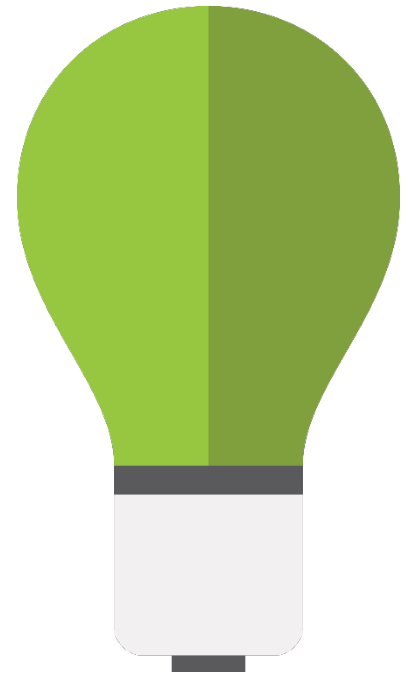






# Problems Sqoop Solves

- Relational database
  - Don't want to remove production data
- Multiple data warehouses
  - Data not shared
- Automating workflows



# Demo Documentation

[Sqoop.apache.org](http://Sqoop.apache.org)



# Data

- NASDAQ daily stock quotes
  - Date
  - Open
  - High
  - Close
  - Volume
  - Adj Close



# Demo MySQL

Setting up test database





# Sqoop Commands

```
sqoop list-tables --connect jdbc:mysql://(MySQL  
Address)/dbname  
//displays tables
```

## List Tables

Display tables in MySQL database

```
sqoop import --connect jdbc:mysql://(MySQL Address)/dbname  
--table tableName -m 1  
//imports data
```

## Import

Import data from MySQL to HDFS



```
sqoop import --connect jdbc:mysql://localhost/test --query  
'SELECT * FROM ... WHERE x > y AND $CONDITIONS -m 1
```

//imports data based off query

## Import Query

Import data from MySQL database based off a query

# Environment

Database → test

Table → nasdaq\_daily

MySQL address → localhost

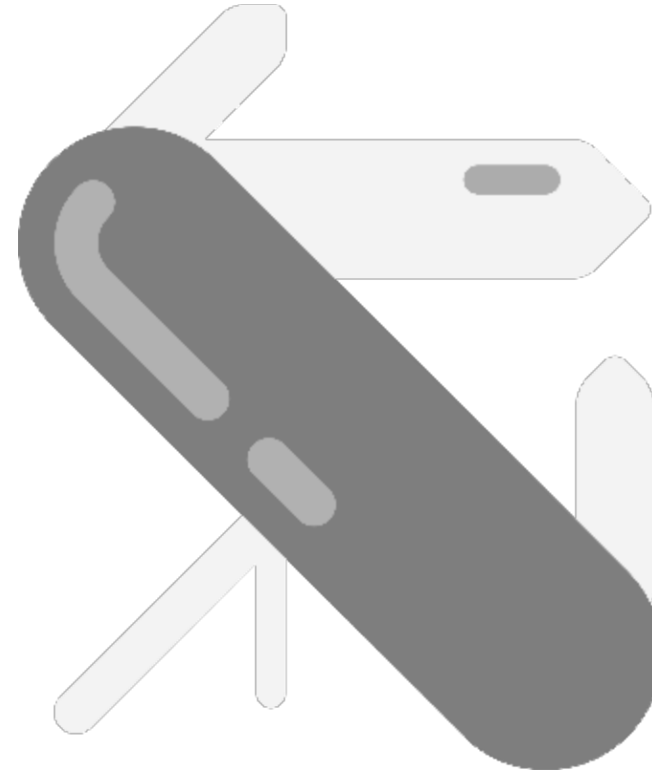
# Demo Moving Data

Moving MySQL table into HDFS

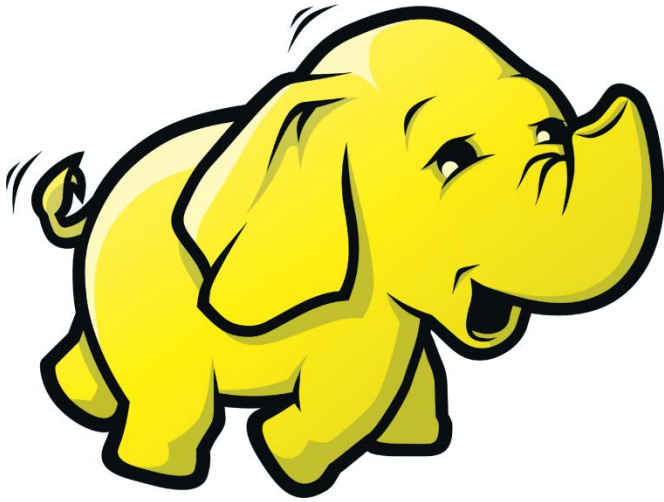


# Learn More

- Official Documentation
  - [Sqoop.apache.org](http://Sqoop.apache.org)
  - IRC channel
- Hadoop Distributors
  - Hortonworks
  - Cloudera
- Get involved
  - Join the mailing list
  - Source code



# Summary



Learned about Sqoop

Talked about Sqoop use cases

Moved data from MySQL to HDFS

Found more resources for Sqoop