

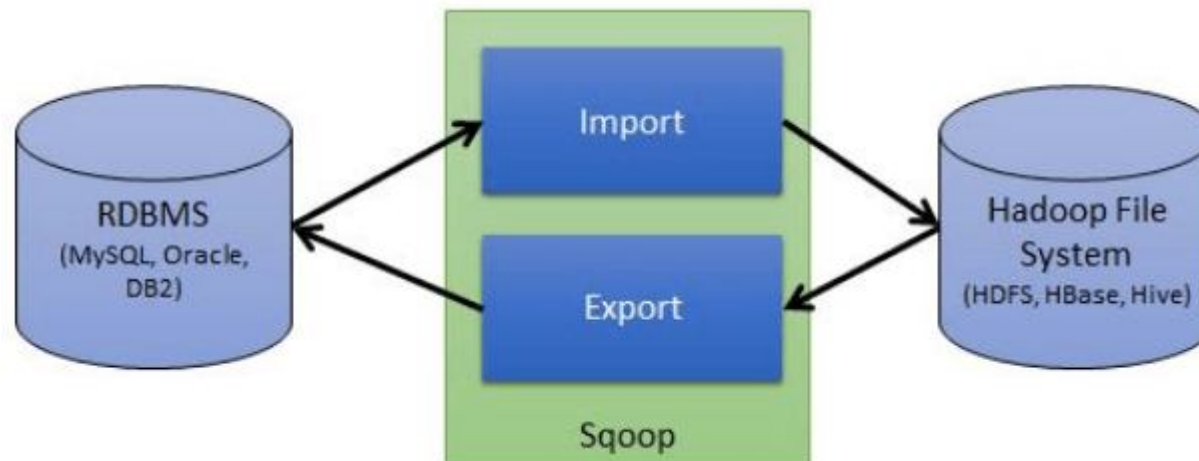
# SQOOP AND FLUME

# Outline

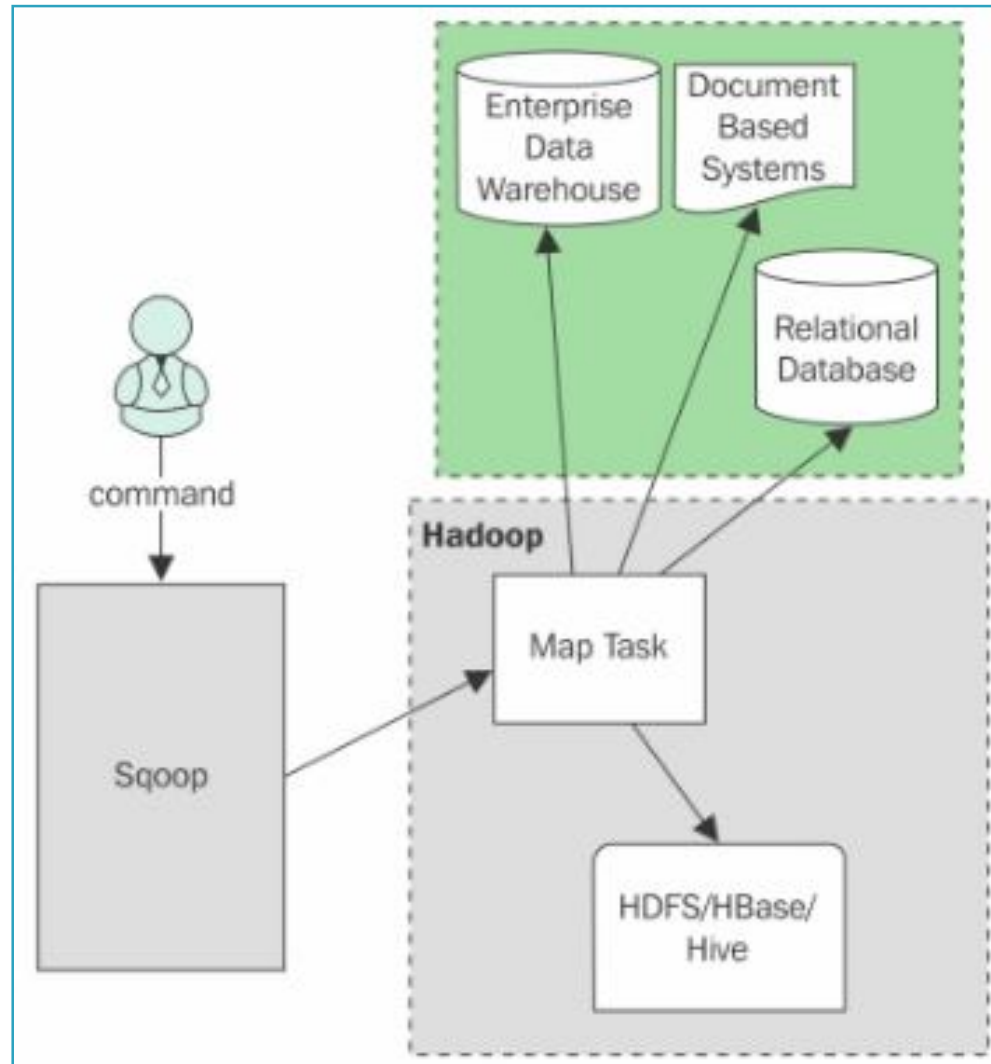
- ▶ Introduction to Sqoop
- ▶ Sqoop Architecture
- ▶ Import / Export Data
- ▶ Introduction to Flume
- ▶ Flume Architecture

# What is Sqoop?

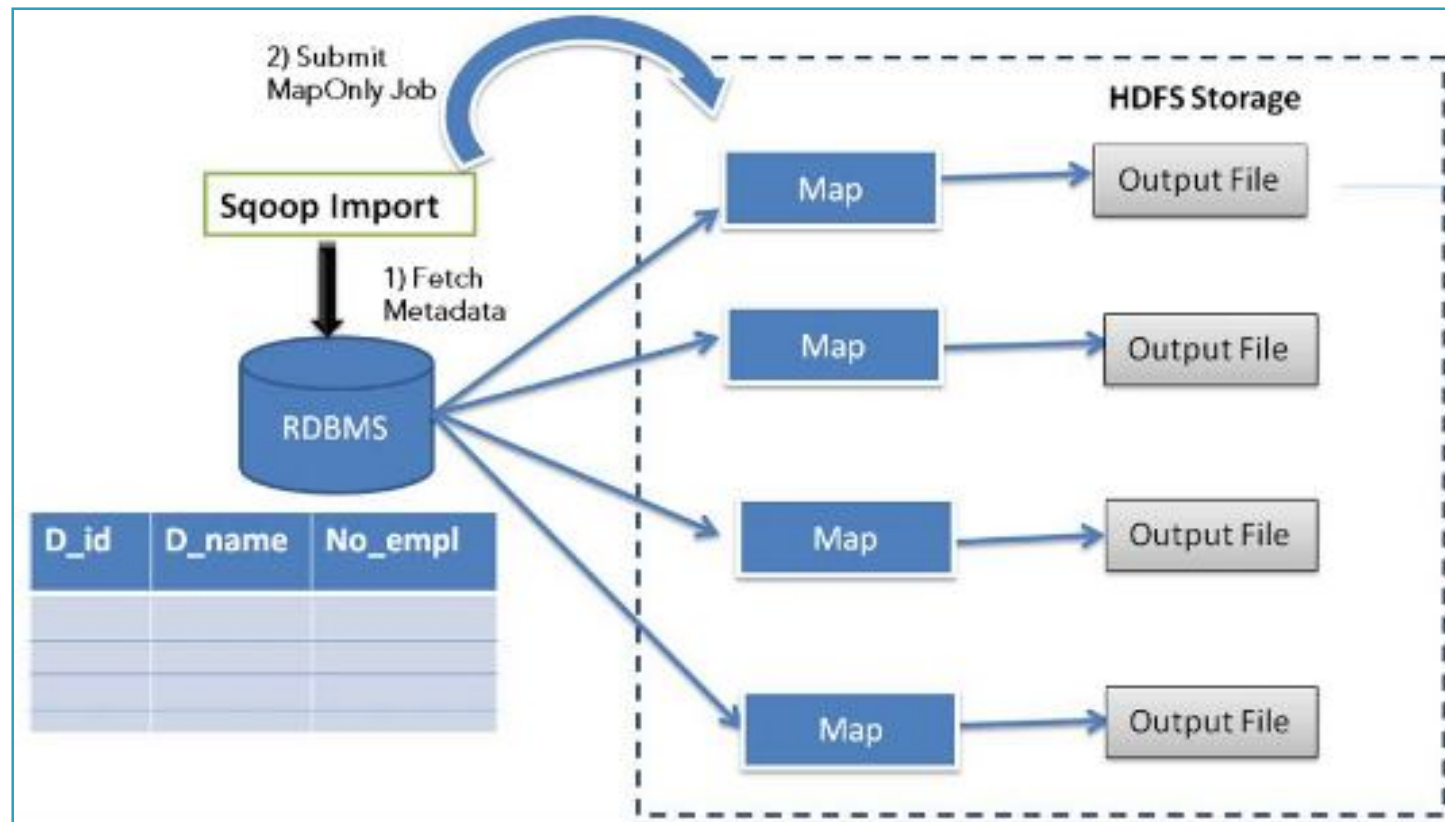
- Tool designed for efficiently transferring bulk data between HDFS and RDBMS
- Import data from RDBMS like MySQL, Oracle into the Hadoop Distributed File System
- Process data in HDFS using MapReduce, PIG, Hive
- Export processed data back into RDBMS
- Sqoop can integrates with Oozie for scheduling and automating import and export tasks
- It supports plugin, so new external system can be integrated easily



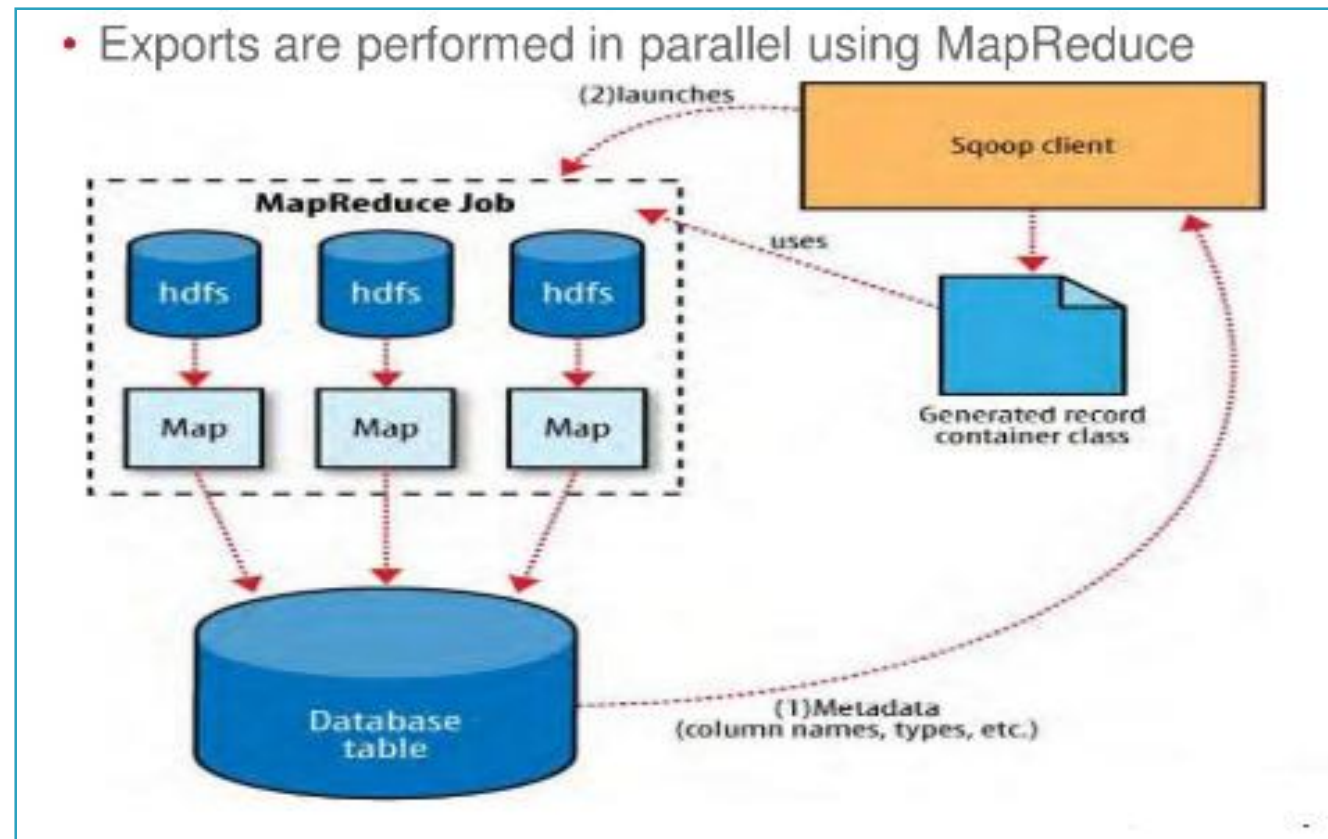
# Sqoop Architecture



# Import Data using Sqoop



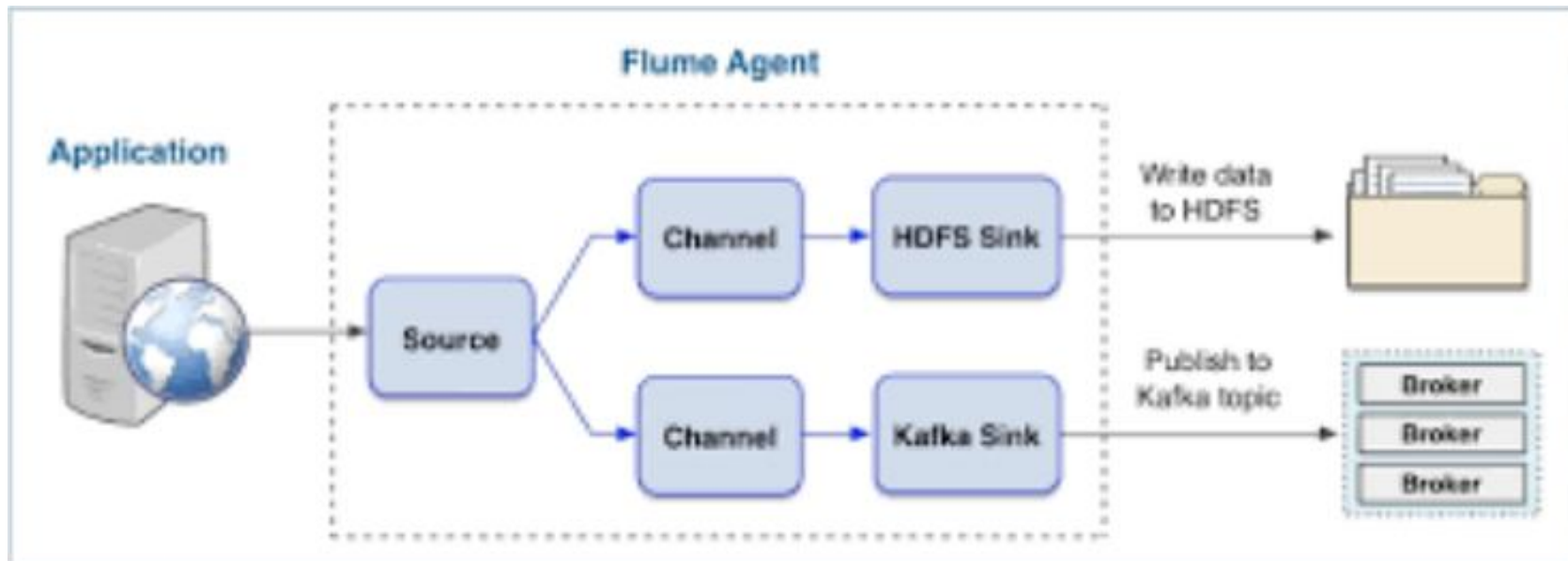
# Export Data using Sqoop



# What is Flume?

- Apache Flume is a distributed and reliable service for efficiently collecting, aggregating, and moving large amounts of log data from one place to another.
- Its main goal is to deliver streaming data from applications to Apache Hadoop's HDFS(Most probably).

# Flume Architecture





**Thank You!**