

## **Flume and Sqoop for Ingesting Big Data**

**Modality:** Self-Paced Learning

**Duration:** 2 Hours

**SATV Value:**

**CLC:**

**NATU:**

**SUBSCRIPTION:** Learn, Master

Import data : Flume and Sqoop play a special role in the Hadoop ecosystem. They transport data from sources like local file systems, HTTP, MySQL and Twitter which hold/produce data to data stores like HDFS, HBase and Hive. Both tools come with built-in functionality and abstract away users from the complexity of transporting data between these systems.

**Flume:** Flume Agents can transport data produced by a streaming application to data stores like HDFS and HBase.

**Sqoop:** Use Sqoop to bulk import data from traditional RDBMS to Hadoop storage architectures like HDFS or Hive.

### **Course Objective:**

Practical implementations for a variety of sources and data stores ..

- Sources : Twitter, MySQL, Spooling Directory, HTTP
- Sinks : HDFS, HBase, Hive

### **Flume features :**

Flume Agents, Flume Events, Event bucketing, Channel selectors, Interceptors

### **Sqoop features :**

Sqoop import from MySQL, Incremental imports using Sqoop Jobs

### **Audience:**

- Engineers building an application with HDFS/HBase/Hive as the data store
- Engineers who want to port data from legacy data stores to HDFS

### **Prerequisite:**

- Knowledge of HDFS is a prerequisite for the course
- HBase and Hive examples assume basic understanding of HBase and Hive shells
- HDFS is required to run most of the examples, so you'll need to have a working installation of

HDFS

## Course Outline:

- You, This Course and Us
- Why do we need Flume and Sqoop?
- Sqoop
- Flume