Tutorials > Getting Started with HDP Sandbox

Getting Started with HDP Sandbox Loading Sensor Data into HDFS

Getting Started with HDP Sandbox **OVERVIEW**

Loading Sensor Data

Concepts

- into HDFS **Hive - Data ETL**
- **Spark Risk Factor Data Reporting With**

Zeppelin

As of January 31, 2021, this tutorial references legacy products that no longer represent Cloudera's current product offerings.

NOTICE

Please visit recommended tutorials: How to Create a CDP Private Cloud Base Development Cluster

Download Sandbox

- All Cloudera Data Platform (CDP) related tutorials
- Introduction

Ready to Get Started?

directories and then load two files into HDFS using the Ambari Files User View. **Prerequisites**

The tutorial is a part of series of hands on tutorial to get you started on HDP using Hortonworks sandbox. Please ensure you complete the prerequisites before proceeding with this tutorial. Downloaded and deployed the Hortonworks Data Platform (HDP) Sandbox

In this section, you will download the sensor data and load that into HDFS using Ambari User Views. You will get introduced to

the Ambari Files User View to manage files. You can perform tasks like create directories, navigate file systems and upload

files to HDFS. In addition, you'll perform a few other file-related tasks as well. Once you get the basics, you will create two

Learning the Ropes of the HDP Sandbox **Outline**

HDFS backdrop

- Summary
- Further Reading

Download and Extract Sensor Data Files

HDFS backdrop

Load the Sensor Data into HDFS

A single physical machine gets saturated with its storage capacity as the data grows. This growth drives the need to partition

Sandbox

₽0

your data across separate machines. This type of File system that manages storage of data across a network of machines is called Distributed File Systems. HDFS is a core component of Apache Hadoop and is designed to store large files with

is now expanded to support **heterogeneous storage** media within the HDFS cluster. **Download and Extract Sensor Data Files**

2. Save the **Geolocation.zip** file to your computer, then extract the files. You should see a Geolocation folder that contains the following files: **geolocation.csv** - This is the collected geolocation data from the trucks. It contains records showing *truck* location, date, time, type of event, speed, etc.

trucks.csv - This is data was exported from a relational database and it shows information on truck models,

streaming data access patterns, running on clusters of commodity hardware. With Hortonworks Data Platform (HDP), HDFS

1. Download the sample sensor data contained in a compressed (.zip) folder here: Geolocation.zip

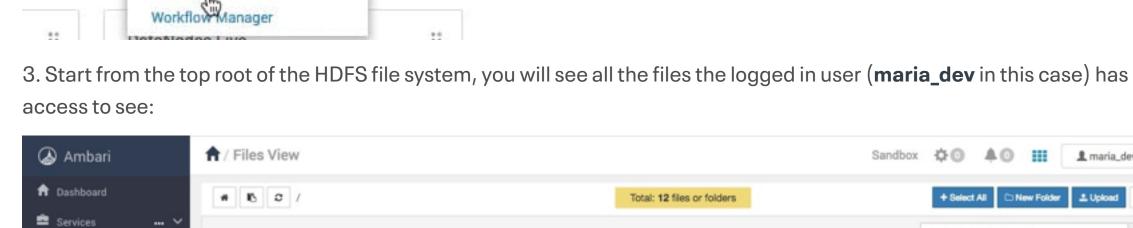
1. Logon to Ambari using: maria_dev/maria_dev 2. Go to Ambari Dashboard and open Files View.

maria_dev -

driverid, truckid, and aggregated mileage info.

Load the Sensor Data into HDFS

ST 1 HOUR -Views



Name >

2018-09-20 08:09 mapred : hdfs drwxr-xr-x mapred hadoop 2018-09-20 08:43 No app-logs

Last Modified >

Size >

Total: 12 files or folders

Owner >

Group >

Permission

+ Select All

Currently supports single file upload

Modified Nov 10, 2013, 4:49 PM

A Q

Erasure Coding

+ Select All

drwxr-xr-x

drwxrwxr-x

drwxr-xr-x

drwx-wx-wx

Sandbox

♣ maria_dev -

Encrypted

Encrypted

No

No

No

No

Next

No

No

Q

Last opened Nov 10, 2013, 4:49 PM Add Tags...

x Cancel

☆○ ▲○ Ⅲ

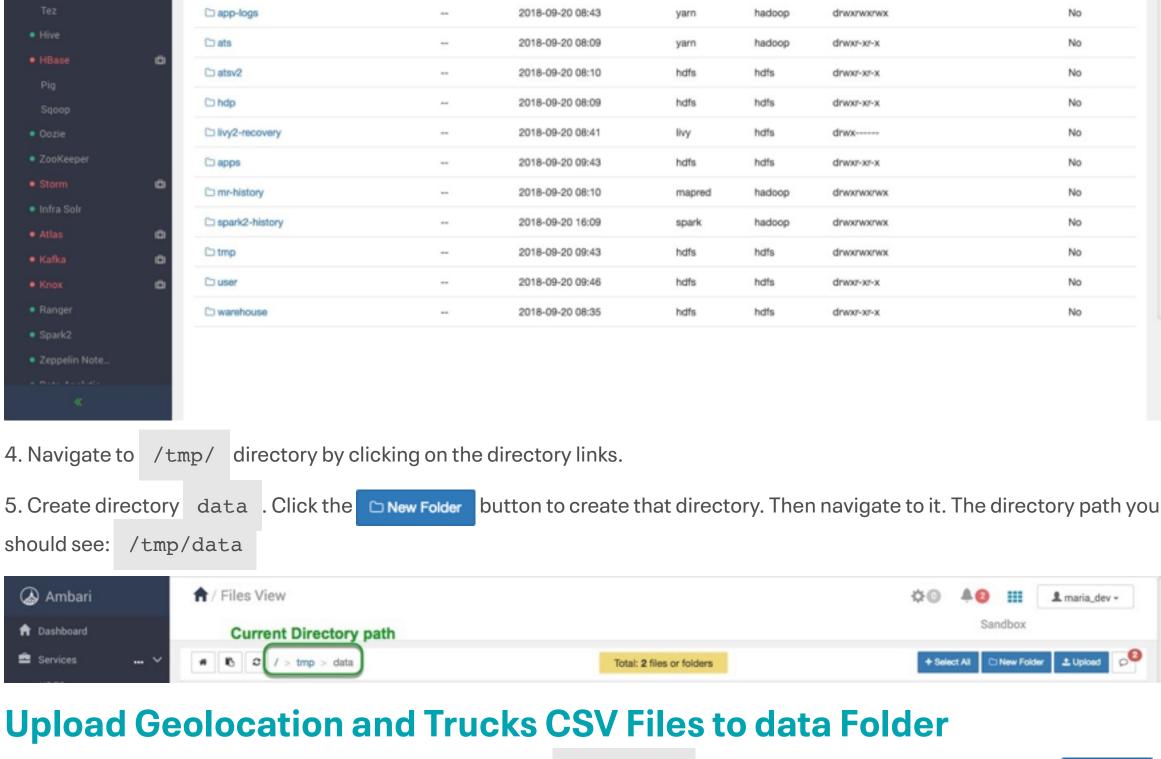
Search in current directory.

Erasure Coding

maria_dev -

Encrypted

Q



2. An **Upload file** window will appear, click on the cloud symbol. lb 2 / > tmp > data Total: 0 files or folders

button to upload the corresponding geolocation.csv and trucks.csv files into it.

× Upload file to /tmp/data Q Name > Encrypted

Drag file to upload or click to browse

1. If you're not already in your newly created directory path /tmp/data, go to the data folder. Then click on the

3. Another window will appear, navigate to the destination the two csv files were downloaded. Click on one at a time, press open to complete the upload. Repeat the process until both files are uploaded. ## **■ Ⅲ** Geolocation 0 Q Search geolocation.csv **Favorites** trucks.csv iCloud Drive Applications Desktop Documents Documents trucks.csv Downloads Movies 61 KB Created Nov 10, 2013, 4:49 PM Music

Format: All Files Options Cancel Open Both files are uploaded to HDFS as shown in the Files View UI: Ambari ♠ / Files View n Dashboard **Current Directory path** Services # 15 2 / > tmp > data Total: 2 files or folders Size > Last Modified > Name > Owner > Group > 514.3 kB 2018-11-26 13:08 geolocation.csv • HBase trucks.csv 59.9 kB 2018-11-26 13:08 Newly uploaded files

You can also perform the following operations on a file or folder by clicking on the entity's row: Open, Rename, Permissions, Delete, Copy, Move, Download and Concatenate.

2. Click Permissions.

₽ 2 / > tmp

□ data

☐ druid-indexing

netity-file-history

Refer to image for a visual explanation.

Open Rename Permissions 2 lete Copy

Set Write Permissions to Write to data Folder

3. Make sure that the background of all the write boxes are checked (blue).

1. click on the data folder's row, which is contained within the directory path /tmp/.

0 Files, 1 Folders selected

Write

Congratulations! Let's summarize the skills and knowledge we acquired from this tutorial. We learned **Hadoop Distributed**

File System (HDFS) was built to manage storing data across multiple machines. Now we can upload data into the HDFS

Execute

Read

Pictures

imedel jmedel

Size Name Permission Erasure Coding Write Read Execute Read Write Execute Group

201

201

201

Other

- hive 201 × Cancel ☑ Save Summary
 - HDFS HDFS User Guide HDFS Architecture Guide

using Ambari's HDFS Files view.

Further Reading

- **Previous**

y in

Contact Us

English

US: +1 888 789 1488

Outside the US: +1 650 362 0488

Partners

Support Community **Documentation**

HDP OPERATIONS: HADOOP ADMINISTRATION

Company **Get started** Resources About us Certification Blog Contact sales Careers CDP resources Diversity, Equality & Inclusion Downloads **CDP Trust Center** Find a partner Community **Events** Find a solution Documentation Leadership

Training

Tutorials

How can I help you today?

Resources library

Support

Careers

© 2021 Cloudera, Inc. All rights reserved. Terms & Conditions | Privacy Statement and Data Policy | Unsubscribe / Do Not Sell My Personal Information | Cook

Apache Hadoop and associated open source project names are trademarks of the Apache Software Foundation. For a complete list of trademarks, click here.

Locations

Newsroom