**ETL Pipeline Testing screenshot**

Contents

[Setup Hadoop Environment 1](#_Toc7621184)

[Run instructions 1](#_Toc7621185)

[https://github.com/fazilaijaz/msd.git 2](#_Toc7621186)

[Run Data ingestion in hadoop and hive 2](#_Toc7621187)

[Run loading script 2](#_Toc7621188)

[Verify table and data 2](#_Toc7621189)

[Run spark data aggregation and store result in hive 2](#_Toc7621190)

[Run spark app 2](#_Toc7621191)

[Verify tables and data 2](#_Toc7621192)

[Run Data Visualization 3](#_Toc7621193)

# Setup Hadoop Environment

Follow the “Setup\_hadoop\_environment.docx” guide to setup the hortonworks sandbox using VM Player on windows.

# Run instructions

1. Run Sandbox Hortonworks Sandbox through VM Player Open SSH connection to VM in Browser.

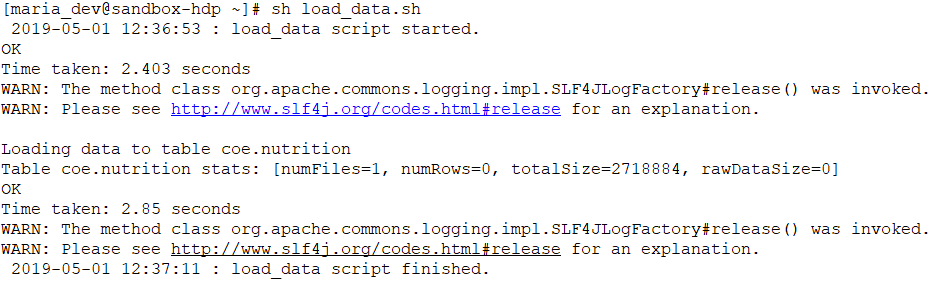
192.168.122.128:4200

1. Login using **maria\_**dev
2. Clone or download

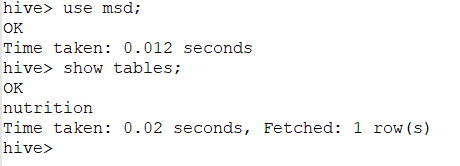
# <https://github.com/fazilaijaz/msd.git>

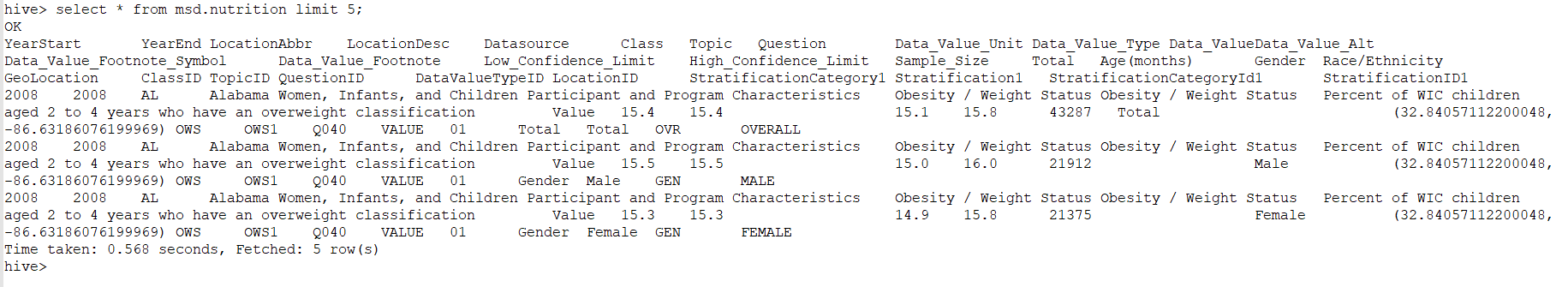
# Run Data ingestion in hadoop and hive

## Run loading script



## Verify table and data



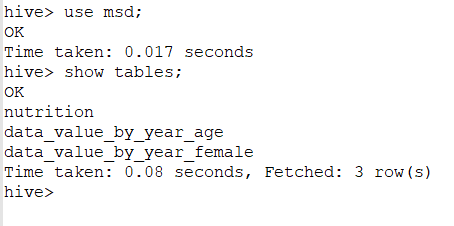


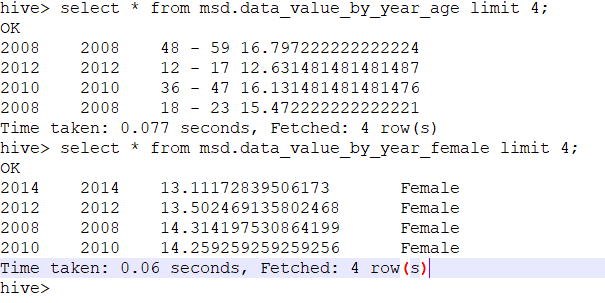
# Run spark data aggregation and store result in hive

## Run spark app

sh run\_spark.sh

## Verify tables and data





# Run Data Visualization

Open Visualization directory in windows.

Open **index.html** with **Firefox**

**Click on the button to display required data.**

