

SCENARIO: You get a CSV file by email and you want to put it on your cluster. We need to copy this file into our Impala cluster through the shell.

1. Move file into cluster. Usually through SSH/SFTP, we do it through Docker

- Retrieve Docker container id (think of it as server IP):

```
sudo docker ps
```

- Copy file:

```
sudo docker cp myfile.csv [FIRST DIGITS OF  
IMAGE]:/myfile.csv
```

2. Copy into HDFS filesystem (Hadoop) so that Impala can read it.

- `hdfs dfs -mkdir -p /user/username/sample_data/airbnb`
- `hdfs dfs -put airbnb.csv`
`/user/username/sample_data/airbnb`

3. Make connection between file in HDFS and associated table in Metastore (`CREATE TABLE`).