============Text file==================

1. **Creating table**

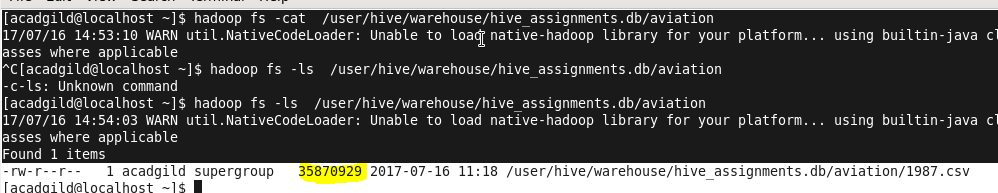
CREATE TABLE AVIATION(year int, month int, flight\_no int, org string, dest string, distance int, cancelled int, cancellation\_code string, diverted int, dep\_delay string)

row format delimited

fields terminated by ','

stored as textfile;

1. **Load data**
2. **File size:**



**Original text file size is around 30 MB.**

============RC file With compression ==================

To store with compression we need to set following properties:

Enabling compression provides performance gains in most cases and is supported for RCFile and SequenceFile tables.

For example, to enable Snappy compression, you would specify the following additional settings when loading data through the Hive shell.

SET hive.exec.comprxess.output=true;

SET mapred.max.split.size=256000000;

SET mapred.output.compression.type=BLOCK; -- block compression for sequence file

SET mapred.output.compression.codec=org.apache.hadoop.io.compress.SnappyCodec;

CREATE TABLE AVIATION\_RC\_COMPRESSION(year int, month int, flight\_no int, org string, dest string, distance int, cancelled int, cancellation\_code string, diverted int, dep\_delay string)

row format delimited

fields terminated by '\t'

stored as rcfile;

Inserting into AVIATION\_RC\_COMPRESSION :

INSERT OVERWRITE TABLE AVIATION\_RC\_COMPRESSION

SELECT \* FROM AVIATION;

================ORC file=======================

1. **Creating table :**

CREATE TABLE AVIATION\_ORC(year int, month int, flight\_no int, org string, dest string, distance int, cancelled int, cancellation\_code string, diverted int, dep\_delay string)

row format delimited

fields terminated by ','

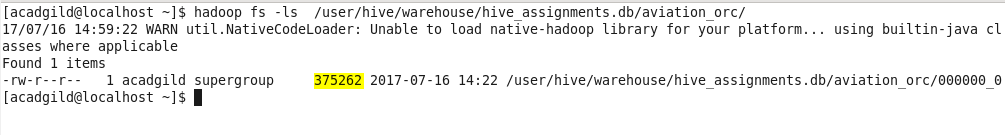
stored as orcfile;

1. **Insert data:**

INSERT OVERWRITE TABLE AVIATION\_ORC

SELECT \* FROM AVIATION;

1. **File size :**



**For ORC file it is just 0.3 MB**