

## SKIPPING HEADER AND FOOTER ROWS IN HIVE

Posted by Naveen P.N | Data Engineering, Hadoop

In this blog post we will explain you how to “skip header and footer rows in hive”.

In Hive we can ignore N number of rows from top and bottom from a file using TBLPROPTIES clause. The TBLPROPERTIES clause provides various feature which can be set as per our need.

```
tool name: quality measure utility tool, tool id: #20823
file created on: 2017-dec-07 15:12:33
file number for the month: 368
20171207|1|pkt 11010|green|small|high
20171207|2|pkt 11011|red|small|low
20171207|3|pkt 11012|black|small|low
20171207|4|pkt 11013|green|large|high
20171207|5|pkt 11014|red|medium|high
total rows: 5
last modified at: 2017-dec-07 15:35:50
```

The diagram illustrates a text file structure with three distinct sections. The top section, labeled 'Headers to be skipped', contains three lines of metadata: 'tool name: quality measure utility tool, tool id: #20823', 'file created on: 2017-dec-07 15:12:33', and 'file number for the month: 368'. The middle section, labeled 'Data to be read', contains five lines of data: '20171207|1|pkt 11010|green|small|high', '20171207|2|pkt 11011|red|small|low', '20171207|3|pkt 11012|black|small|low', '20171207|4|pkt 11013|green|large|high', and '20171207|5|pkt 11014|red|medium|high'. The bottom section, labeled 'Footers to be skipped', contains two lines: 'total rows: 5' and 'last modified at: 2017-dec-07 15:35:50'. Red boxes and arrows highlight these sections.

Now assume that we are dealing with above text file which is being generated by an automated tool which adds 3 additional rows at the very beginning of the file and 2 additional rows at the end of the file. These extra rows are added to provide additional details about the file itself.

The first 3 rows contain the below details:

- Name of the tool which generates the file,
- Date along with time on which the file was generated and
- The file number (a sequential number which resets at the start of every month) for that particular month.

Similarly, the last 2 rows contain the below details:



- i. The total number of rows in the file (which can be used to verify missing records) and
- ii. The date along with time on which the file was last modified.

### Download the data set

The total number of rows in the file (which can be used to verify missing records) and

The date along with time on which the file was last modified.

```
[code lang="sql"]
CREATE DATABASE IF NOT EXISTS testdb;
USE testDB;

DROP TABLE IF EXISTS testdb.sample_table;
CREATE TABLE IF NOT EXISTS testdb.sample_table
(
    log_date string,
    row_num int,
    pkt_number string,
    pkt_color string,
    pkt_size string,
    pkt_quality string
)
ROW FORMAT DELIMITED FIELDS TERMINATED BY '|'
LINES TERMINATED BY '\n' STORED AS TEXTFILE
TBLPROPERTIES('skip.header.line.count'='3', 'skip.footer.line.count')
```



Learn and Master **Big Data Architect** course from **NPN Training** which includes **(Hadoop + Apache Storm + Spark + Scala + MongoDB + Kafka)** all in at just one course from **NPN Training** which is designed to enhance you knowledge and skill to become a successful Big Data Architect.

[Check out our Big Data course](#)

 Post Views: 1,225

SHARE:   



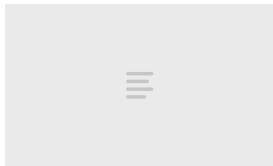
## ABOUT THE AUTHOR



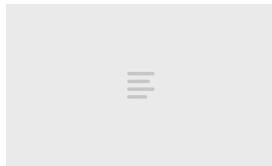
### Naveen P.N

12+ years of experience in IT with vast experience in executing complex projects using Java, Micro Services , Big Data and Cloud Platforms. I found **NPN Training** Pvt Ltd a India based startup to provide high quality training for IT professionals. I have trained more than 3000+ IT professionals and helped them to succeed in their career in different technologies. I am very passionate about Technology and Training. I have spent 12 years at **Siemens, Yahoo, Amazon and Cisco**, developing and managing technology.

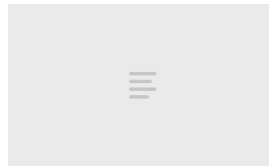
## RELATED POSTS



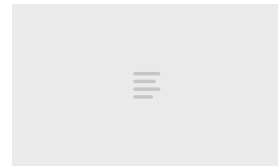
**Hive Architecture**  
July 26, 2016



**Hive Optimization Techniques in Hadoop 2.x**  
October 16, 2016



**Can Reducer always be reused for Combiner?**  
July 26, 2016



**Hive Interview Questions**  
February 19, 2020

## CATEGORIES

[Apache Spark](#)[BDD-Cucumber](#)[Data Engineering](#)

## COURSES WE OFFER

[Python Training](#)[Selenium Training](#)

Data Science	Big Data Hadoop Training
Hadoop	Apache Spark And Scala Training
Java Programming	Big Data Architect Masters Program
Kafka	Data Science Training
Machine Learning	
Microservices	
Scala Programming	
Selenium	
Statistics	
Test Automation	
Uncategorized	

## CONTACT US

**Website :** [www.npntraining.com](http://www.npntraining.com)

**Email Id :** [info@npntraining.com](mailto:info@npntraining.com)

**Address :** #35 31st main BTM 2nd Stage,  
Tank Shore / Tank Bund Road, Bengaluru, Karnataka 560068.

**Land Mark:** Near to AXA Building Complex

**Mobile No :** 9535584691 / 8095918383



