



Hive-Word Count Examples



Hive Word Count Examples

Step1: Create input table contain input file

```
CREATE TABLE IF NOT EXISTS input(line string);
```

Step 2: Load input data

```
LOAD DATA LOCAL INPATH '/home/coolbaski/wc.txt'  
OVERWRITE INTO TABLE input;
```

Step 3: WordCount

```
CREATE TABLE word_counts AS SELECT word,count(1)  
AS count FROM (SELECT explode(split(line, ' ')) AS  
word FROM input) w GROUP BY word ORDER BY word;
```

Hive Word Count Examples-

Understand step3 in more details

Step3.a: Convert sentence into words

- The input file in sentences format
- First we have to convert that it into words
- Applying space as delimiter.
- Use **split function** of hive. Ex:-**split(line, ' ')**

Step3.b: Convert column into rows

- Output of step3.a is in array of strings
- we have to convert it into multiple rows
- explode function in hive, convert every line of data into multiple rows
- we have built in table generation UDTF explode function

SELECT explode(split(line, ' ')) AS word FROM input

- create above output as intermediate table as follows

SELECT explode(split(line, ' ')) AS word FROM input) w

Hive Word Count Examples-

Understand step3 in more details

Step3.c: **apply group by to count word occurrences**

```
SELECT word,count(1) AS count FROM (SELECT explode(split(line, ' '))  
AS word FROM input) w GROUP BY word ORDER BY word;
```

Step3.d: Store the results in another table

```
CREATE TABLE word_counts AS SELECT word,count(1) AS count FROM  
(SELECT explode(split(line, ' ')) AS word FROM input) w GROUP BY  
word ORDER BY word;
```

```
Select * from word_counts;
```