


Word Count In Hive

 [hadooplessons.info/2014/12/in-this-post-i-am-going-to-discuss-how.html](http://www.hadooplessons.info/2014/12/in-this-post-i-am-going-to-discuss-how.html)

In this post I am going to discuss how to write word count program in Hive.

Assume we have data in our table like below

This is a Hadoop Post
and Hadoop is a big data technology

and we want to generate word count like below

a 2
and 1
Big 1
data 1
Hadoop 2
is 2
Post 1
technology 1
This 1

Now we will learn how to write program for the same.

1.Convert sentence into words

the data we have is in sentences,first we have to convert that it into words applying space as delimiter.we have to use [split function](#) of hive.

```
split (sentence , ' ')
```

2.Convert column into rows

Now we have array of strings like this
[This,is,a,hadoop,Post]
but we have to convert it into multiple rows like below

This
is
a
hadoop
Post

I mean we have to convert every line of data into multiple rows ,for this we have function called [explode in hive](#) and this is also called table generating function.

```
SELECT explode(split(sentence, ' ')) AS word FROM texttable
```

and create above output as intermediate table.

```
(SELECT explode(split(sentence, ' ')) AS word FROM texttable)tempTable
```

after second step you should get output like below

```
a
a
and
Big
data
Hadoop
Hadoop
is
is
Post
technology
This
```

3.Apply group by

after second step , it is straight forward ,we have to apply group by to count word occurrences.

```
select word,count(1) as count from
(SELECT explode(split(sentence, ' ')) AS word FROM texttable)tempTable
group by word
```