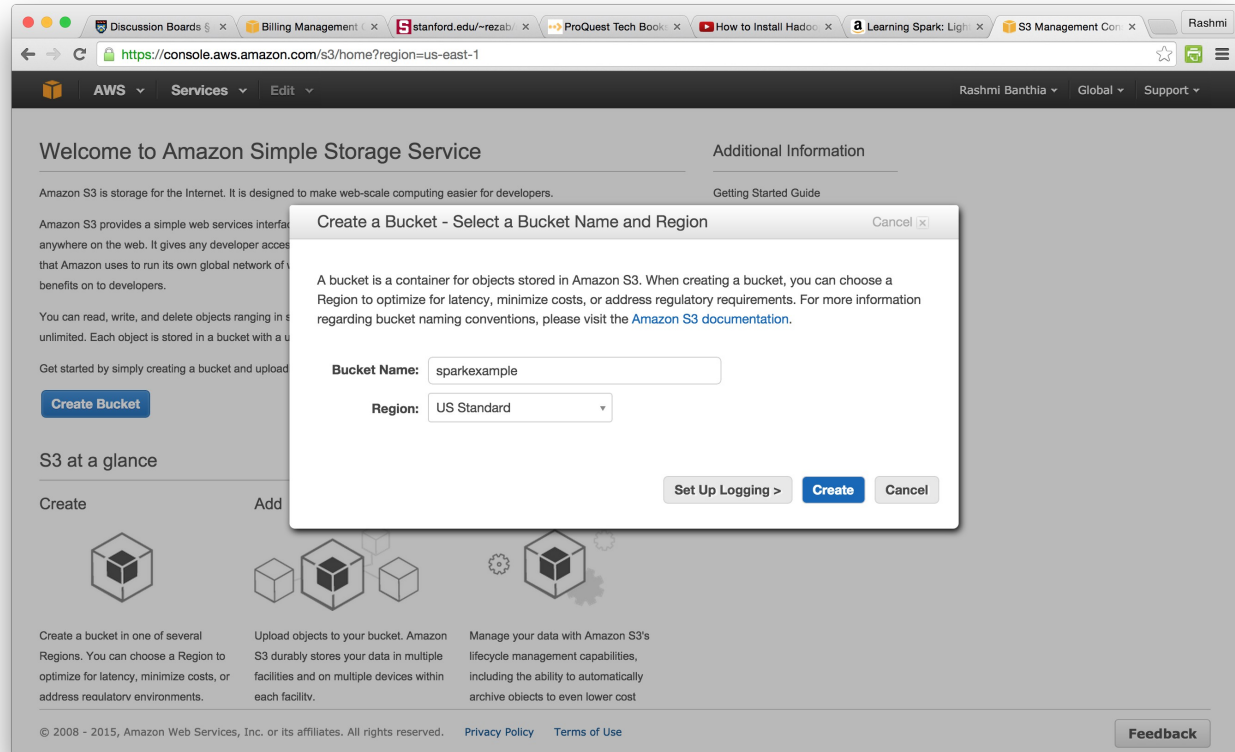


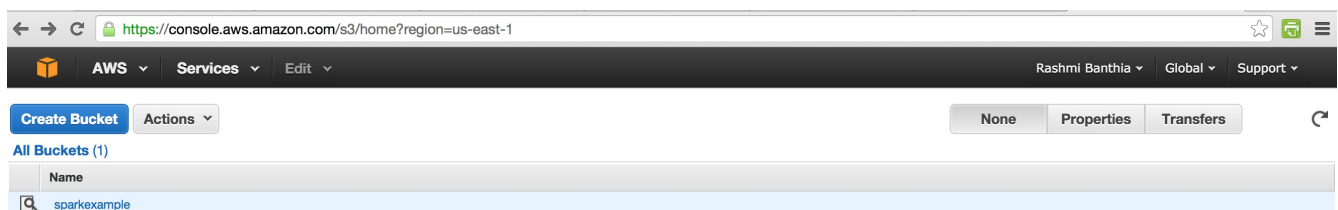
## Apache Spark Examples

### Creating RDD using S3 bucket

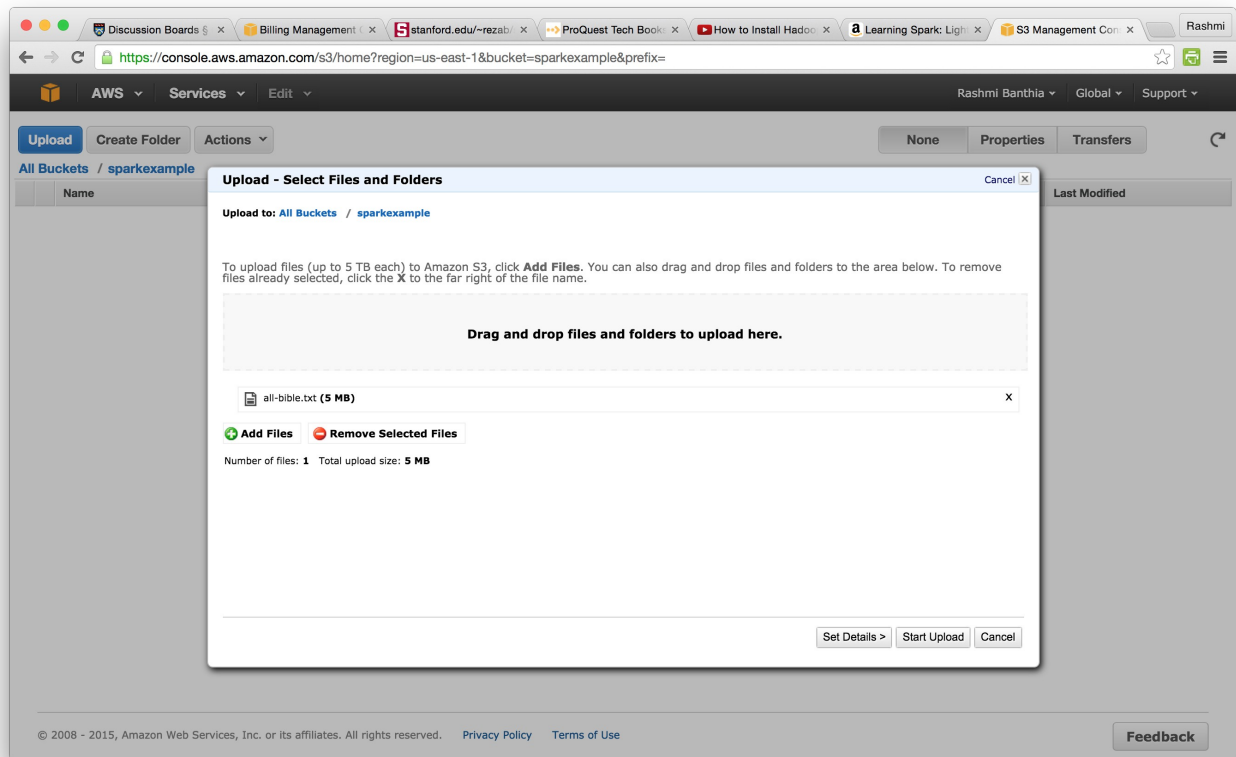
Create a S3 bucket

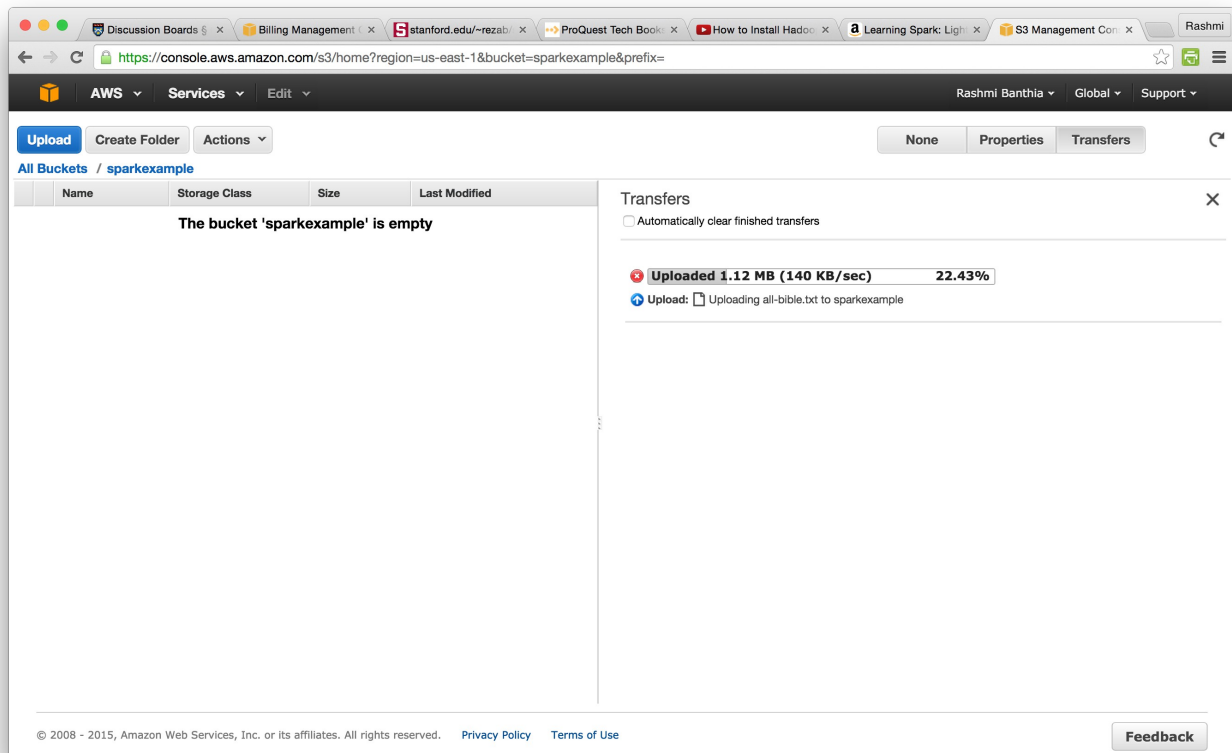


Click on sparkexample



Click on Upload

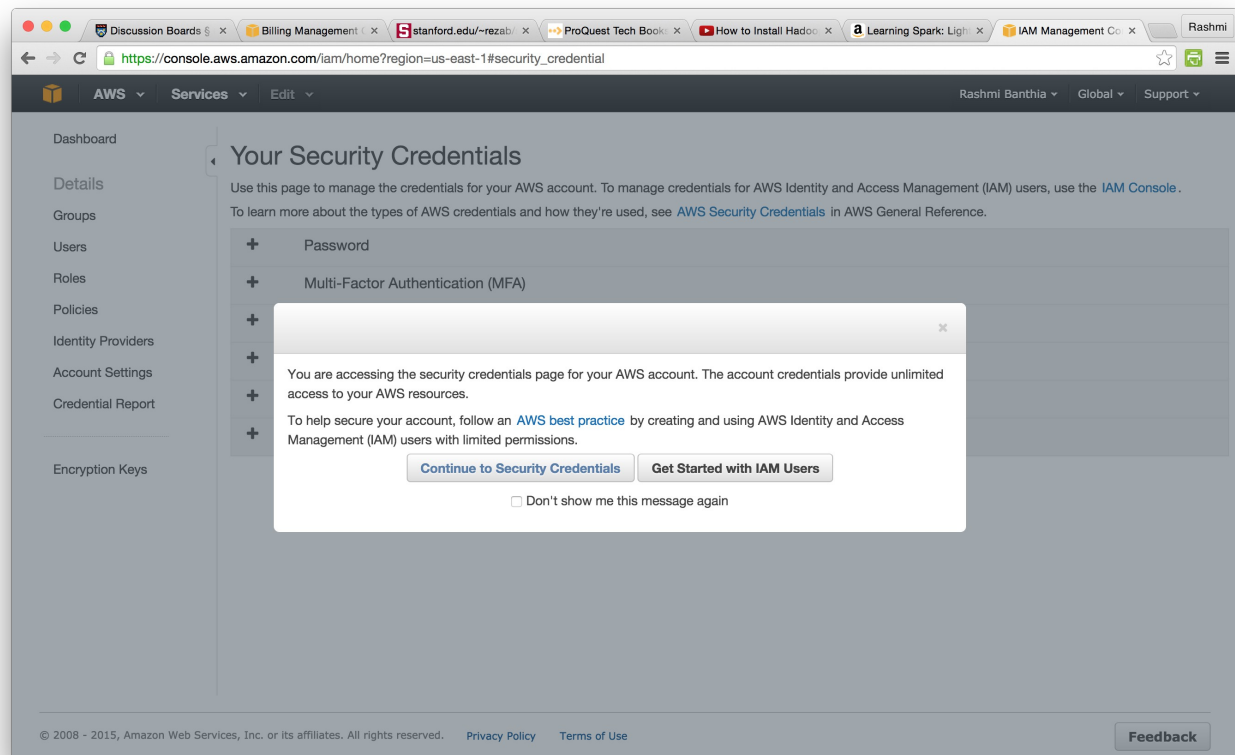




URL of the file:

<https://s3.amazonaws.com/sparkexample/all-bible.txt>

Click on Security Credentials (top right corner)



Create new Access keys and download it.

**Access Keys (Access Key ID and Secret Access Key)**

You use access keys to sign programmatic requests to AWS services. To learn how to sign requests using your access keys, see the [signing documentation](#). For your protection, store your access keys securely and do not share them. In addition, AWS recommends that you rotate your access keys every 90 days.

Note: You can have a maximum of two access keys (active or inactive) at a time.

Created	Deleted	Access Key ID	Status	Actions
Feb 16th 2015		AKIAJYK3XCNVROD5L4JA	Active	<a href="#">Make Inactive</a>   <a href="#">Delete</a>
Apr 13th 2015		AKIAJENL2KJ77JCLJZ2A	Active	<a href="#">Make Inactive</a>   <a href="#">Delete</a>
Apr 13th 2015	Apr 13th 2015	AKIAJEIWFZK6VRFNTS4A	Deleted	

[Create New Access Key](#)

**Important Change - Managing Your AWS Secret Access Keys**

Rootkey.csv

AWSAccessKeyId=AKIAJENL2KJ77JCLJZ2A

AWSSecretKey=poFAOg0dnykUsNNKHxaj+vXG5z2VZCX3aSTR8YyV



On VM, (or you can put it in .bash\_profile)

```
export AWS_ACCESS_KEY_ID=AKIAJENL2KJ77JCLJZ2A  
export AWS_SECRET_ACCESS_KEY=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
```

(On python spark shell)

\$pyspark

```
>>> s3lines = sc.textFile("s3n://sparkexample/all-bible.txt")  
>>> s3lines.count()  
117154
```

Stop/Delete the S3 bucket