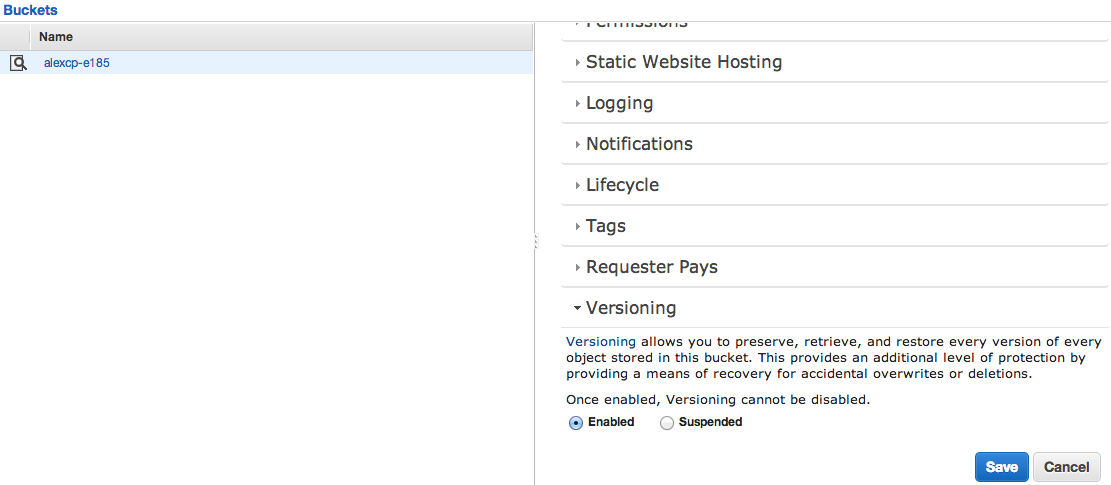
**HU Extension Assignment 03 E-185 Big Data Analytics**

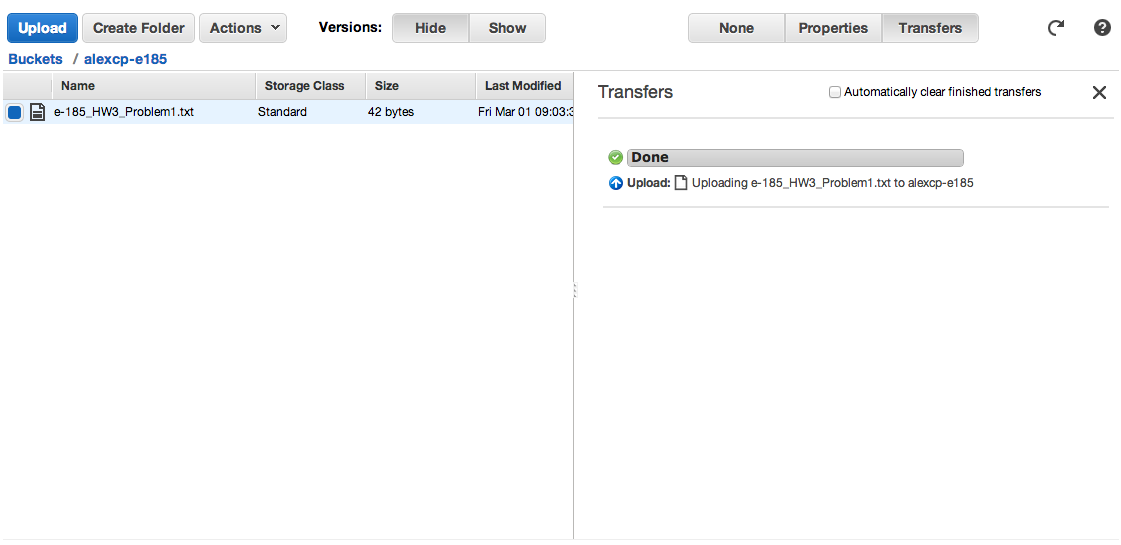
**Handed out: 02/23/2013 Due by 17:25PM on Friday, 03/01/2013**

**Problem 1. Create a bucket. Enable versioning. Upload a single text document. A few minutes later modify the document a tiny bit. Upload modified document. Demonstrate that you can retrieve both versions of the document.**

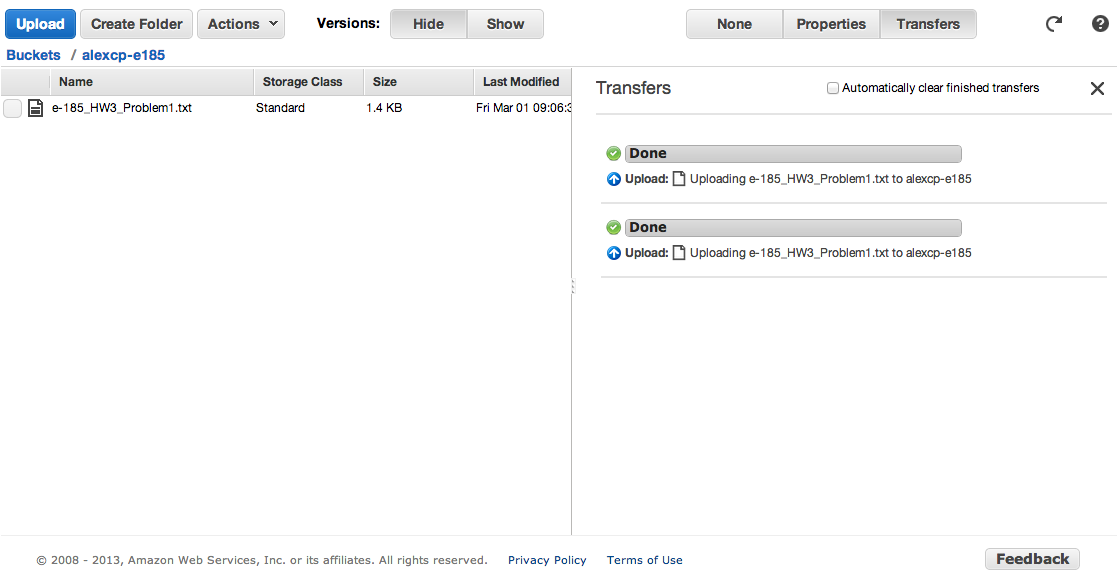
First, we sign in to AWS and request a bucket to be created at S3, and then we configure it to enable versioning



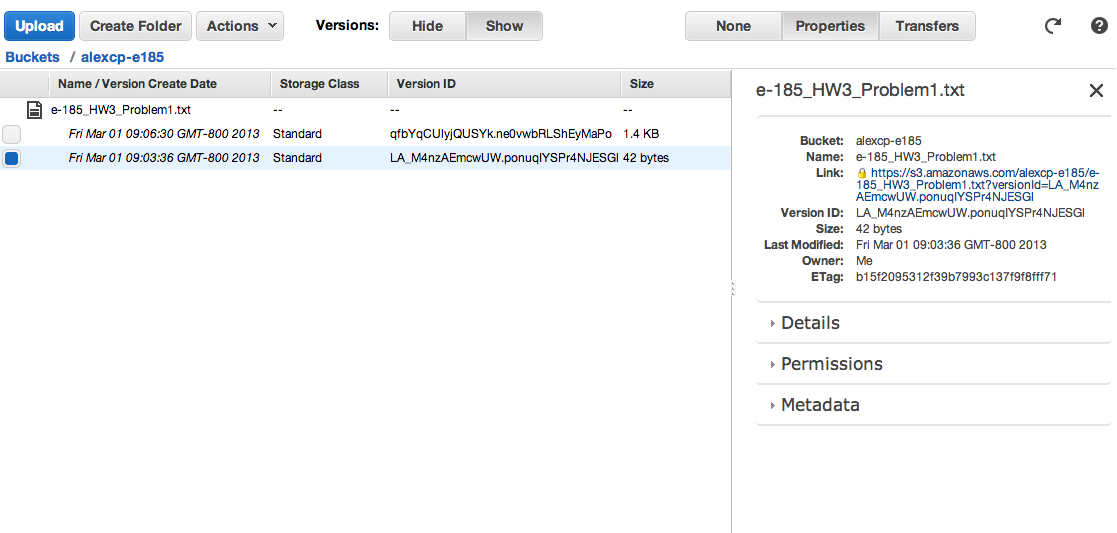
We upload the initial version of the document (42 bytes)



And later, the second, bigger one (1.4kbytes)

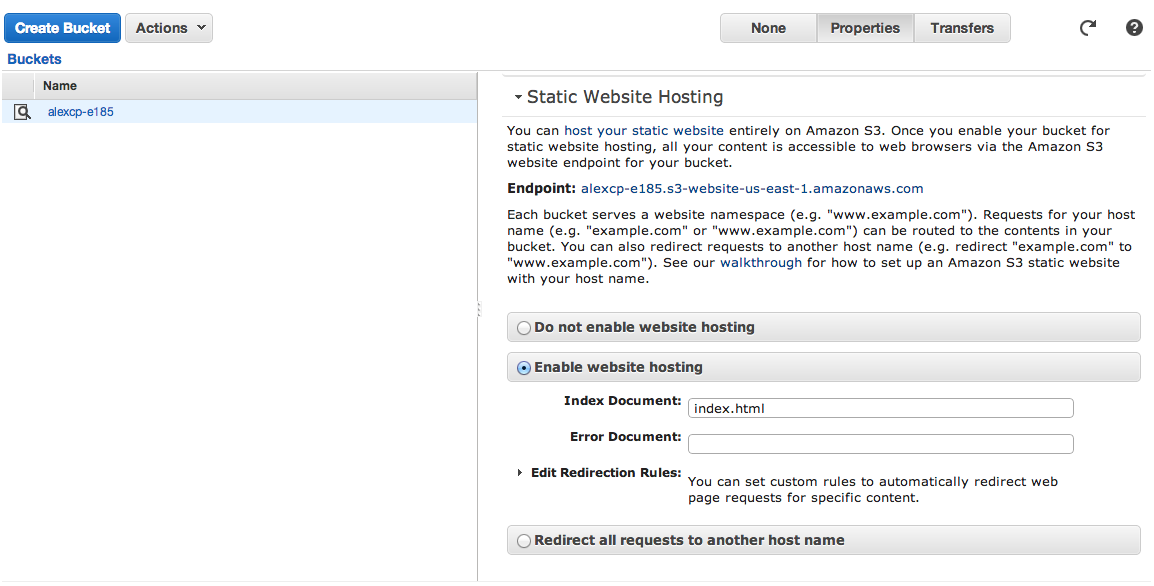


By asking the Management Console to "Show" version, we are able to examine the properties and download both versions of the file.



**Problem 2. Create a folder for your dog’s Web site. This is a very simple Web site and has only two HTML pages but you can navigate from one to the other. Demonstrate that you can serve your Web site from an AWS S3 bucket.**

First, we enable the "Static Website Hosting" option in our bucket:

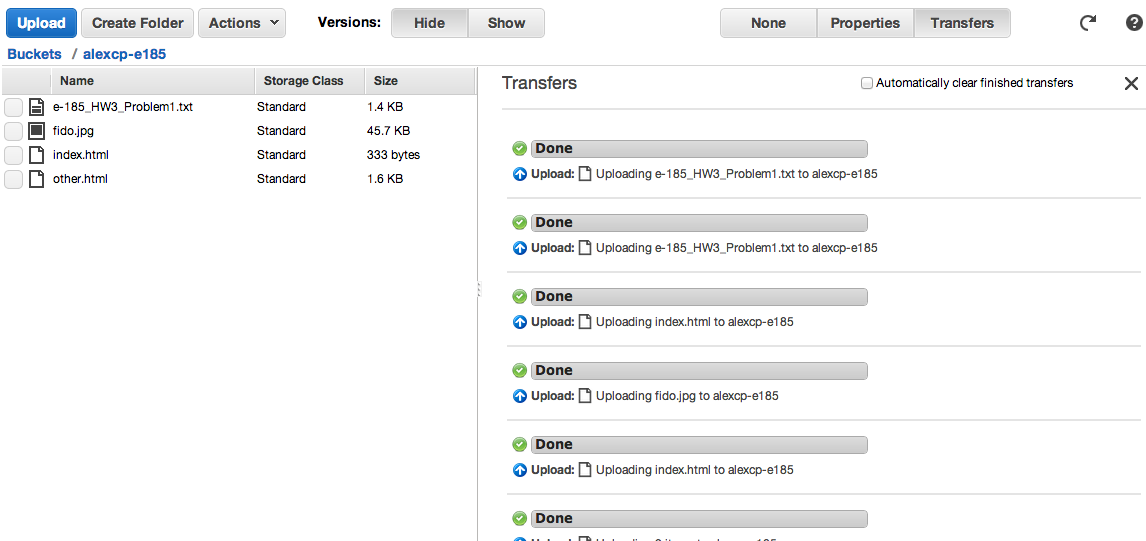


Then we upload a Bucket Policy that will allow anonymous GET access to the files in this bucket (kindly provided by AWS here:

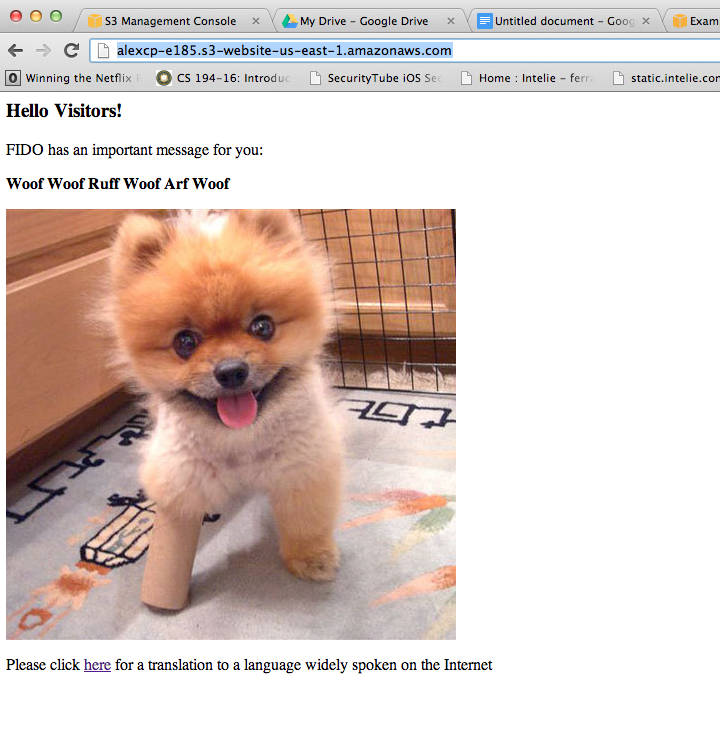
<http://docs.aws.amazon.com/AmazonS3/latest/dev/HostingWebsiteOnS3Setup.html>)



Finally we upload the files for the website.



And then the website is available at <http://alexcp-e185.s3-website-us-east-1.amazonaws.com/>. Please visit it; it would make my dog very happy.



**Problem 3. Install Ruby 1.8.7 and the elastic map reduce command interface tools. Demonstrate that you can run word count example we did in class from the command line of your PC or Mac. The command looks like:**

**ruby elastic-mapreduce --create –stream \  
 --mapper s3://elasticmapreduce/samples/wordcount/wordSplitter.py \**

**--input s3://elasticmapreduce/samples/wordcount/input \**

**--output s3n://yourbucket \ # A path to your bucket own on Amazon S3 --reducer aggregate**

**The output will look similar to:**

**Created job flow JobFlowID.**

**Download results to your PC or Mac and verify the content. Show a few lines to us. Use facilities of hadoop distributed file system shell to fetch for use input file 00002. We always look at 00001 and are quite curious what is in 00002. In your report, please display the first 20 and the last 20 lines of that file.**

The command-line options were installed on Mac OS X by using homebrew (<http://mxcl.github.com/homebrew/>) to install the latest versions of ruby, ec2-api-tools and ec2-ami-tools. These packages already automate much of the needed PATH configuration to simplify the process.

Then, the following procedures were used to download the necessary credentials and configure the environment

<http://docs.aws.amazon.com/AWSEC2/latest/UserGuide/SettingUp_CommandLine.html>

After the configuration of the API keys, we can demonstrate that the tools are operational:

aperture:~ alexcp$ ec2-describe-regions

REGION eu-west-1 ec2.eu-west-1.amazonaws.com

REGION sa-east-1 ec2.sa-east-1.amazonaws.com

REGION us-east-1 ec2.us-east-1.amazonaws.com

REGION ap-northeast-1 ec2.ap-northeast-1.amazonaws.com

REGION us-west-2 ec2.us-west-2.amazonaws.com

REGION us-west-1 ec2.us-west-1.amazonaws.com

REGION ap-southeast-1 ec2.ap-southeast-1.amazonaws.com

REGION ap-southeast-2 ec2.ap-southeast-2.amazonaws.com

Then, we proceeded to configure the Elastic MapReduce tools according to another tutorial provided by AWS on

<http://docs.aws.amazon.com/ElasticMapReduce/latest/GettingStartedGuide/SignUp.html>.

After some fun and games with incompatible ruby versions, the tool was up and running.

aperture:~ alexcp$ elastic-mapreduce --version

Version 2012-12-17

With the tools configured, we execute the job requested adding an option to keep the instance alive after the termination of the job, so that it can be accessed later.

aperture:~ alexcp$ elastic-mapreduce --create --stream --mapper s3://elasticmapreduce/samples/wordcount/wordSplitter.py --input s3://elasticmapreduce/samples/wordcount/input --output s3n://alexcp-e185-assign03-p3/output --reducer aggregate --alive

Created job flow j-2RRKZQ7RNBY2E

We check from time to time to see if the job has completed

aperture:~ alexcp$ elastic-mapreduce --list

j-2RRKZQ7RNBY2E STARTING Development Job Flow (requires manual termination)

PENDING Example Streaming Step

After a while, the job completes

aperture:~ alexcp$ elastic-mapreduce --list

j-2RRKZQ7RNBY2E WAITING ec2-50-19-11-52.compute-1.amazonaws.com Development Job Flow (requires manual termination)

COMPLETED Example Streaming Step

The result was downloaded as **problem3\_output.txt** and is included along with this document.

aperture:~ alexcp$ head -15 problem3-output.txt

a 14716

aa 52

aac 3

aai 3

aaiun 3

aak 9

aakar 3

aalborg 3

aargau 3

aarhus 3

aatash 3

ababa 19

abad 3

abadgaran 3

abaiang 3

aperture:~ alexcp$ tail -15 problem3-output.txt

zuzana 3

zuzemberk 3

zvecan 3

zviad 3

zvonko 3

zvornik 5

zvrskovec 3

zw 5

zwd 6

zwe 2

zwelinzima 3

zyanon 3

zylfi 3

zyuganov 6

zzs 12

In order to review the input file, we connect to the master instance in EC2 that we kept alive from the previous job.

aperture:~ alexcp$ ssh -i .ec2/ec2-keypair hadoop@ec2-50-19-11-52.compute-1.amazonaws.com

We get a copy of the file from the S3 bucket.

hadoop@domU-12-31-39-0A-4C-A4:~$ hadoop fs -copyToLocal s3://elasticmapreduce/samples/wordcount/input/0002 wordcount-0002

13/03/01 20:56:35 INFO s3native.NativeS3FileSystem: Opening 's3://elasticmapreduce/samples/wordcount/input/0002' for reading

And we check its contents:

hadoop@domU-12-31-39-0A-4C-A4:~$ head -20 wordcount-0002

Transnational Issues

Nepal

Disputes - international:

joint border commission continues to work on contested sections of boundary with India, including the 400 square kilometer dispute over the source of the Kalapani River; India has instituted a stricter border regime to restrict transit of Maoist insurgents and illegal cross-border activities; approximately 106,000 Bhutanese Lhotshampas (Hindus) have been confined in refugee camps in southeastern Nepal since 1990

Refugees and internally displaced persons:

refugees (country of origin): 107,803 (Bhutan); 20,153 (Tibet/China)

IDPs: 50,000-70,000 (remaining from ten-year Maoist insurgency that officially ended in 2006; displacement spread across the country) (2007)

Illicit drugs:

illicit producer of cannabis and hashish for the domestic and international drug markets; transit point for opiates from Southeast Asia to the West

This page was last updated on 1 January 2003

This page was last updated on 19 March, 2009

&nbsp;

CIA - The World Factbook -- Fiji

a { font-family: Verdana, Arial, Helvetica, sans-serif; font-size: 12px; color: #000000; text-decoration: none}

&nbsp;

Country

List | World Factbook Home

The World Factbook

&nbsp;

Fiji

hadoop@domU-12-31-39-0A-4C-A4:~$ tail -20 wordcount-0002

Angola is rebuilding its country after the end of a 27-year civil war in 2002. Fighting between the Popular Movement for the Liberation of Angola (MPLA), led by Jose Eduardo DOS SANTOS, and the National Union for the Total Independence of Angola (UNITA), led by Jonas SAVIMBI, followed independence from Portugal in 1975. Peace seemed imminent in 1992 when Angola held national elections, but fighting picked up again by 1996. Up to 1.5 million lives may have been lost - and 4 million people displaced - in the quarter century of fighting. SAVIMBI's death in 2002 ended UNITA's insurgency and strengthened the MPLA's hold on power. President DOS SANTOS held legislative elections in September 2008, and announced plans to hold presidential elections in 2009.

Anguilla

Colonized by English settlers from Saint Kitts in 1650, Anguilla was administered by Great Britain until the early 19th century, when the island - against the wishes of the inhabitants - was incorporated into a single British dependency, along with Saint Kitts and Nevis. Several attempts at separation failed. In 1971, two years after a revolt, Anguilla was finally allowed to secede; this arrangement was formally recognized in 1980, with Anguilla becoming a separate British dependency.

Antarctica

Speculation over the existence of a "southern land" was not confirmed until the early 1820s when British and American commercial operators and British and Russian national expeditions began exploring the Antarctic Peninsula region and other areas south of the Antarctic Circle. Not until 1840 was it established that Antarctica was indeed a continent and not just a group of islands. Several exploration "firsts" were achieved in the early 20th century. Following World War II, there was an upsurge in scientific research on the continent. A number of countries have set up a range of year-round and seasonal stations, camps, and refuges to support scientific research in Antarctica. Seven have made territorial claims, but not all countries recognize these claims. In order to form a legal framework for the activities of nations on the continent, an Antarctic Treaty was negotiated that neither denies nor gives recognition to existing territorial claims; signed in 1959, it entered into force in 1961.

Antigua and Barbuda

The Siboney were the first to inhabit the islands of Antigua and Barbuda in 2400 B.C., but Arawak Indians populated the islands when COLUMBUS landed on his second voyage in 1493. Early settlements by the Spanish and French were succeeded by the English who formed a colony in 1667. Slavery, established to run the sugar plantations on Antigua, was abolished in 1834. The islands became an independent state within the British Commonwealth of Nations in 1981.

Arctic Ocean

The Arctic Ocean is the smallest of the world's five oceans (after the Pacific Ocean, Atlantic Ocean, Indian Ocean, and the recently delimited Southern Ocean). The Northwest Passage (US and Canada) and Northern Sea Route (Norway and Russia) are two important seasonal waterways. A sparse network of air, ocean, river, and land routes circumscribes the Arctic Ocean.

Argentina

In 1816, the United Provinces of the Rio Plata declared their independence from Spain. After Bolivia, Paraguay, and Uruguay went their separate ways, the area that remained became Argentina. The country's population and culture were heavily shaped by immigrants from throughout Europe, but most particularly Italy and Spain, which provided the largest percentage of newcomers from 1860 to 1930. Up until about the mid-20th century, much of Argentina's history was dominated by periods of internal political conflict between Federalists and Unitarians and between civilian and military factions. After World War II, an era of Peronist populism and direct and indirect military interference in subsequent governments was followed by a military junta that took power in 1976. Democracy returned in 1983 after a failed bid to seize the Falkland (Malvinas) Islands by force, and has persisted despite numerous challenges, the most formidable of which was a severe economic crisis in 2001-02 that led to violent public protests and the resignation of several interim presidents.

Armenia

Armenia prides itself on being the first nation to formally adopt Christianity (early 4th century). Despite periods of autonomy, over the centuries Armenia came under the sway of various empires including the Roman, Byzantine, Arab, Persian, and Ottoman. During World War I in the western portion of Armenia, Ottoman Turkey instituted a policy of forced resettlement coupled with other harsh practices that resulted in an estimated 1 million Armenian deaths. The eastern area of Armenia was ceded by the Ottomans to Russia in 1828; this portion declared its independence in 1918, but was conquered by the Soviet Red Army in 1920. Armenian leaders remain preoccupied by the long conflict with Azerbaijan over Nagorno-Karabakh, a primarily Armenian-populated region, assigned to Soviet Azerbaijan in the 1920s by Moscow. Armenia and Azerbaijan began fighting over the area in 1988; the struggle escalated after both countries attained independence from the Soviet Union in 1991. By May 1994, when a cease-fire took hold, Armenian forces held not only Nagorno-Karabakh but also a significant portion of Azerbaijan proper. The economies of both sides have been hurt by their inability to make substantial progress toward a peaceful resolution. Turkey imposed an economic blockade on Armenia and closed the common border because of the Armenian separatists' control of Nagorno-Karabakh and surrounding areas.

Aruba

Discovered and claimed for Spain in 1499, Aruba was acquired by the Dutch in 1636. The island's economy has been dominated by three main industries. A 19th century gold rush was followed by prosperity brought on by the opening in 1924 of an oil refinery. The last decades of the 20th century saw a boom in the tourism industry. Aruba seceded from the Netherlands Antilles in 1986 and became a separate, autonomous member of the Kingdom of the Netherlands. Movement toward full independence was halted at Aruba's request in 1990.

Ashmore and Cartier Islands

These uninhabited islands came under Australian authority in 1931; formal administration began two years later. Ashmore Reef supports a rich and diverse avian and marine habitat; in 1983, it became a National Nature Reserve. Cartier Island, a former bombing range, became a marine reserve in 2000.

Atlantic Ocean

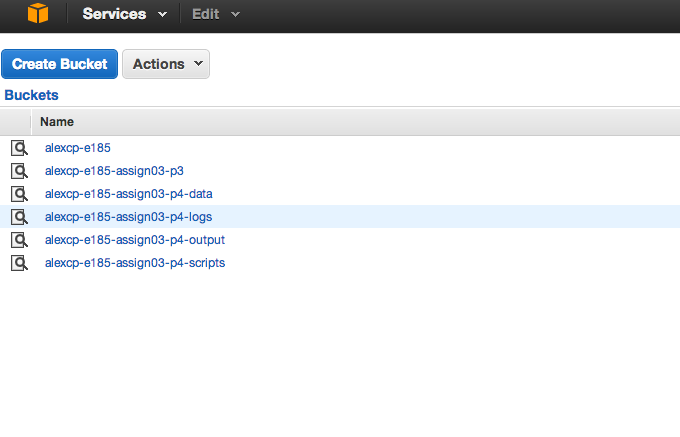
The Atlantic Ocean is the second largest of the world's five oceans (after the Pacific Ocean, but larger than the Indian Ocean, Southern Ocean, and Arctic Ocean). The Kiel Canal (Germany), Oresund (Denmark-Sweden), Bosporus (Turkey), Strait of Gibraltar (Morocco-Spain), and the Saint Lawrence Seaway (Canada-US) are important strategic access waterways. The decision by the International Hydrographic Organization in the spring of 2000 to delimit a fifth world ocean, the Southern Ocean, removed the portion of the Atlantic Ocean south of 60 degrees south latitude.

Australia

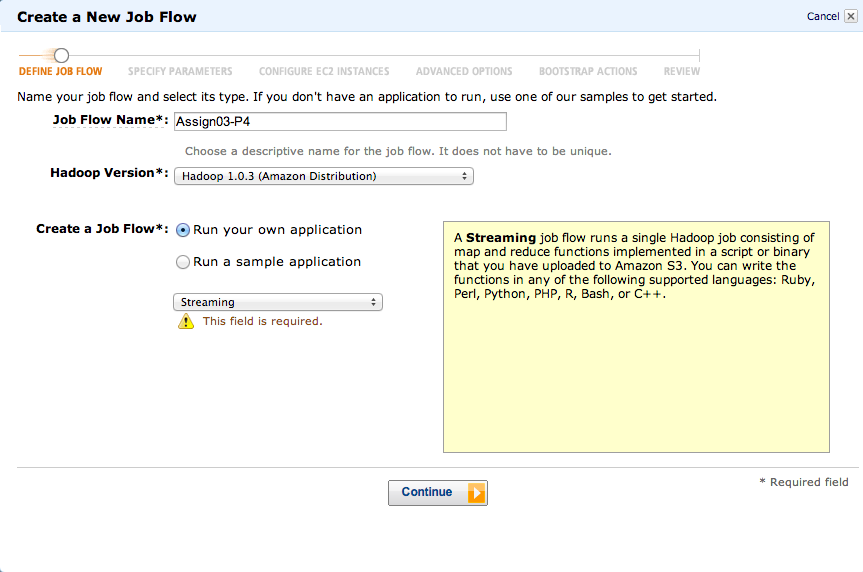
**Problem 4. Attached are two ruby scripts: max\_temperature\_map.rb and max\_temperature\_reduce.rb. Also are attached two sample file containing recordings of metheorogical data from years 1901 and 1902. Your scripts will extract the year and the temperature in Celsius from the every line of those files and then determine what was the highest temperature for each year. Both values are buried in the lines. Years are spelled**

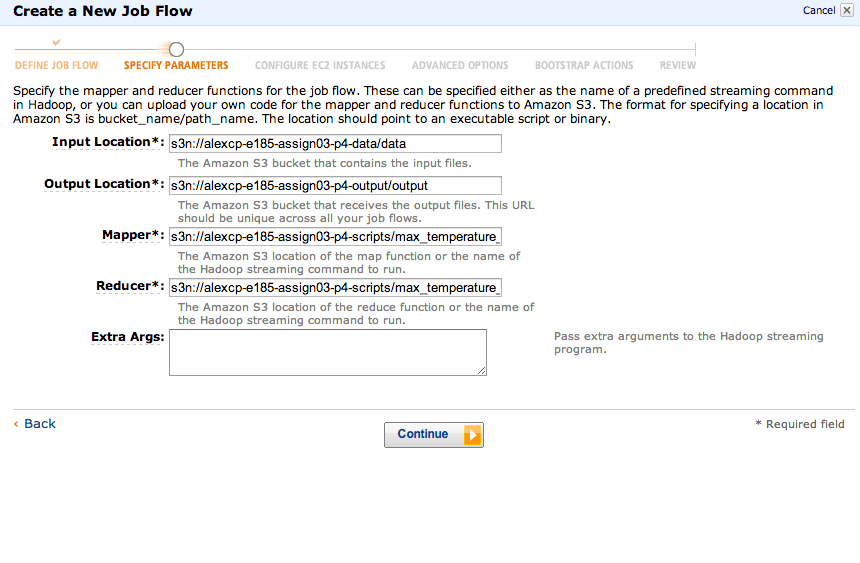
**out as 1901,and 1902. Temperatures are presented as 100 time the actual temperature in Celsius. So, -6.11oC is written as N9-00611+.. Year is extracted as 4 digits starting from position 15 and the temperature as five digits starting at position 87. We are actually not that much interested into those data. We are just familiarizing ourselves with the Elastic MapReduce environment. Upload both scripts to a folder in one of your S3 buckets. Upload two data files to perhaps another bucket. Direct the output to a third bucket and logs to yet another. Run an Elastic Map Reduce job flow as your own application. As the Job Type select Streaming. Retrieve the results and the logs and submit. Capture the interaction with AWS console.**

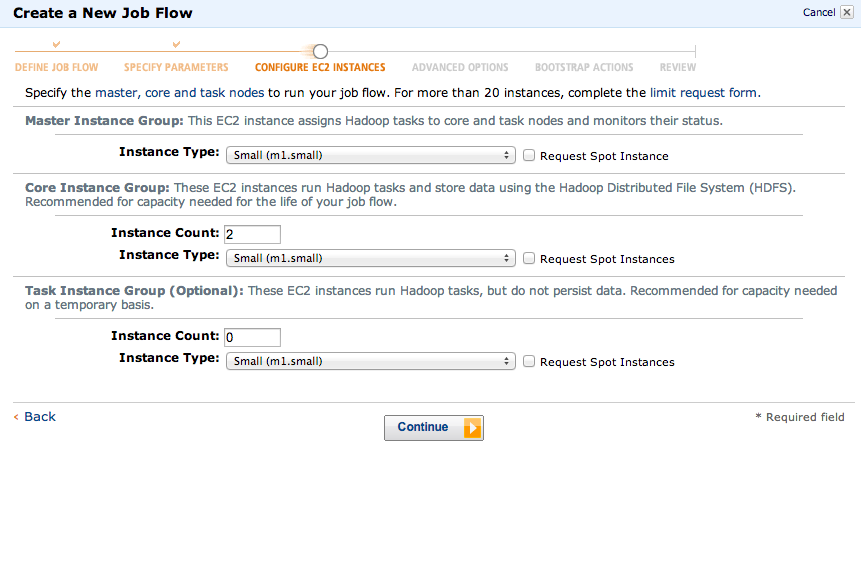
Initially, we create and set the appropriate permissions on the necessary buckets:



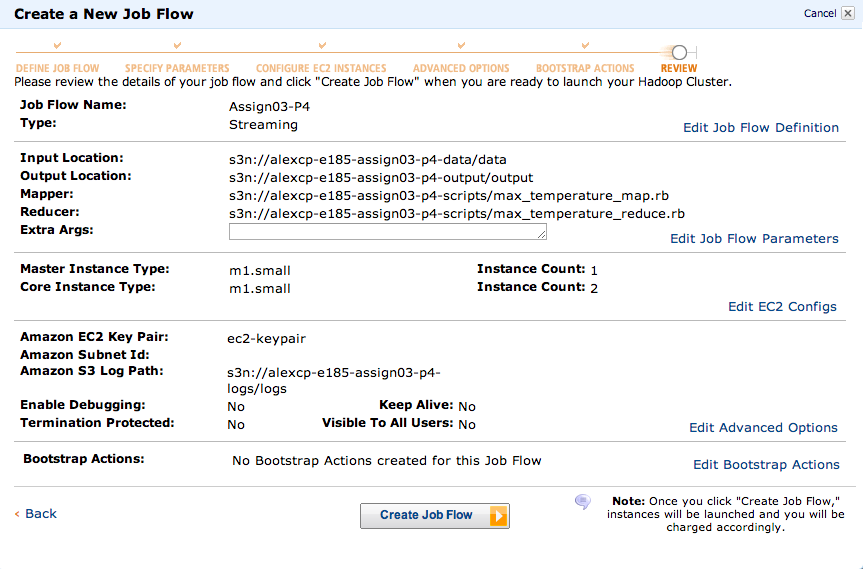
The files were then uploaded to the correct locations in preparation to the next step. We used the AWS Elastic Map Reduce interface as described on the images below to schedule the job.











After the job was completed, the output and logs files were copied from the S3 buckets and have been included in this assignment response. The files were downloaded using the s3cmd package (<http://s3tools.org/s3cmd>) available through Homebrew.