Running Scripts in the Cluster

Author: Saveen Reddy

Date: 8/6/2013

# Overview

## Summary

Now we will run a script in the cluster. First you will need to get access to a cluster and VC. We will use **Cosmos08** cluster and the sandbox VC on that cluster. Then we will familiarize ourselves with the basic mechanics of submitting jobs.

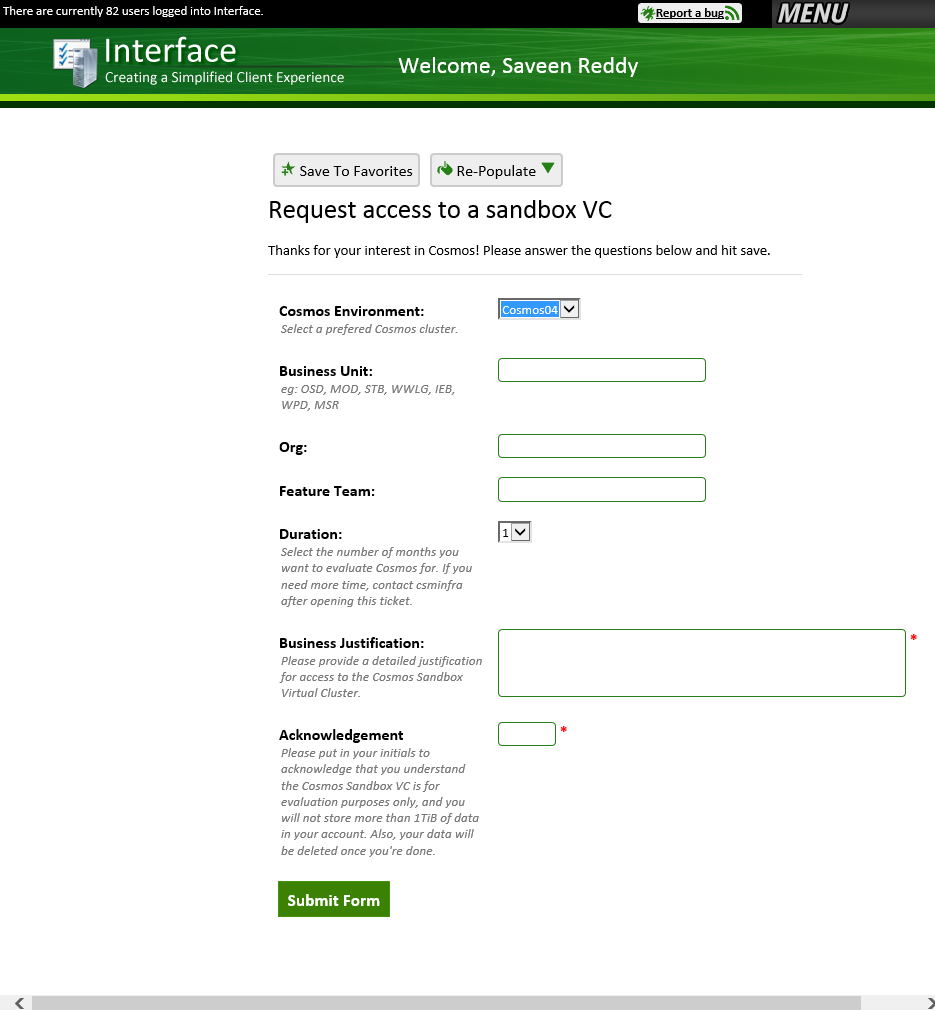
## Please Read This

I refer to the SandBox VC in this document because it's convenient for the tutorial. However, you can use any VC you want. In fact, if your team already has access to a VC then go ahead and use that VC.

## Getting Access to the Sandbox VC

Go here: <http://aka.ms/CosmosSandbox>

It will take you to a very simple interface to request access as shown below:



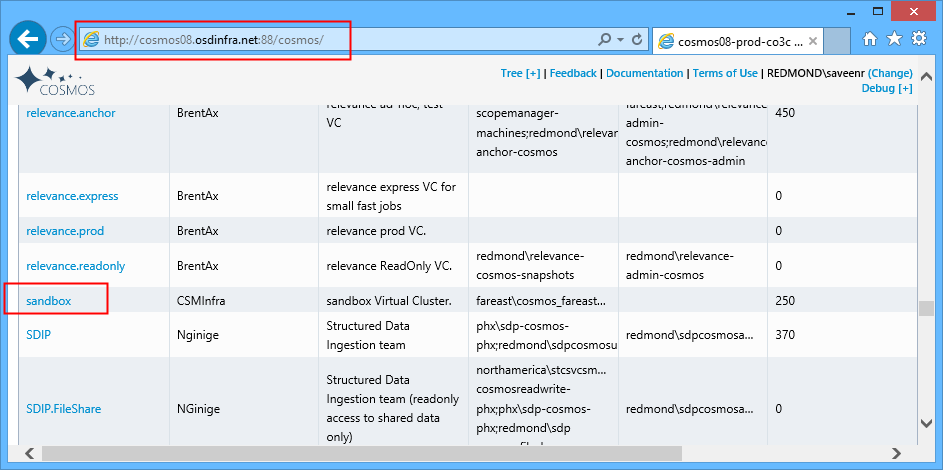
The Cosmos Portal

The Cosmos Portal is how we will upload/download streams and monitor the progress of jobs. We will cover the basic mechanics in this tutorial but if you want to learn more consult the Cosmos Portal User Guide at this location: <http://aka.ms/CosmosPortalUserGuide>

Accessing the Sandbox

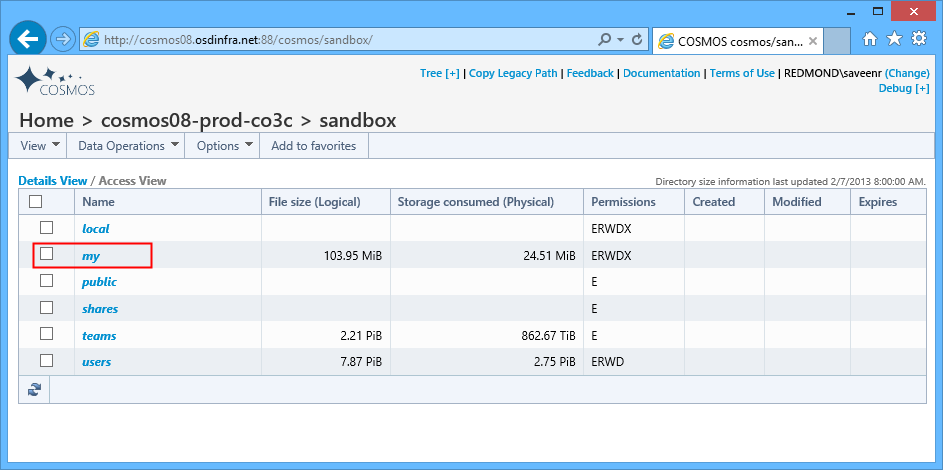
First open the link to the Cosmos08 cluster: <http://cosmos08.osdinfra.net:88/cosmos/>

On the left is the list of virtual clusters. Scroll down and click the link to "sandbox."



Now you are viewing the sandbox VC on Cosmos08.

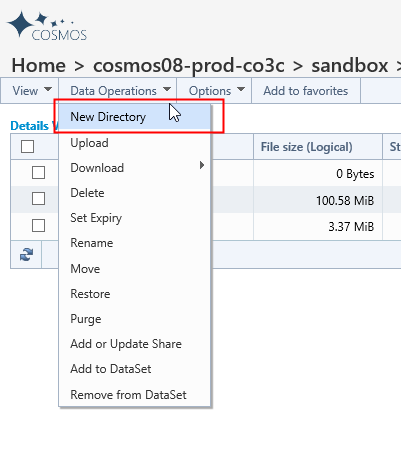
As a sandbox user you have a **my** folder that is reserved for you.

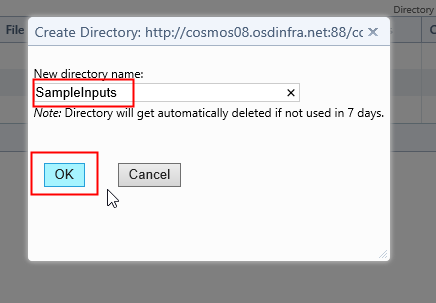


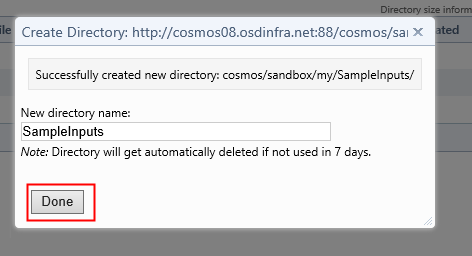
## Uploading a Stream

We will use the **my** folder for our tutorial. Click on **my** folder.

Now create a folder called "SampleInputs" via **Data Operations > New Folder.**

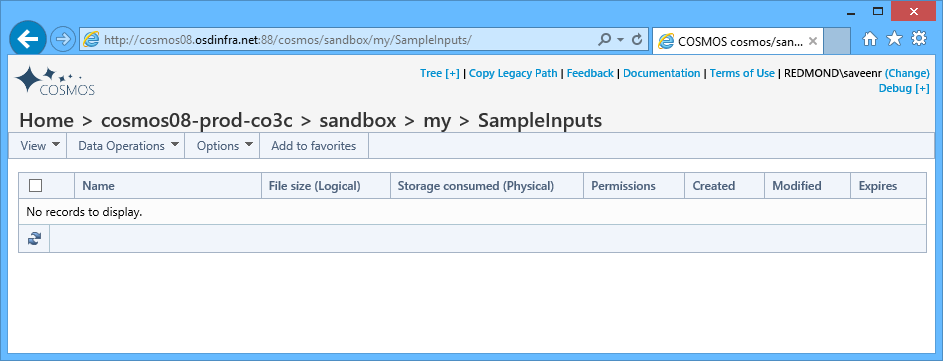




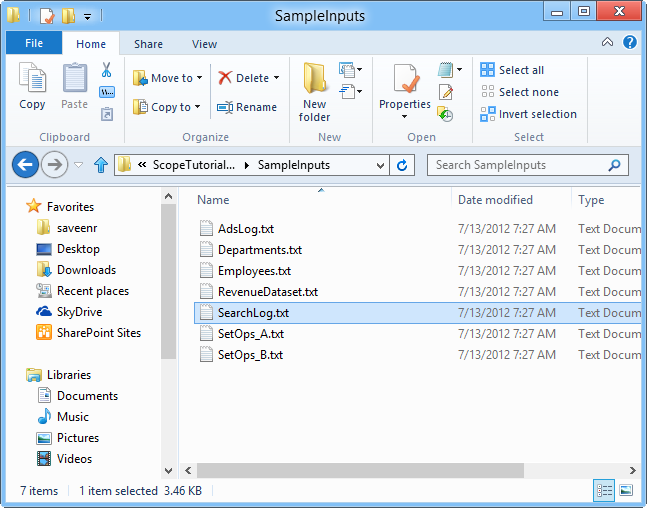


Navigate to the **SampleInputs** folder.

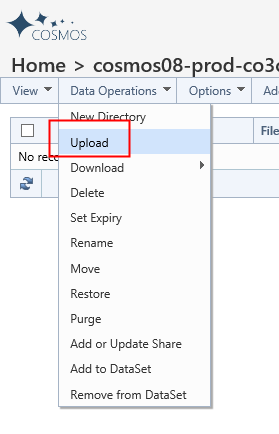
You can see there is nothing in it.



We will upload a stream that comes in the Scope Tutorial Samples ZIP file. Specifically we will upload the SearchLog.txt file shown below:

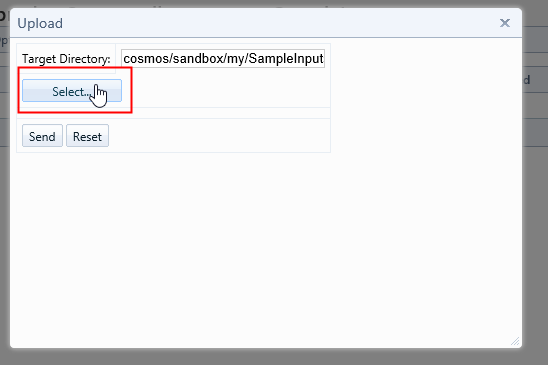


In the **SampleInputs** folder, Select **Data Operations > Upload.**

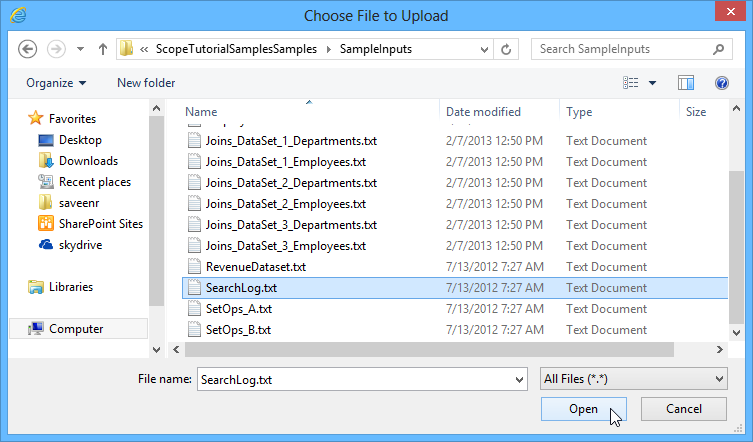


The following UI will come up.

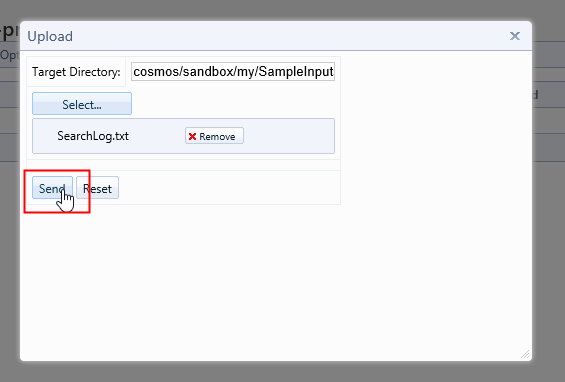
Click **Select**.



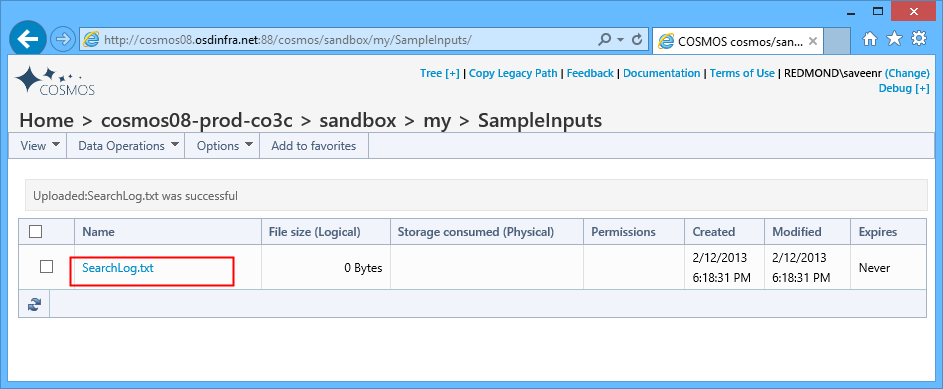
Then pick the SearchLog.txt file.



Then click **Send**.

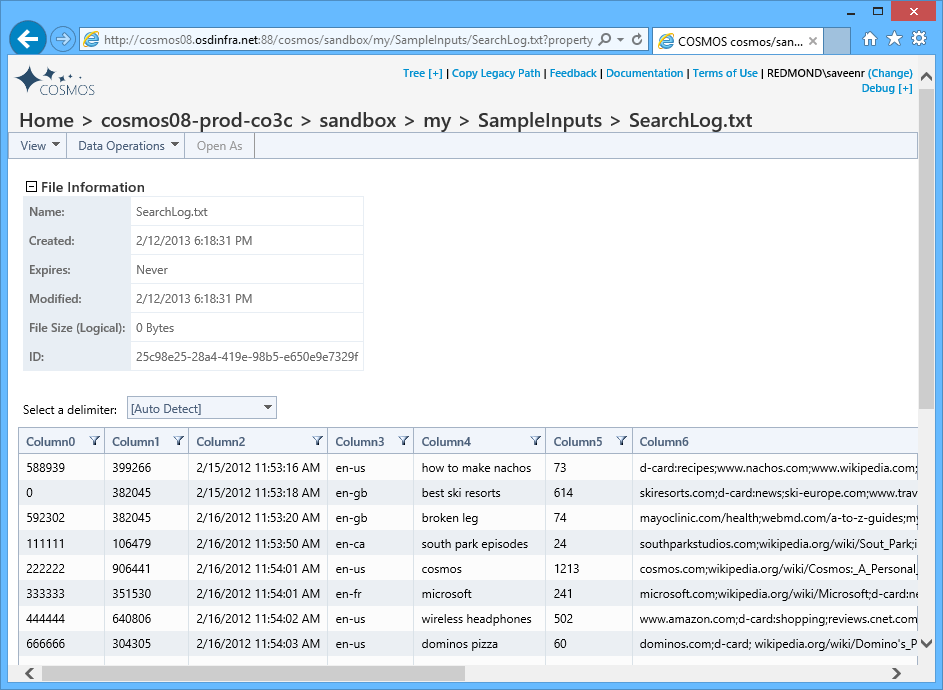


This operation could take a while so just wait. Once done you'll see this:



Notice that the logical file size is "0 Bytes". Don't worry, at some point it'll update and be correct.

If you click on the file you will see a preview.



## The Script

We'll choose a very simple script shown below:

searchlog =

EXTRACT IId:int, UId:int, Start:DateTime, Market:string, Query:string, DwellTime:int, Results:string, ClickedUrls:string

FROM @"my/SampleInputs/SearchLog.txt"

USING DefaultTextExtractor();

rs1 = SELECT Market, SUM(DwellTime) AS TotalDwellTime

FROM searchlog

GROUP BY Market;

OUTPUT rs1

TO @"/my/ScopeTutorial/SampleOutputs/output.txt";

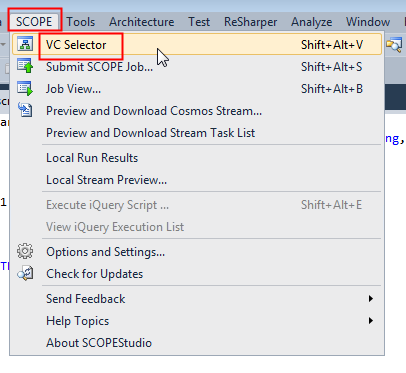
NOTE that the path separator character is "/" and not "\".

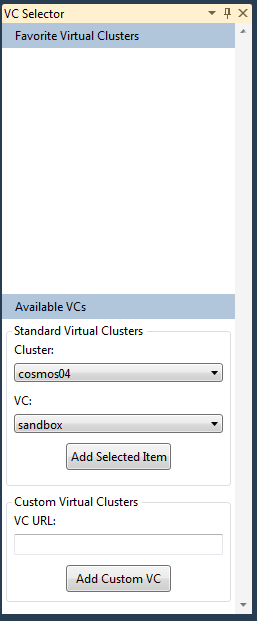
Enter this script into Scope Studio.

## Setting up the Virtual Cluster

Before we can submit a job we have to configure VCs in Scope Studio.

First bring up the VC Selector

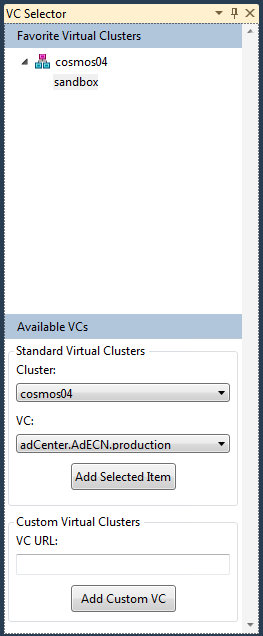




Notice that the Favorite Virtual Clusters box is empty. We will need to populate it.

Go to Available VCs. Under Cluster select "cosmos08" and then under VC select "sandbox". Then click **Add Selected Item**.

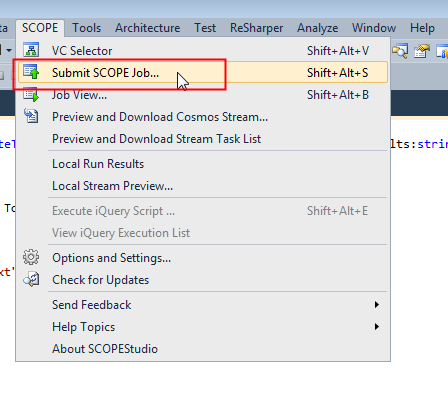
Now you will see sandbox available in **Favorite Virtual Clusters**.



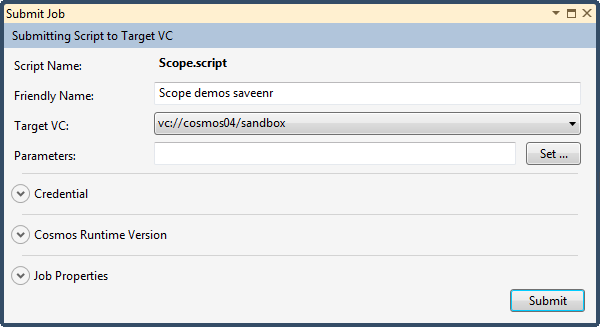
We don't need the VC Selector anymore so you can dismiss it.

## Submitting the Job

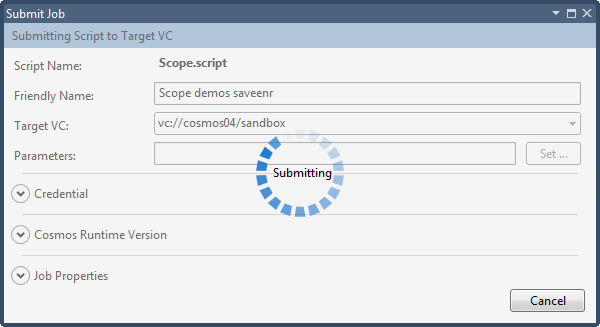
To submit the Job use **Scope > Submit Scope Job.**



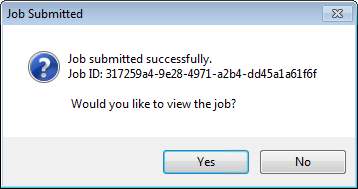
The **Submit Job** UI will come up. Give the job a friendly name and then click **Submit**.



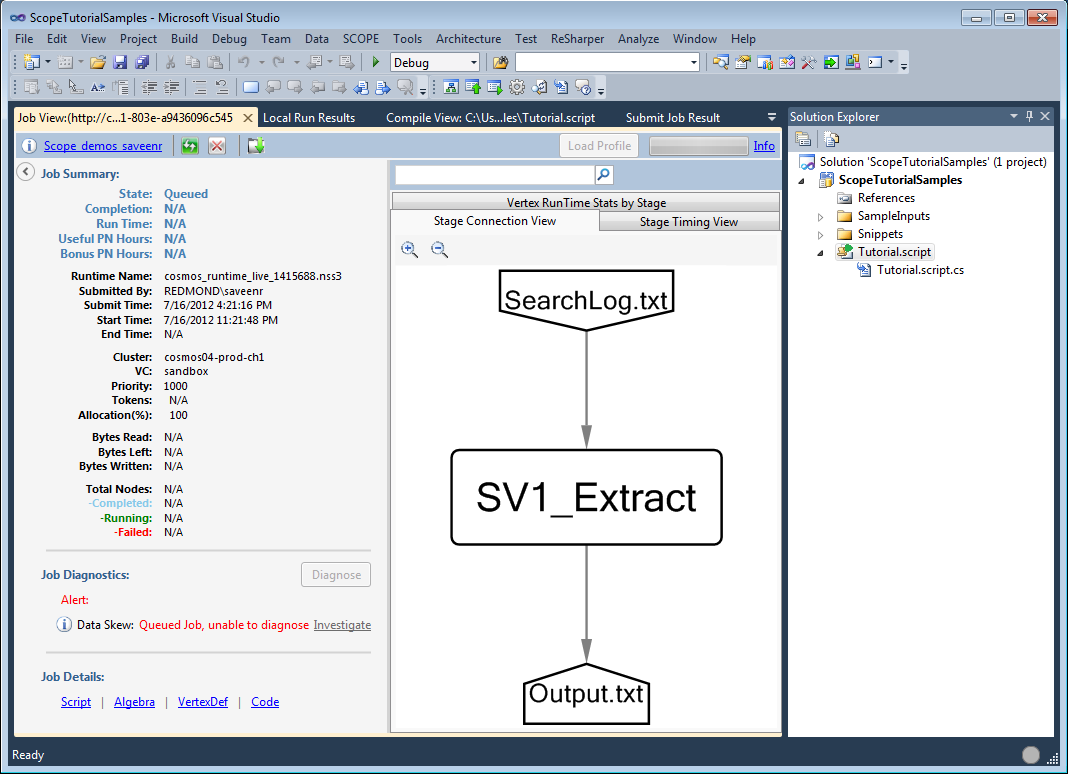
During this submission process you'll see it wait as shown below:



And then when the submission has completed you will see this:



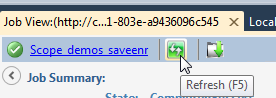
Click **Yes** and you'll see this:



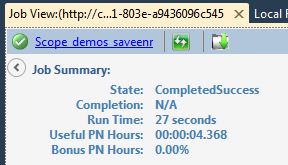
Notice that the Job is queued.



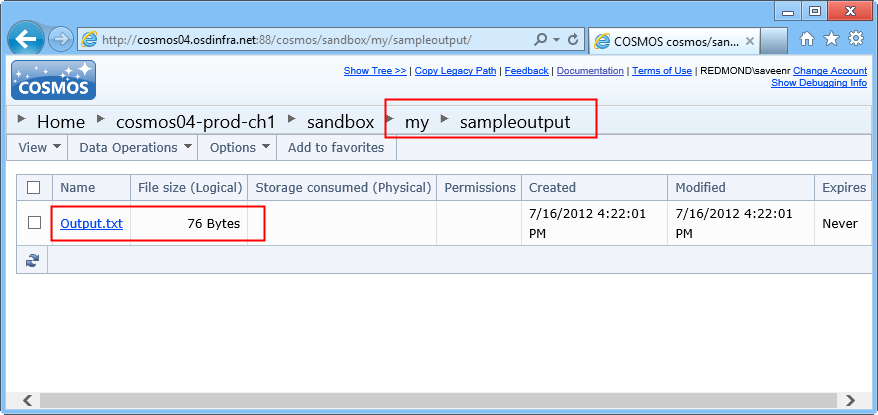
Click the Refresh button…



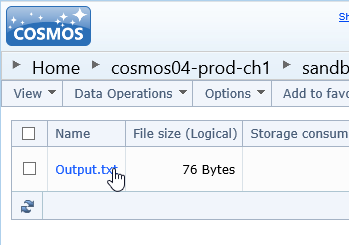
…and you'll eventually see the job was completed successfully.



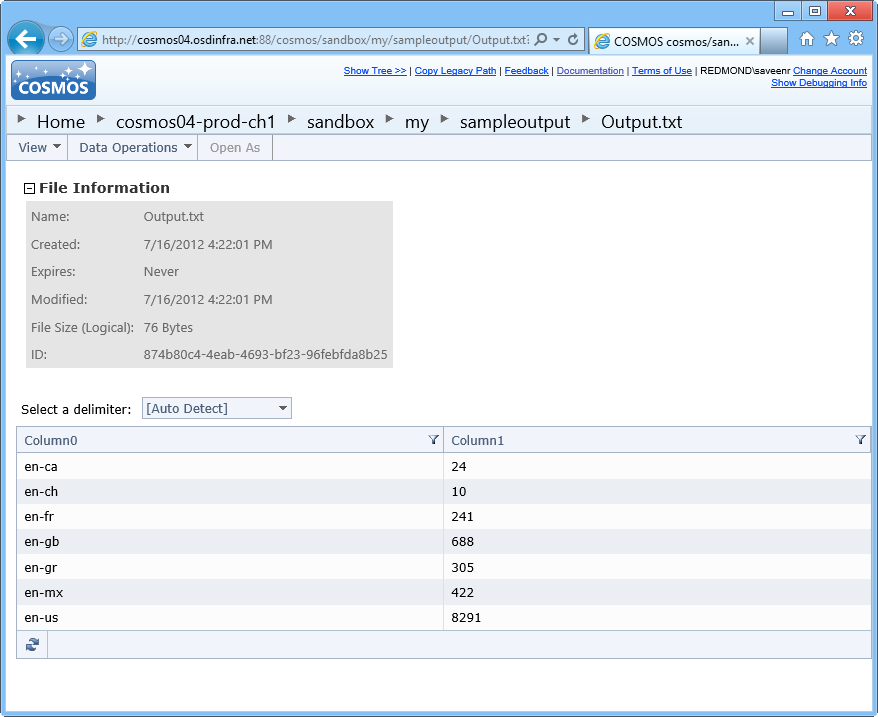
To see the output go back to the ***/my*** folder and navigate to sampleoutput. Here you will see the stream.



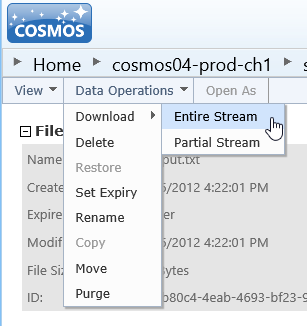
Now click on the output.txt link.



And now you'll see details for the stream and even a preview of some content.



To download the stream, see below:



## Monitoring Jobs

If you want to monitor jobs, go to your my/jobs folder and you can see a list of the jobs you have recently submitted.

