Running MOVES in the Cloud

EPA does not endorse the products or services of commercial companies. Any reference to a specific commercial product or service by trade name, trademark, manufacturer, company, or otherwise does not constitute or imply the endorsement or recommendation of the U.S. Environmental Protection Agency.

William Aikman
Wes Faler

Air Quality and Modeling Center
Office of Transportation and Air Quality
U.S. Environmental Protection Agency

MOVES

EPA Rule Making and Air Quality MOVES Requirements

- **The Challenge:**
  - Run MOVES for 3200+ US counties for rates (rate per vehicle, rate per distance, rate per profile)
  - For seven, slightly different scenarios
- **Running MOVES for rates for one county for 2 months (January and July) takes 50 – 80 hours, depending on county temperature extremes, and requires 160 to 240 individual MOVES runs**
- **The math, if done serially**
  - 3200 (counties) * 65 hrs (average) * 7 Scenarios= 1,456,000 hours, or 60,666.67 days, 166.2 years

MOVES

The Solution: Step 1: "Representative" Countries

- **Representative Counties, methodology**
  - Resulted in 103 Representative counties
- **Adding runs for January and July, resulted in 206 county/month combinations**
- **Given the SMOKE/MOVES temperature profiles, each county/month required 50 – 120 individual MOVES runs**
- **For each scenario, 16,604 MOVES Runs were required**
- **In early testing, processing 1 county/month on average took about 30 hours**

The Solution: Step 2: To the Cloud!

- If only we could launch 206 computers on demand, then we could do **ALL** 206 county/month combinations in parallel in about 30 hours

MOVES in the Cloud

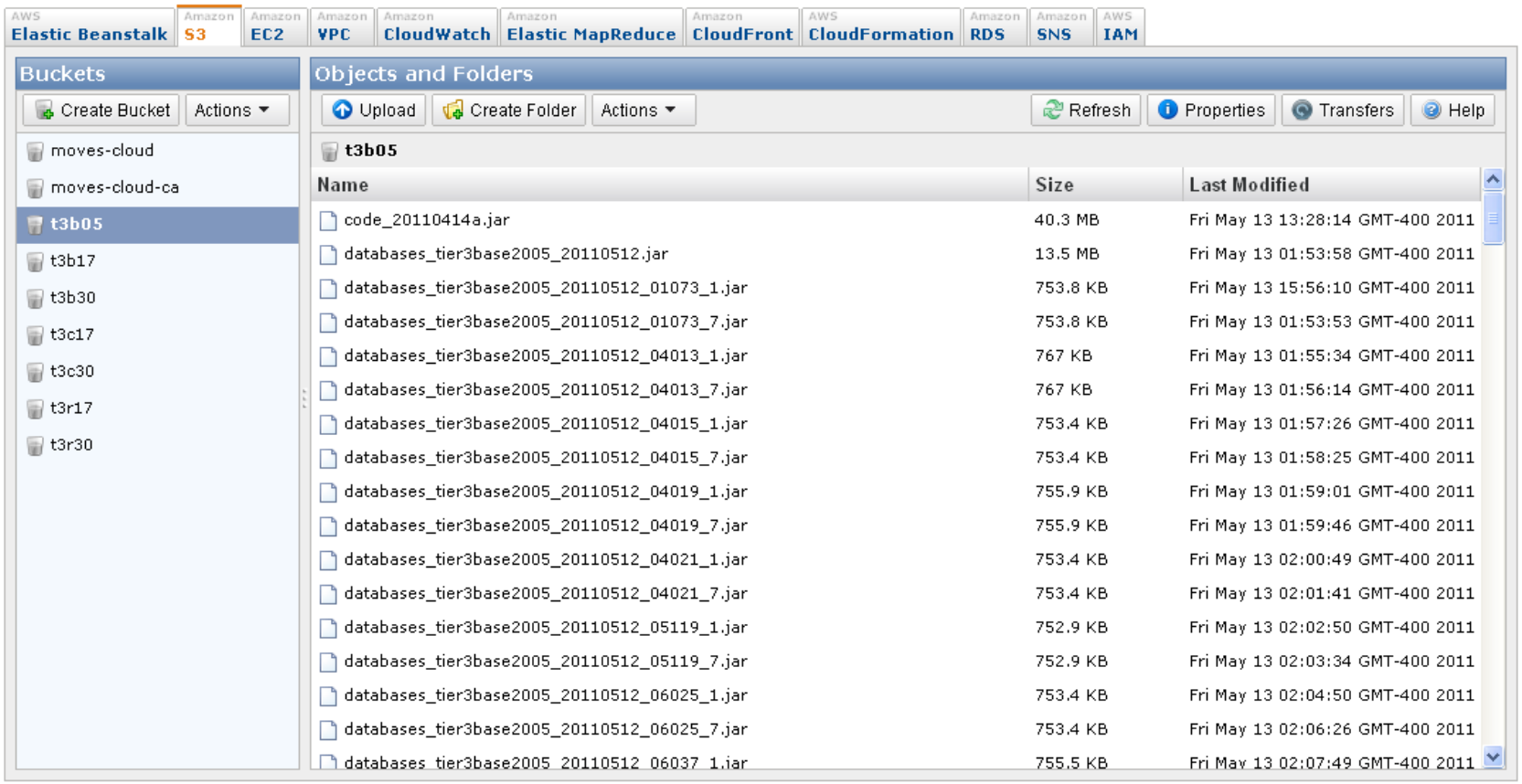
- **Through an EPA contractor and based upon their recommendation, we established an account at Amazon Web Services**
- **Based upon the SMOKE/MOVES requirements, we developed a series of scripts that would allow us to launch 206 instances (virtual computers) in parallel to**
  - process all the MOVES runs required for a county/month and then to process
  - another 206 instances to “post process” MOVES output for input to EPA Air Quality modeling systems

MOVES

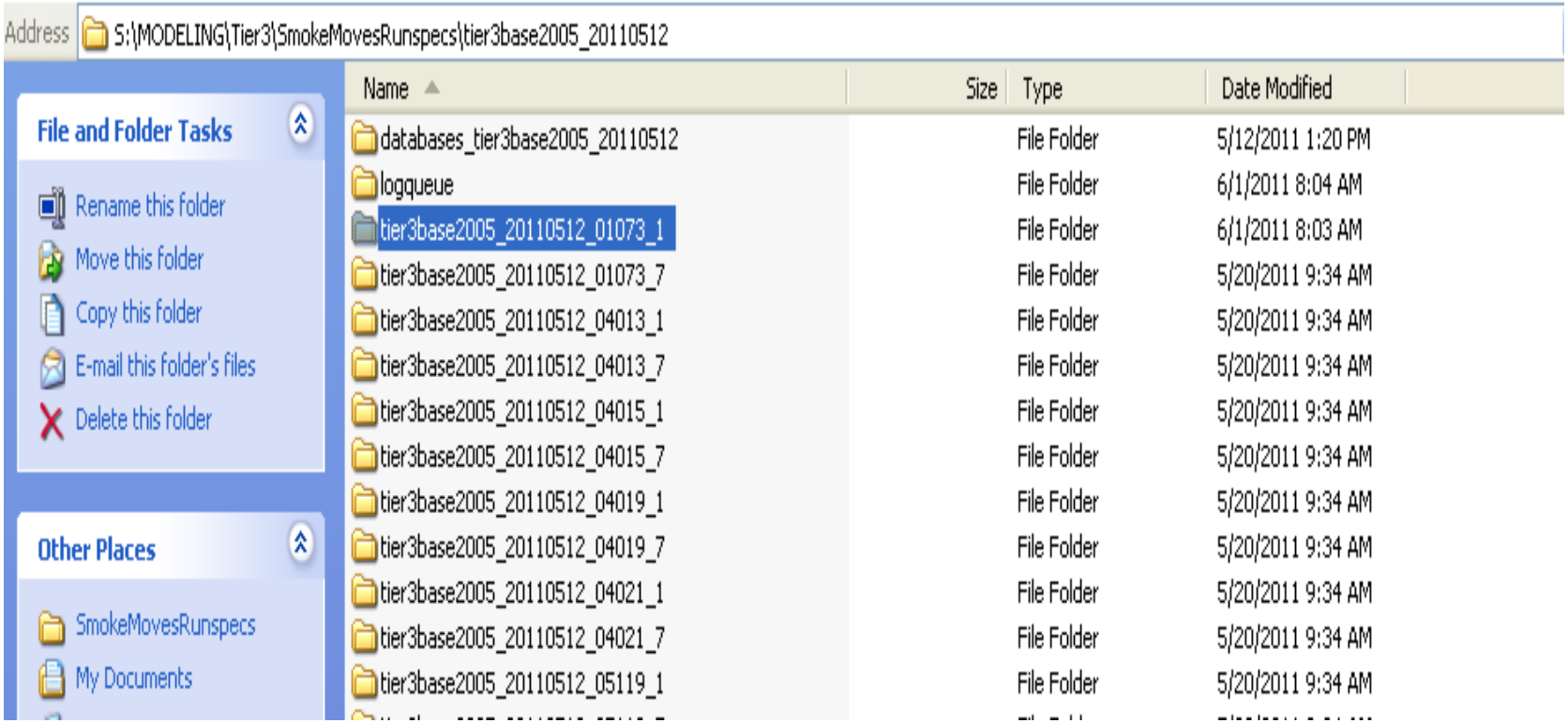
MOVES Processing in the Cloud

- **Data organization and management**
  - Scenario (seven in total)
  - Batch (county/month, 206 in total)
  - Jobs (individual runspecs, 16,604 in total)
- **All MOVES code, default data bases, and runspecs would reside in Amazon "buckets" (1 bucket per scenario)**

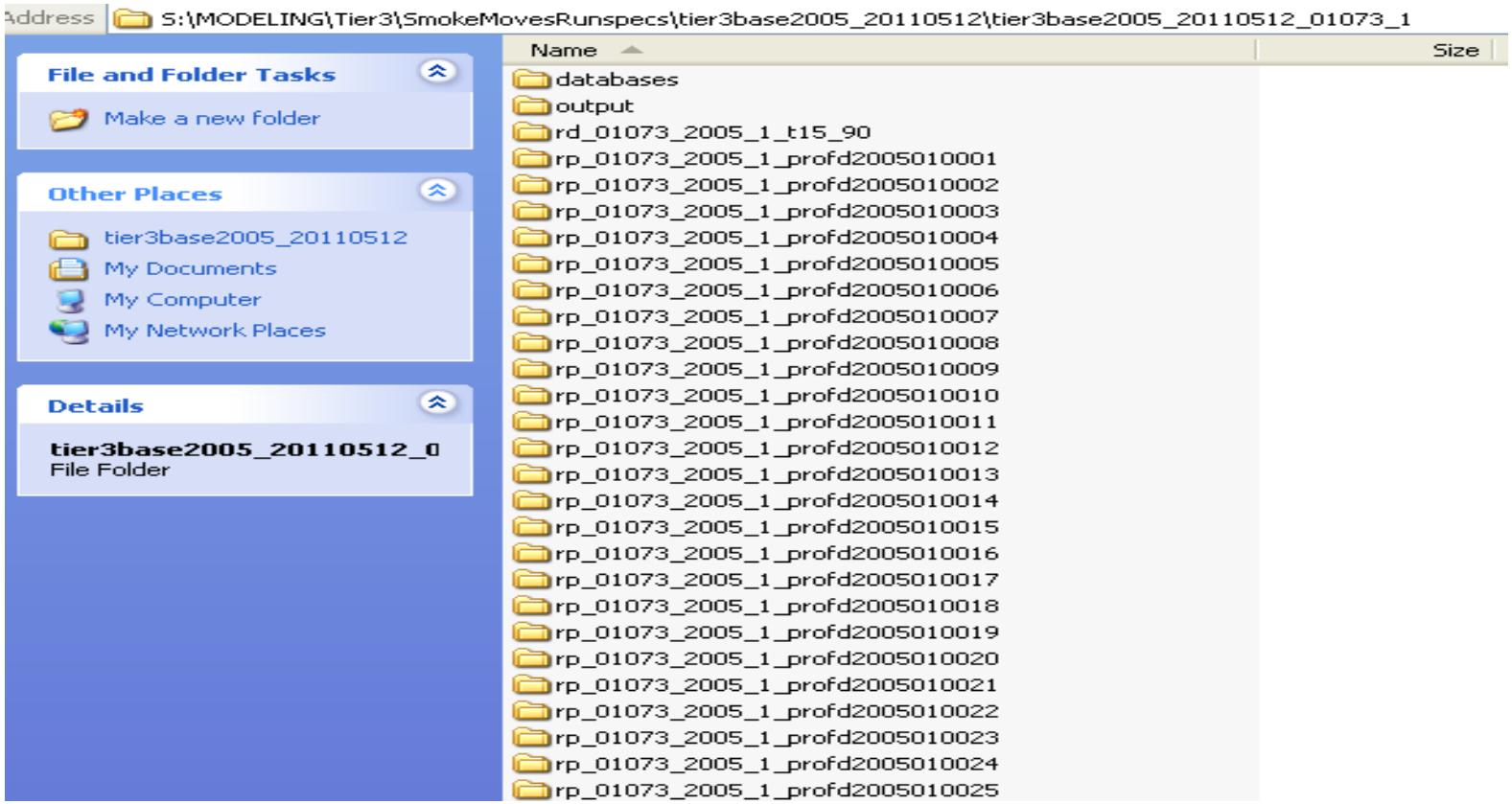
MOVES Processing in the Cloud – S3



MOVES Processing in the Cloud – local store



MOVES Processing in the Cloud – local store



MOVES Processing in the Cloud

- Operating system is standard Amazon provided Linux (Centos)
- Instance management done by Amazon Elastic Compute Cloud (EC2)
- Storage done by Amazon Simple Storage Service (S3)
- Queueing done by Amazon's Simple Queue Service (SQS)

Amazon Processing

- **Scripts were developed to**
  - Upload code, databases, batches to a “bucket”
  - Establish queues, 1 set per scenario (1 queue for jobs and 1 queue for stats)
  - Add batches (county/month) to a queue
  - Download results to local shared drive
  - Check status of each batch
  - Re-add jobs for incompletely processed batches
- **An Amazon instance will process all jobs in a batch and then quit**

Amazon Processing

- **Scripts are .bat files and are run from a local directory and the command line on a local EPA computer**
- **The scripts**
  - manipulate data and services at Amazon and
  - manage the transfer of data between Amazon and the local EPA file store

Amazon Processing

S:\MODELING\Tier3\SmokeMovesRunspecs\PopulateAmazonDataStructure

and Folder Tasks

Make a new folder

er Places

SmokeMovesRunspecs
My Documents
My Computer
My Network Places

ails

ulateAmazonDataStruct
Folder

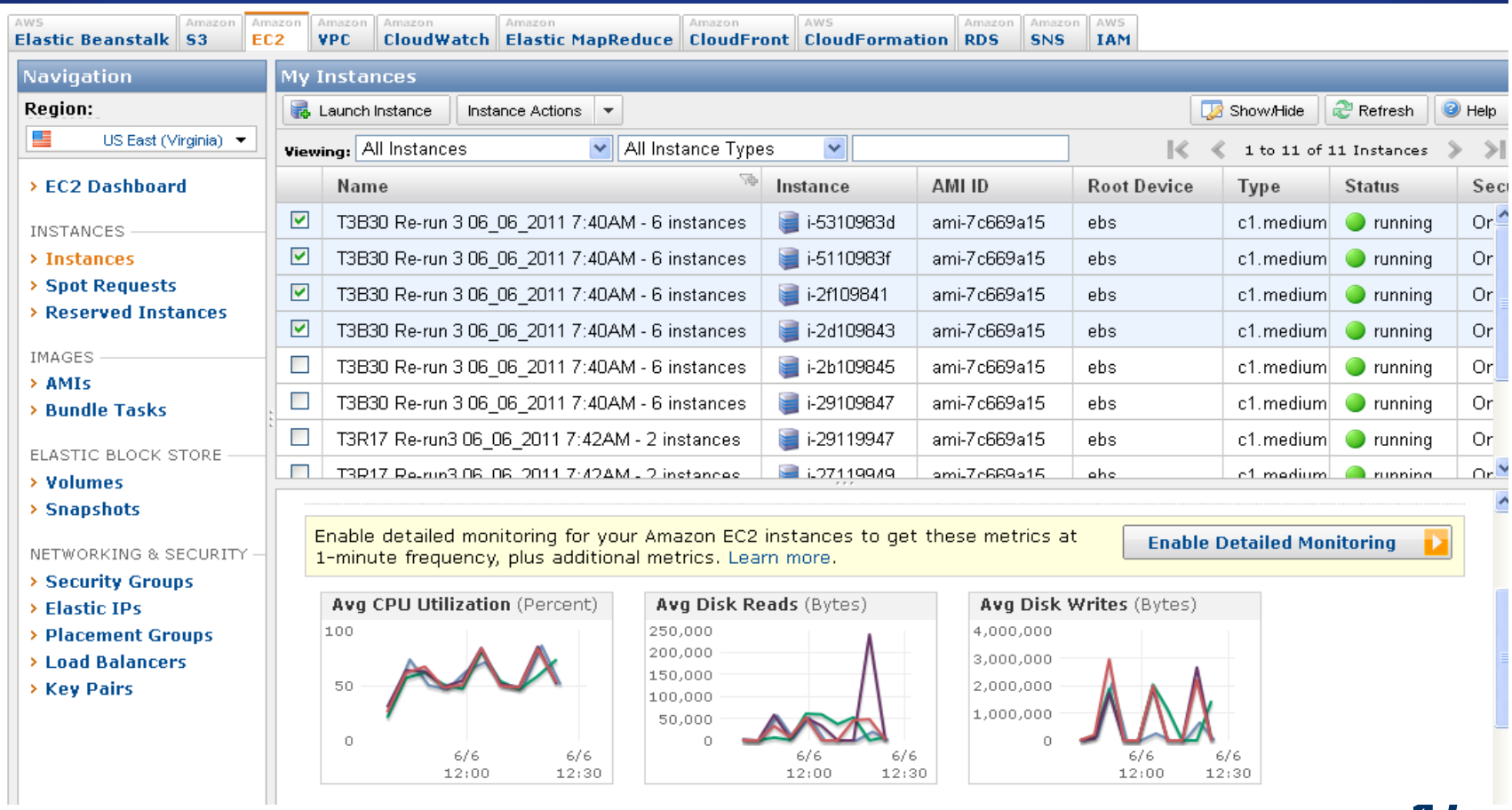
Name	Size	Type
TIER3LOWE2030_20110520.txt	1 KB	Text Document
tier3lowe2030_20110520_addjobs.bat	77 KB	MS-DOS Batch File
tier3lowe2030_20110520_addjobs.log	421 KB	Text Document
tier3lowe2030_20110520_addpostprocess.bat	72 KB	MS-DOS Batch File
tier3lowe2030_20110520_addpostprocess.log	379 KB	Text Document
tier3lowe2030_20110520_batchstatus.bat	77 KB	MS-DOS Batch File
tier3lowe2030_20110520_CreateQueues.bat	1 KB	MS-DOS Batch File
tier3lowe2030_20110520_deleteQueues.bat	1 KB	MS-DOS Batch File
tier3lowe2030_20110520_downloaddbresults.bat	55 KB	MS-DOS Batch File
tier3lowe2030_20110520_downloadpostresults.bat	56 KB	MS-DOS Batch File
tier3lowe2030_20110520_downloadpostresults.bat.bak	56 KB	BAK File
tier3lowe2030_20110520_downloadpostresults.log	324 KB	Text Document
tier3lowe2030_20110520_downloadresults.bat	54 KB	MS-DOS Batch File
tier3lowe2030_20110520_downloadresults.bat.bak	54 KB	BAK File
tier3lowe2030_20110520_downloadresults.log	39,186 KB	Text Document
tier3lowe2030_20110520_flushQueues.bat	1 KB	MS-DOS Batch File
tier3lowe2030_20110520_getstatus_MOVES.bat	1 KB	MS-DOS Batch File
tier3lowe2030_20110520_getstatus_Post.bat	1 KB	MS-DOS Batch File
tier3lowe2030_20110520_jarjobs.bat	48 KB	MS-DOS Batch File
tier3lowe2030_20110520_poststatus.bat	1 KB	MS-DOS Batch File
tier3lowe2030_20110520_poststatus.txt	18 KB	Text Document
tier3lowe2030_20110520_readdjobs.bat	77 KB	MS-DOS Batch File
tier3lowe2030_20110520_readdjobs.log	215 KB	Text Document
tier3lowe2030_20110520_readdpostprocess.bat	72 KB	MS-DOS Batch File
tier3lowe2030_20110520_readdpostprocess.bat.bak	72 KB	BAK File
tier3lowe2030_20110520_readdpostprocess.log	128 KB	Text Document
tier3lowe2030_20110520_TextToStartInstances-MOVES.txt	1 KB	Text Document
tier3lowe2030_20110520_TextToStartInstances-post.txt	1 KB	Text Document
tier3lowe2030_20110520_uploadjobs.bat	52 KB	MS-DOS Batch File
tier3lowe2030_20110520_uploadjobs.log	11,341 KB	Text Document

Amazon Processing

- The Amazon AWS console (browser based) is used to initiate instances and monitor instance status

MOVES

Amazon Processing



Experiences to date

" …Thy fate is the common fate of all, **Into each life a little rain must fall**, Some days must be dark and dreary"

- Longfellow

The week we ran our first large scale scenario

NY Times – April 21, 2011
Amazon Cloud Failure Takes Down Web Sites
By CLAIRE CAIN MILLER **10:28 a.m. | Updated** *to reflect status of the problem on Friday.*

A widespread failure in Amazon.com's Web services business was still affecting many Internet sites on Friday morning, highlighting the risks involved when companies rely on so-called cloud computing.

(Note that our presence was purely coincidental!!!)

Then, Attack of the "Zombies"

- We experience a 10 – 20 percent instance failure rate when launching instances
- Instances stop responding, CPU goes to zero
  - but not "dead", still "running" and still incurring charges
- Necessitates manually terminating the instance, re-queueing incomplete batches, and launching new instances
- To complete all batches, 3 to 4 re-runs have been required, resulting in 2-3 times the elapsed time 60-90 clock hrs) vs originally envisioned (30 hrs)
- Amazon very concerned and working with us to find the cause of "zombie"instances

Conclusions to-date

- Keys to effectiveness are data architecture and easy-to-use scripts for management of cloud processes
- Is attractive versus acquiring local hardware and incurring corresponding support costs
- Cloud computing an extremely cost effective (and only the feasible) way to process large scale MOVES runs – parallelism is key!
- Reliability issues will be resolved ("zombies" will eventually be defeated)

Conclusions to-date

- In the cloud, there is a cost vs "clock time" tradeoff
- Faster instances may yield shorter elapsed times, However, they are more expensive.

  Hi-CPU, On-Demand Instances
  Medium $0.17 per hour
  Extra Large $0.68 per hour$
  Cluster Compute Instances
  Quadruple Extra Large $1.60 per hour
  GPU Instances Quadruple Extra Large $2.10 per hour

- Each user must consider cost/time tradeoff on their own
- Note: application may not take advantage of enhanced instance capabilities

The Future

- **Future plans**
  - More research, debugging/testing/refinement is required
  - EPA plans to eventually make scripts, MOVES version(s) and default databases “public” in the “cloud”
- **Feedback on interest in a cloud version of MOVES would be welcome**