Using Cloud Computing to Do Large Numbers of MOVES Runs
20th International Emission Inventory Conference
Tampa, Florida, 15 August 2012

Wesley Faler
Fluid & Reason, LLC

Harvey Michaels and William Aikman
U.S. Environmental Protection Agency, OTAQ
Disclaimer

• 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. The content provided represents the work of the authors and does not reflect the policy, guidance, or procedures recommended by the U.S. EPA. This document is disseminated in the interest of information exchange and the U.S. Government assumes no liability for use of the information.
Outline

• SMOKE-MOVES
• Our cloud approach
• Instance structure
• Data flow
• Cloud terminology
• Organizing runspecs and databases
• Amazon tips
• Run timing and statistics
• Summary and conclusions
SMOKE and Emissions for Air Quality Modeling

Pollutant emissions from a variety of sources in a variety of formats in a variety of temporal and spatial resolutions

SMOKE

Gridded, hourly emissions ready for air quality model

AQ Model

Source: Zubrow & Baek
MOVES Workshop, June 2011
SMOKE-MOVES Part 1

- 3109 Counties x 12 months = 37,308 county-months
  
  Group counties by IM, fuels, age distribution
  
- 146 county groups x 2 fuel-months = 292 county-months
  
  Runspect Generator*
  
  Gridded, hourly meteorology for air quality modeling
  
  Met4moves*
  
  Met input for Runspect generator
  
  27,513 Runspects
  27,513 ZoneMonthHour tables

*SMOKE-MOVES Script
SMOKE-MOVES Part 2

- **27,513 Runs specs**
  - 27,513 ZoneMonthHour tables
- **Run MOVES in the Cloud**
  - 292 batches (~100 runs specs/batch)
- **MOVES Rate Tables**
- **Run moves2smk* in the Cloud**
- **SMOKE Rate Tables**
- **Movesmrg***
  - Gridded, hourly emissions for air quality modeling
- **Gridded, hourly meteorology for air quality modeling**
- **County VMT, speed, and Vehicle Population by month**

*SMOKE-MOVES Script*
What is the Cloud?

• “The Cloud” refers to computers that we access via the internet. It includes Storage, Instances, and Queues.

• Storage
  – The Bucket, a common storage area for the user and the Instances.
  – The Image, a virtual computer built to our specifications. It contains framework software (Java and scripts) and the correct versions of Java and MySQL.

• Instances, as many as we need, are copies of the Image.

• Queues
  – Job queues tell the instances which jobs to do.
  – Status queues hold diagnostic messages for us.
Simplest Cloud Approach

1 Computer

1 Master
1 Worker
1 Run
Our Cloud Approach

1 Computer

1 Master

1 Worker
Our Cloud Approach
Cloud Instance Structure
Cloud Machine Instances

Amazon EC2 AMI

- Files
- Commands
- Result Files
- Instances
- Messages
Cloud Queues

- Files
- Instances
- Messages
  - Commands
  - Result Files
  - Amazon SQS Queue
  - AWS SQS Queue

15
Organizing Runspecs and Databases

[scenario]/
databases_[scenario]/
[anydatabase]/
[scenario]_[batch]/
databases/
[job]/
databases/
[job].mrs
output/
Amazon Tips - Security

Access Credentials

There are three types of access credentials used to authenticate your requests to AWS services: (a) access keys, (b) X.509 certificates, and (c) key pairs. Each access credential type is explained below.

Use access keys to make secure REST or Query protocol requests to any AWS service API. We create one for you when your account is created — see your access key below.

Your Access Keys

<table>
<thead>
<tr>
<th>Created</th>
<th>Access Key ID</th>
<th>Secret Access Key</th>
<th>Status</th>
</tr>
</thead>
<tbody>
<tr>
<td>March 2, 2011</td>
<td></td>
<td>Show</td>
<td>Active  (Make Inactive)</td>
</tr>
</tbody>
</table>

Create a new Access Key

For your protection, you should never share your secret access key. Best practice recommends frequent key rotation.

Learn more about Access Keys
Amazon Tips - Instances

Request Instances Wizard

Choose an Amazon Machine Image (AMI) from one of the tabbed lists below by clicking its Select button.

- Quick Start
- My AMIs
- Community AMIs

Viewing: Private Images

<table>
<thead>
<tr>
<th>AMI ID</th>
<th>Root Device</th>
<th>Name</th>
<th>Platform</th>
<th>Stars</th>
</tr>
</thead>
<tbody>
<tr>
<td>ami-a2190bcb</td>
<td>obs</td>
<td>607601799279/us-ops-moves-ami-20110317a</td>
<td>Other Linux</td>
<td>🟢</td>
</tr>
</tbody>
</table>

Free tier eligible if used with a micro instance. See AWS free tier for complete details and terms.
Amazon Tips - Instances

Request Instances Wizard

CHOOSE AN AMI  INSTANCE DETAILS  CREATE KEY PAIR  CONFIGURE FIREWALL  REVIEW

Provide the details for your instance(s). You may also decide whether you want to launch your instances as "on-demand" or "spot" instances.

Number of Instances: 1  Availability Zone: No Preference

Instance Type: High-CPU Medium (c1.medium, 1.7 GB)

Launch Instances

EC2 Instances let you pay for compute capacity by the hour with no long term commitments. This transforms what are commonly large fixed costs into much smaller variable costs.

Request Spot Instances

LaunchInstances Into Your Virtual Private Cloud

< Back

Continue
Amazon Tips - Instances

Request Instances Wizard

Number of Instances: 1
Availability Zone: No Preference

Advanced Instance Options
Here you can choose a specific kernel or RAM disk to use with your instances. You can also choose to enable CloudWatch Detailed Monitoring or enter data that will be available from your instances once they launch.

Kernel ID: Use Default
RAM Disk ID: Use Default

Monitoring:
- Enable CloudWatch detailed monitoring for this instance (additional charges will apply)

User Data:
- AUTO_TERMINATE=1
- ACCESSKEY=*
- SECRETKEY=*
- STATUSQUEUE=
- JOBQUEUE=
- base64 encoded

Termination Protection:
- Prevention against accidental termination.

Shutdown Behavior:
- Choose the behavior when the instance is shutdown from within the instance.

Back Continue
Amazon Tips - Instances

Public/private key pairs allow you to securely connect to your instance after it launches. To create a key pair, enter a name and click Create & Download your Key Pair. You will then be prompted to save the private key to your computer. Note, you only need to generate a key pair once - not each time you want to deploy an Amazon EC2 instance.

Choose from your existing Key Pairs

Your existing Key Pairs*: EpaCloudKey1

Create a new Key Pair

Proceed without a Key Pair

Continue
Amazon Tips - Instances

Request Instances Wizard

Security groups determine whether a network port is open or blocked on your instances. You may use an existing security group, or we can help you create a new security group to allow access to your instances using the suggested ports below. Add additional ports now or update your security group anytime using the Security Groups page.

Choose one or more of your existing Security Groups

- sg-ba8419d3 - Only SSH
- sg-5afa6d33 - default

(Selected groups: sg-ba8419d3)

Create a new Security Group

< Back

Continue
Example Runs

<table>
<thead>
<tr>
<th></th>
<th>Scenario</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>2020</td>
</tr>
<tr>
<td>Batches</td>
<td>292</td>
</tr>
<tr>
<td>Jobs</td>
<td>28,377</td>
</tr>
</tbody>
</table>

These two runs were done together at the same time and provide some measure of the variability in timing.
## Timing

<table>
<thead>
<tr>
<th>Process</th>
<th>Time (hours)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>2020</td>
</tr>
<tr>
<td>Loadzmh.plx</td>
<td>2.75</td>
</tr>
<tr>
<td>CreateAndPopulate.plx</td>
<td>5.98</td>
</tr>
<tr>
<td>Uploadjobs</td>
<td>5.28</td>
</tr>
<tr>
<td>Addjobs</td>
<td>0.71</td>
</tr>
<tr>
<td>Shortest Batch (42 jobs)</td>
<td></td>
</tr>
<tr>
<td>Longest Batch (167 jobs)</td>
<td></td>
</tr>
<tr>
<td>Downloadresults</td>
<td>13.76</td>
</tr>
<tr>
<td>Downloadpostresults</td>
<td>8.82</td>
</tr>
<tr>
<td>Batchstatus</td>
<td>0.15</td>
</tr>
</tbody>
</table>
Bottom Line

$0.17/computer/hour
* 32 average hours/batch
* 292 batches
* 1 batch/computer
$1,600 for the run

With restarts, and repeats of failures, the run took about a week, including data transfer time not included above. On a single computer, it would have taken more than a year.
Conclusions

• Cloud computing has been successfully employed to generate MOVES inventories and the lookup tables needed by SMOKE-MOVES.

• Cloud computing provides cheap, abundant computing resources on demand.

• We are able to specify the configuration we want: operating system, memory, storage, and installed software.

• We don’t need to buy, install, maintain, repair, or upgrade the computers.
Thank you!

Wesley Faler
Fluid & Reason, LLC
3319 Greenfield Rd #352, Dearborn, MI 48120
wes.faler@gmail.com

Harvey Michaels and William Aikman
U.S. Environmental Protection Agency, OTAQ
2565 Plymouth Rd, Ann Arbor, MI 48105
michaels.harvey@epa.gov, aikman.william@epa.gov

http://www.epa.gov/otaq/models/moves/

Background image by Chris Waits
http://www.flickr.com/photos/chriswaits/6109678426/