

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

The logo for MOVES, featuring the word 'MOVES' in a bold, metallic, three-dimensional font with a glowing effect. The letters are set against a dark, rectangular background with a light gradient.

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

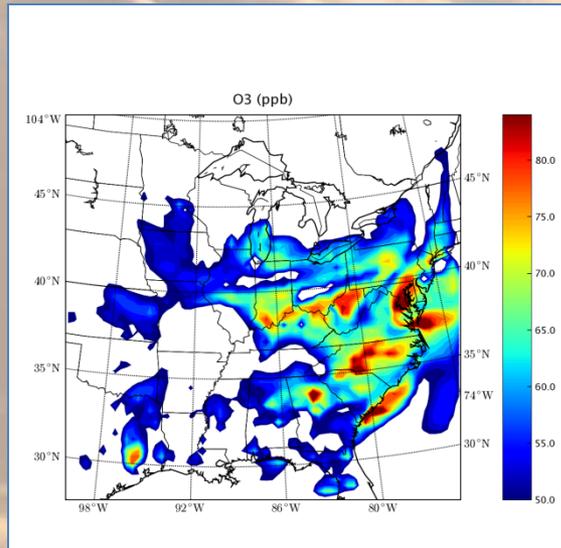
- SMOKE-MOVES
- Our cloud approach
- Instance structure
- Data flow
- Cloud terminology
- Organizing runspecs and databases
- Amazon tips
- Run timing and statistics
- Summary and conclusions

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

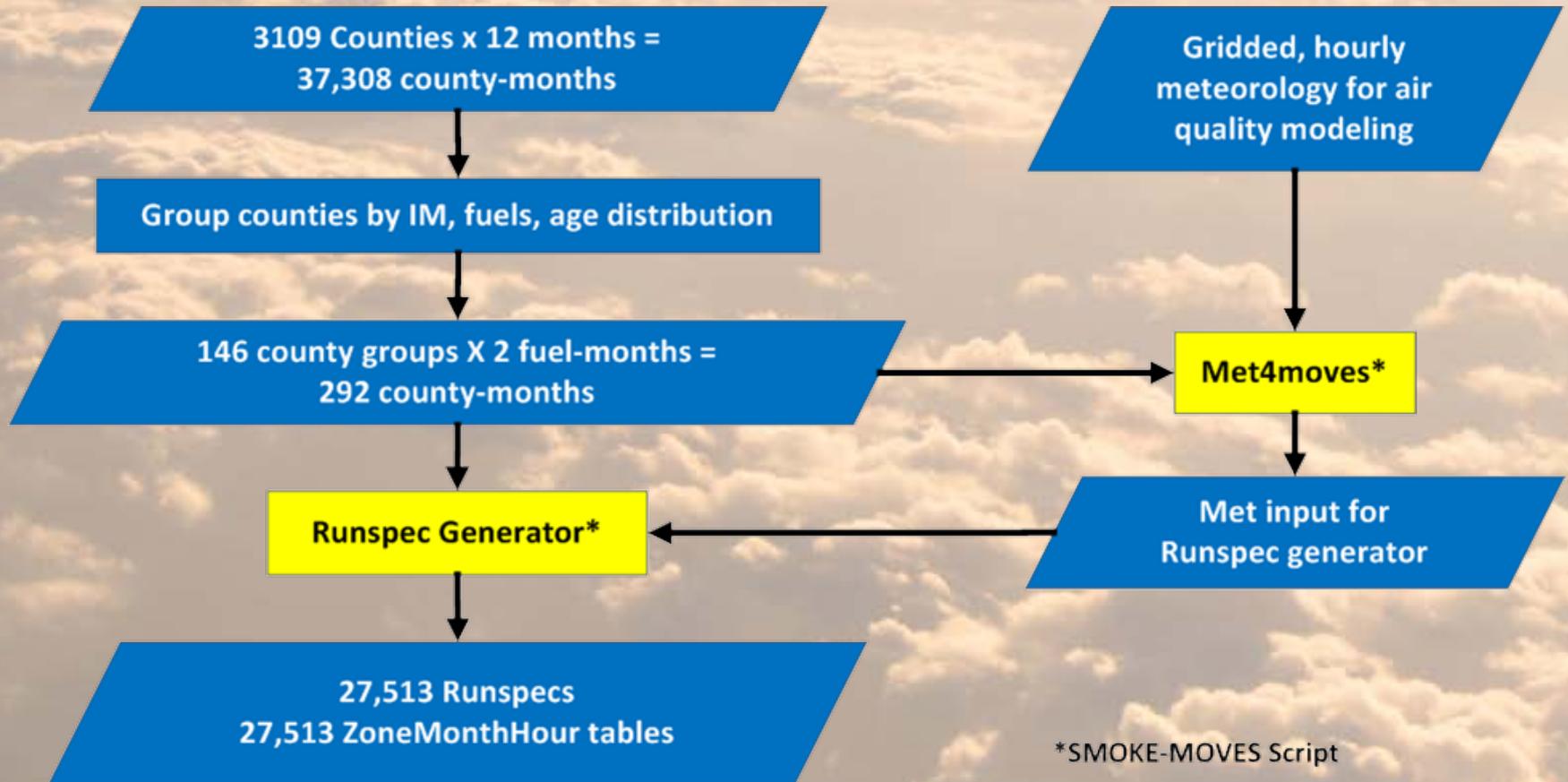
Gridded, hourly emissions ready for air quality model



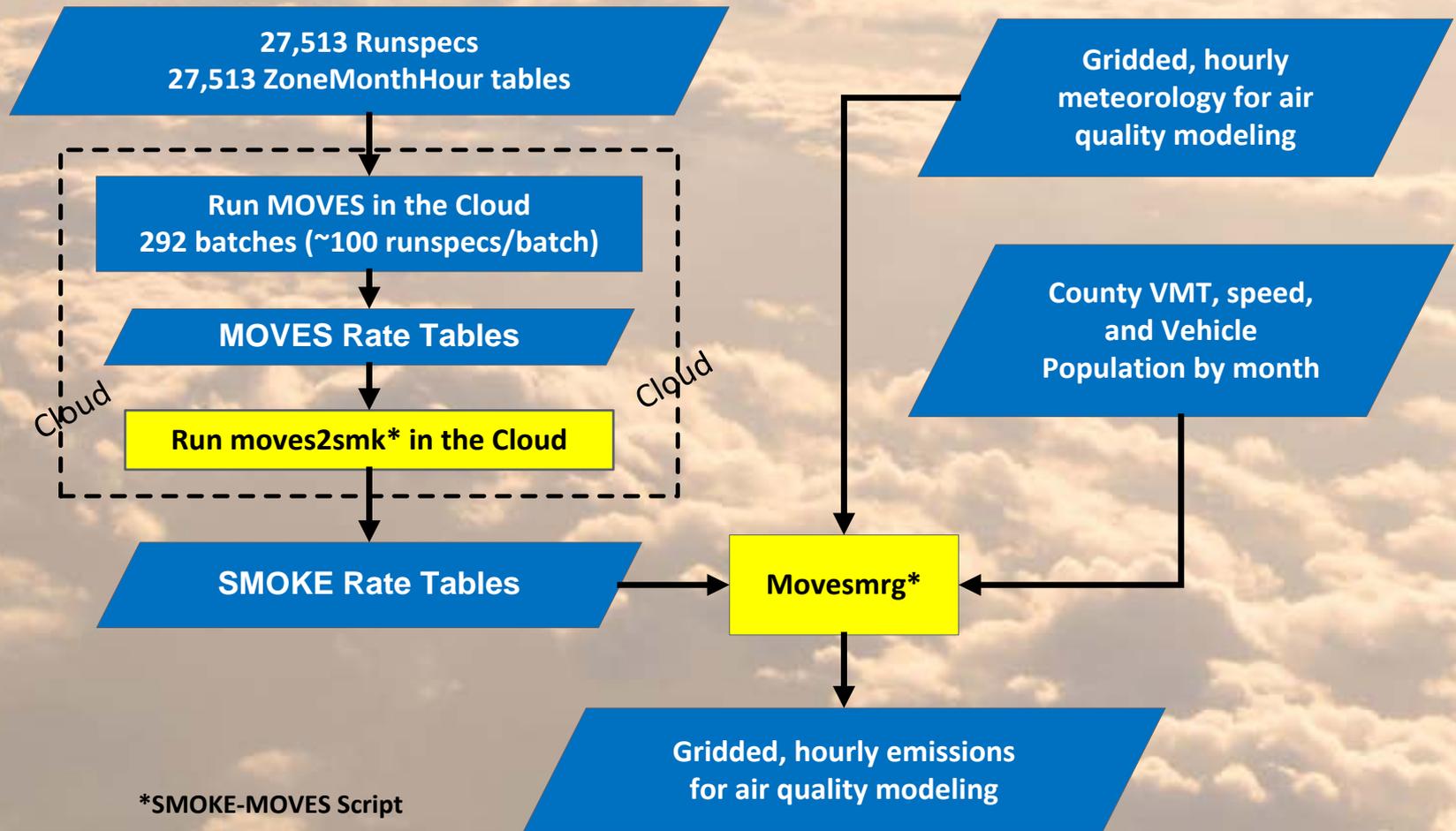
AQ Model

Source: Zubrow & Baek
MOVES Workshop, June 2011

SMOKE-MOVES Part 1



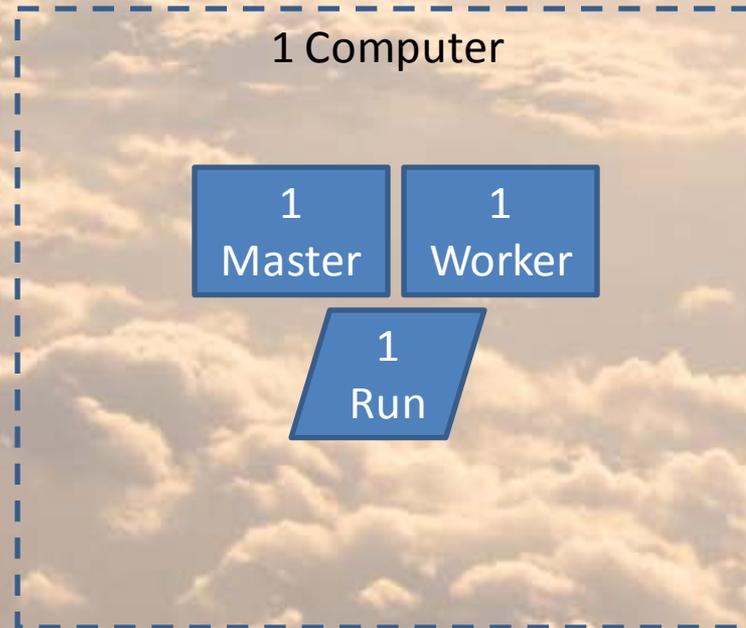
SMOKE-MOVES Part 2



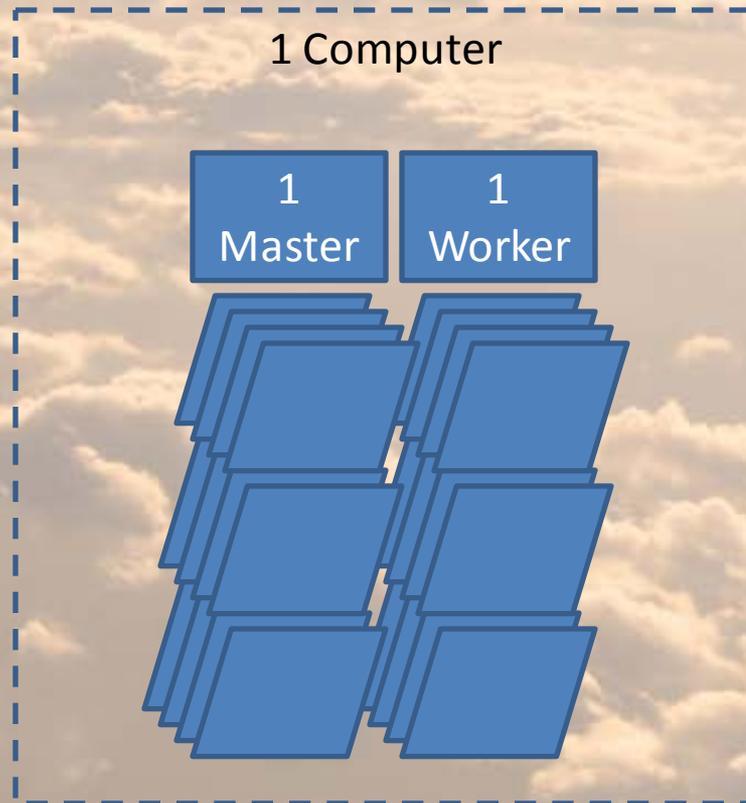
What is the Cloud?

- “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

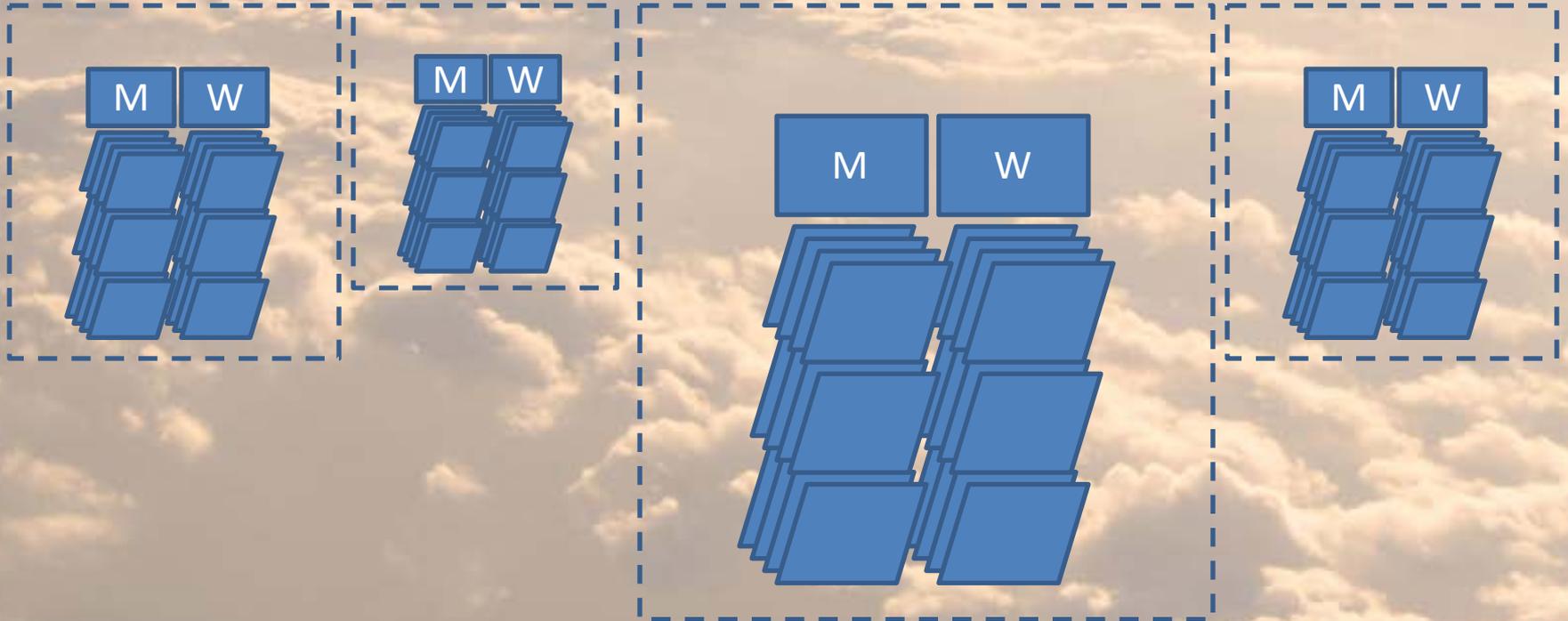
Simplest Cloud Approach



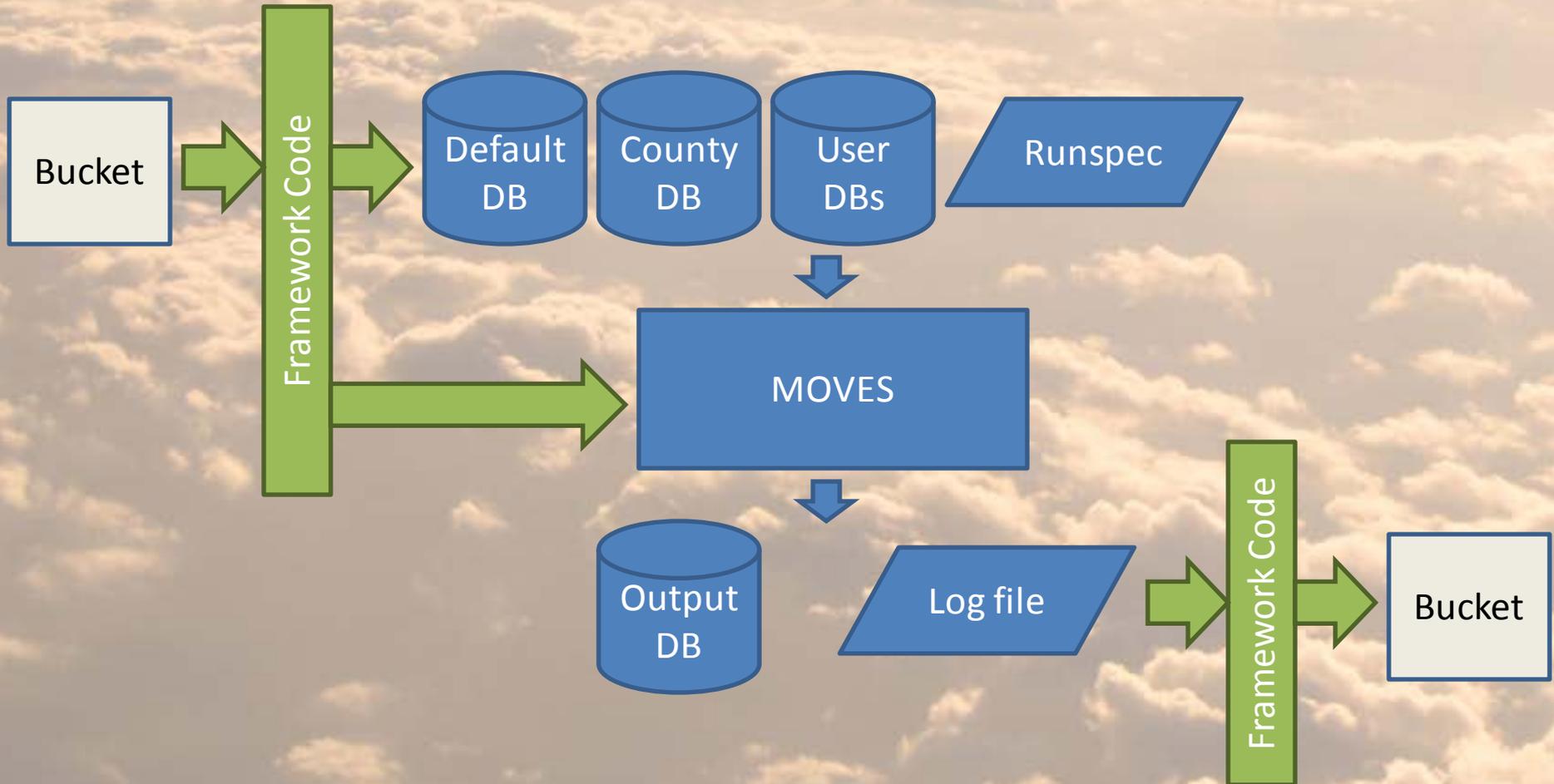
Our Cloud Approach



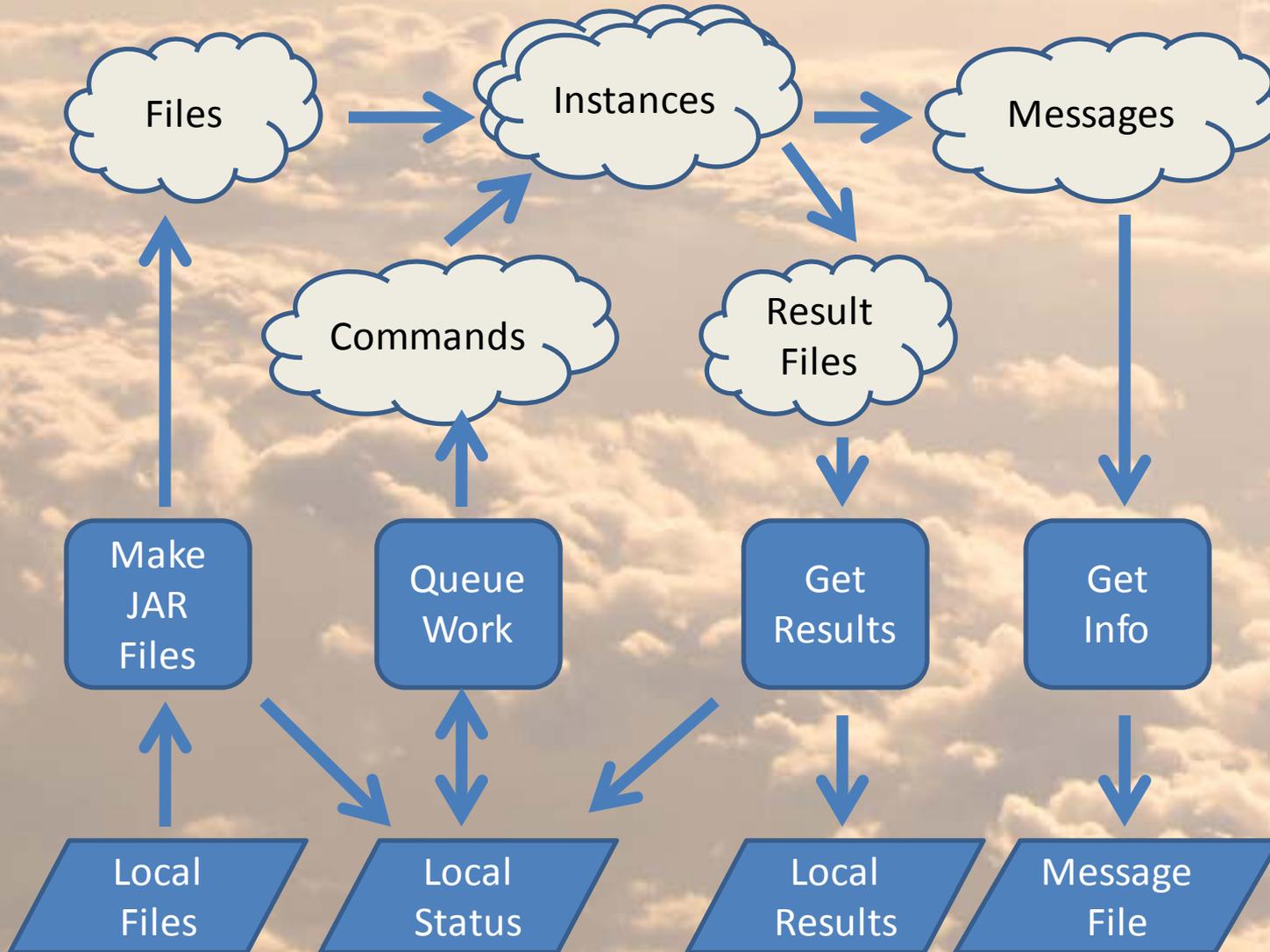
Our Cloud Approach



Cloud Instance Structure

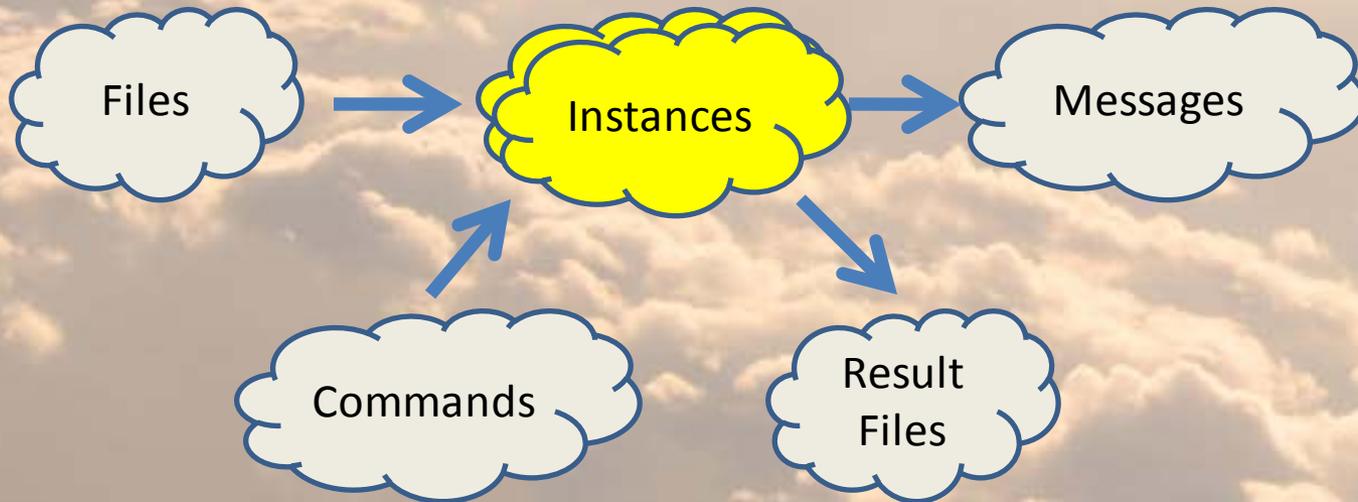


Data Flow

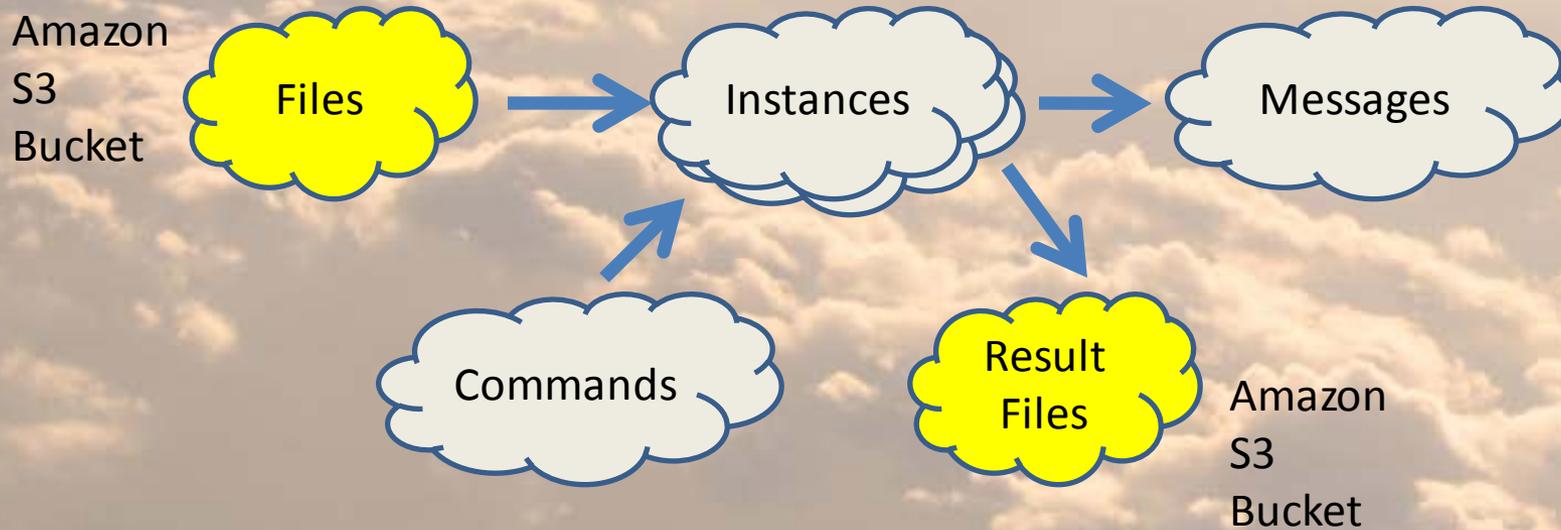


Cloud Machine Instances

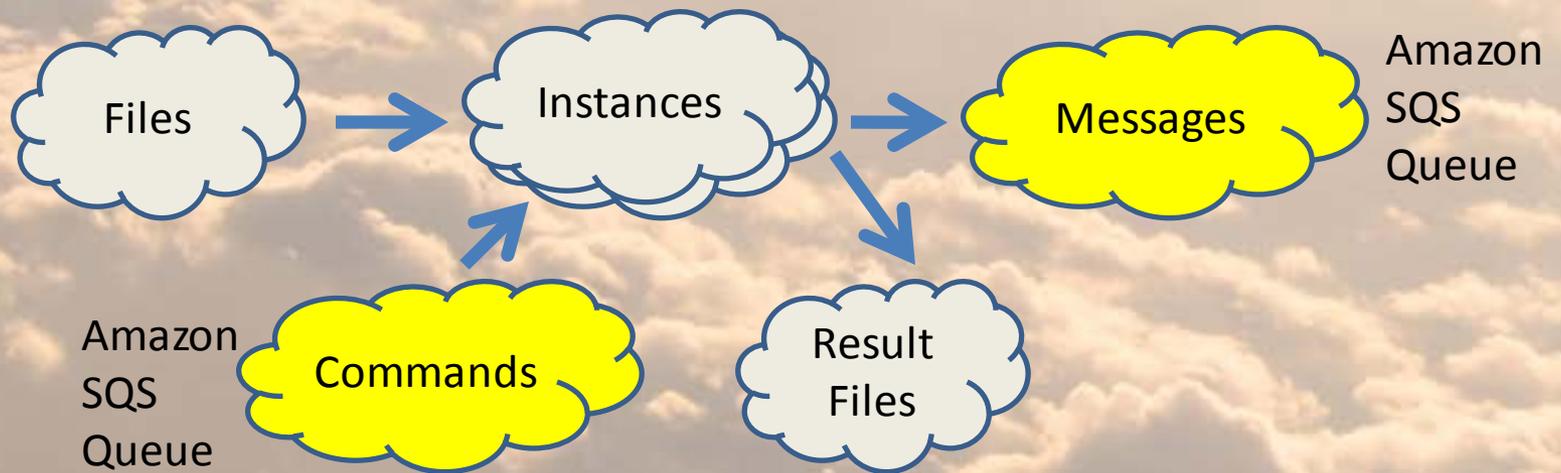
Amazon
EC2 AMI



Cloud Files



Cloud Queues



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.

 **Access Keys**  X.509 Certificates  Key Pairs

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

Created	Access Key ID	Secret Access Key	Status
March 2, 2011	[Redacted]	Show	Active (Make Inactive)

[Create a new Access Key](#)

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

 [Learn more about Access Keys](#)

Secret Access Key ✕
[Redacted]

Amazon Tips - Instances

Request Instances Wizard Cancel X

CHOOSE AN AMI INSTANCE DETAILS CREATE KEY PAIR CONFIGURE FIREWALL REVIEW

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 1 to 1 of 1 Items

AMI ID	Root Device	Name	Platform		
ami-a219ebcb	ebs	607601799279/us-epa-moves-ami-20110317a	Other Linux		Select

Free tier eligible if used with a micro instance. See [AWS free tier](#) for complete details and terms.

© 2008 - 2011, Amazon Web Services LLC or its affiliates. All right reserved. [Feedback](#) [Support](#) [Privacy Policy](#) [Terms of Use](#) An [amazon.com](#) company

Done

Amazon Tips - Instances

The screenshot shows the AWS Management Console interface with a 'Request Instances Wizard' dialog box open. The wizard is titled 'Request Instances Wizard' and has a 'Cancel' button in the top right corner. It features a progress bar with four steps: 'CHOOSE AN AMI', 'INSTANCE DETAILS', 'CREATE KEY PAIR', and 'CONFIGURE FIREWALL', with 'INSTANCE DETAILS' currently selected. Below the progress bar, there is a text prompt: 'Provide the details for your instance(s). You may also decide whether you want to launch your instances as "on-demand" or "spot" instances.' The 'Number of Instances' is set to '1' and the 'Availability Zone' is set to 'No Preference'. The 'Instance Type' is set to 'High-CPU Medium (c1.medium, 1.7 GB)'. There are three radio button options for instance types: 'Launch Instances' (selected), 'Request Spot Instances', and 'Launch Instances Into Your Virtual Private Cloud'. At the bottom of the dialog, there are 'Back' and 'Continue' buttons. The background shows the AWS console navigation menu and a search bar.

Request Instances Wizard Cancel

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: **Availability Zone:**

Instance Type:

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

Launch Instances Into Your Virtual Private Cloud

[< Back](#) [Continue >](#)

Amazon Tips - Instances

Request Instances Wizard Cancel X

CHOOSE AN AMI **INSTANCE DETAILS** CREATE KEY PAIR CONFIGURE FIREWALL REVIEW

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: AUTOTERMINATE=1, ACCESSKEY=, SECRETKEY=, JOBQUEUE=, STATUSQUEUE=

base64 encoded

Termination Protection: Prevention against accidental termination.

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

[Back](#) [Continue](#)

Amazon Tips - Instances

Request Instances Wizard Cancel

CHOOSE AN AMI INSTANCE DETAILS **CREATE KEY PAIR** CONFIGURE FIREWALL REVIEW

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*:

Create a new Key Pair

Proceed without a Key Pair

[< Back](#) [Continue >](#)

Amazon Tips - Instances

Request Instances Wizard Cancel

CHOOSE AN AMI **INSTANCE DETAILS** **CREATE KEY PAIR** **CONFIGURE FIREWALL** REVIEW

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 - OnlySSH
- sg-5afa6d33 - default

(Selected groups: sg-ba8419d3)

Create a new Security Group

[< Back](#) [Continue >](#)

© 2008 - [zotero](#)

Example Runs

	Scenario	
	2020	2025
Batches	292	292
Jobs	28,377	28,377

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

Timing

Process	Time (hours)	
	2020	2025
Loadzmh.plx	2.75	6.35
CreateAndPopulate.plx	5.98	7.13
Uploadjobs	5.28	5.51
Addjobs	0.71	0.83
Shortest Batch (42 jobs)		13.91
Longest Batch (167 jobs)		44.16
Downloadresults	13.76	13.14
Downloadpostresults	8.82	7.65
Batchstatus	0.15	0.18

Bottom Line

\$0.17/computer/hour

* 32 average hours/batch

* 292 batches

* 1 batch/computer

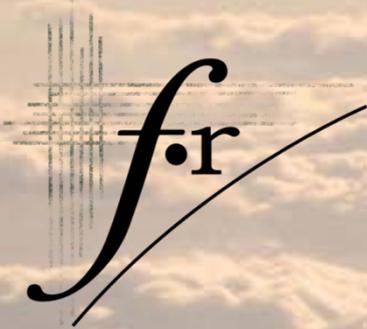
\$1,600 for the run

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

The logo for MOVES, featuring the word 'MOVES' in a bold, metallic, sans-serif font with a glowing effect, set against a dark rectangular background.

MOVES

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

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