Data Sharing Options for Scientific Workflows on Amazon EC2
Gideon Juve, Ewa Deelman, Karan Vahi, Gaurang Mehta, Benjamin P. Berman, Bruce Berriman, Phil Maechling

Francesco Allertsen
Vrije Universiteit
March 16, 2012
Introduction

- Workflow applications need to exchange lots of files between the tasks
- Workflow applications are starting to be deployed in the cloud
- What is the best (for costs and efficiency) way of store files for workflow applications in the cloud?
Overview

- What is a workflow application?
- Why cloud?
- Workflow applications
  - Epigenome
  - Broadband
  - Montage
- Storage options
  - Amazon S3
  - NFS
  - GlusterFS
  - PVFS
- Performance results
- Cost comparison
- Conclusions
What is a workflow application?

- Workflow application is a parallel application where tasks typically communicate through the use of files.
- Each task in a workflow produces one or more output files that become input files to other tasks.
Why cloud?

Traditionally, large-scale workflows have been run on academic HPC systems such as cluster and grids. Why moving to the cloud?

- root access
- control over the entire software environment
- using of VM images
- on-demand provisioning capabilities
Workflow Applications

- Epigenome (bio informatics application)
- Broadband (seismology application)
- Montage (astronomy application)

<table>
<thead>
<tr>
<th>Application</th>
<th>I/O</th>
<th>Memory</th>
<th>CPU</th>
</tr>
</thead>
<tbody>
<tr>
<td>Epigenome</td>
<td>Low</td>
<td>Medium</td>
<td>High</td>
</tr>
<tr>
<td>Broadband</td>
<td>Medium</td>
<td>High</td>
<td>Medium</td>
</tr>
<tr>
<td>Montage</td>
<td>High</td>
<td>Low</td>
<td>Low</td>
</tr>
</tbody>
</table>
## Epigenome

<table>
<thead>
<tr>
<th>Application</th>
<th>I/O</th>
<th>Memory</th>
<th>CPU</th>
</tr>
</thead>
<tbody>
<tr>
<td>Epigenome</td>
<td>Low</td>
<td>Medium</td>
<td>High</td>
</tr>
<tr>
<td>Broadband</td>
<td>Medium</td>
<td>High</td>
<td>Medium</td>
</tr>
<tr>
<td>Montage</td>
<td>High</td>
<td>Low</td>
<td>Low</td>
</tr>
</tbody>
</table>

- Contains 529 tasks
- Read 1.9GB of input data
- Produces 300MB of output data
- 99% of its runtime in the CPU and only 1% on I/O and other activities
## Broadband

<table>
<thead>
<tr>
<th>Application</th>
<th>I/O</th>
<th>Memory</th>
<th>CPU</th>
</tr>
</thead>
<tbody>
<tr>
<td>Epigenome</td>
<td>Low</td>
<td>Medium</td>
<td>High</td>
</tr>
<tr>
<td><strong>Broadband</strong></td>
<td><strong>Medium</strong></td>
<td><strong>High</strong></td>
<td><strong>Medium</strong></td>
</tr>
<tr>
<td>Montage</td>
<td>High</td>
<td>Low</td>
<td>Low</td>
</tr>
</tbody>
</table>

- Contains 768 tasks
- Read 6GB of input data
- Produces 303MB of output data
- 75% is consumed by tasks requiring more than 1GB of physical memory
Montage

<table>
<thead>
<tr>
<th>Application</th>
<th>I/O</th>
<th>Memory</th>
<th>CPU</th>
</tr>
</thead>
<tbody>
<tr>
<td>Epigenome</td>
<td>Low</td>
<td>Medium</td>
<td>High</td>
</tr>
<tr>
<td>Broadband</td>
<td>Medium</td>
<td>High</td>
<td>Medium</td>
</tr>
<tr>
<td>Montage</td>
<td>High</td>
<td>Low</td>
<td>Low</td>
</tr>
</tbody>
</table>

- Contains 10429 tasks
- Read 4.2GB of input data
- Produces 7.9GB of output data
- 95% of its time waiting on I/O operations
For the experiments it is used the Amazon EC2 infrastructure, because it is the most popular, feature-rich and stable commercial cloud available.
For the experiments it is used the c1.xlarge instance type.

- two quad core 2.33-2.66 GHz Xeon processors
- 7GB RAM
- 1690 GB local disk storage

It delivers the best overall performance for the applications considered.
For each workflow application we need to allocate storage for
- Application executables
- Input data
- Intermediate and output data

For all the experiments we did not perform or measure data transfer to/from the cloud.
Storage Options

- Amazon S3
- NFS
  - Dedicated note using the m1.xlarge instance type (16GB memory)
  - async option
- GlusterFS
  - NUFA (Non-Uniform File Access)
  - distributed
- PVFS
  - Version 2.6 instead of 2.8
- XtreemFS
Performance results for Epigenome

![Graph showing performance results for Epigenome]
Performance results for Broadband

![Graph showing runtime for different file systems with varying number of nodes/cores. The graph compares NFS, GlusterFS (NUFA), GlusterFS (distribute), PVFS2, S3, and Local storage systems. The x-axis represents the number of nodes/cores (1/8, 2/16, 4/32, 8/64), and the y-axis represents runtime in seconds.]
Performance results for Montage

![Graph showing performance results for Montage]
Cost comparison

There are three different cost categories when running an application on EC2

- resource cost
- storage cost
- transfer cost

Amazon charges for resources by the hour, and any partial hours are rounded up. We consider also how much does it cost if Amazon would charge per second.
Cost for Epigenome
Cost for Broadband

Number of Worker Nodes

Cost per Workflow

Cost for Broadband

Number of Worker Nodes

Cost per Workflow
Cost for Montage

![Bar chart showing cost per workflow for different numbers of worker nodes and file systems. The chart compares NFS, GlusterFS (NUFA), GlusterFS (distribute), PVFS2, S3, and Local. The x-axis represents the number of worker nodes (1, 2, 4, 8), and the y-axis represents the cost per workflow. The data indicates that the cost increases with the number of worker nodes for all file systems, but the specific costs vary depending on the file system.](chart.png)
Conclusions

- For different type of workflow application the performance of each file system can be different.
- Using a cost-per-second model can be cheaper than the cost-per-hour, so it’s better if you use one virtual cluster for more workflow applications instead of creating one for each application.
Take Home Message

Test your workflow application with different storage options to find the optimal solution for you, because there isn’t the perfect one.