Samvera and Fedora 4
Performance

Yinlin Chen, Ph.D.
Software Engineer
ylchen@vt.edu

University Libraries
Virginia Tech
Agenda

• Overview
• Fedora 4 performance team
• Performance Tests Overview
• Software and Tools
• Experiment Setup
• Experiment Results
• Conclusion
• Future Works
Overview

- Measure Fedora 4 performance and explore limit
- Compare Fedora 4 performance between versions
- Use a uniform platform (AWS) to conduct Fedora 4 performance tests
- Able to conduct repeatable performance tests
- Perform tests and evaluate results
- Supported by the AWS Cloud Credits for Research program
Fedora 4 performance team

• Dynamic team
  – Yinlin Chen (Virginia Tech)
  – Andrew Woods (DuraSpace)
  – Colin Gross (University of Michigan)
  – Joshua Westgard (University of Maryland)
  – Danny Bernstein (DuraSpace)

• Fedora Performance - Scale meeting
• You are welcome to participate anytime!
Performance Tests Overview

• Currently 6 performance tests
  – Demonstrate how the performance of the Fedora 4 repository changes as the size of the repository grows under various but discrete uses
• Developed by the Fedora 4 performance team
• 5 performance tests are implemented in Jmeter
• Made a performance request after every 1000 requests
## Performance Tests

<table>
<thead>
<tr>
<th>Test</th>
<th>Type</th>
<th>File Size</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Upload Binary File - large</td>
<td>240-250MB</td>
</tr>
<tr>
<td>2</td>
<td>Upload Binary File - small</td>
<td>&lt; 4096 Byte</td>
</tr>
<tr>
<td>3</td>
<td>Upload Binary File - medium</td>
<td>10KB – 100KB</td>
</tr>
<tr>
<td>4</td>
<td>Create containers</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>Create containers with RDF bodies</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>Create a mix of resources (Containers and files)</td>
<td></td>
</tr>
</tbody>
</table>

*Note: RDF stands for Resource Description Framework.*
Software and Tools

• Fedora 4: https://github.com/fcrepo4/fcrepo4
• Fedora 4 Ansible: https://github.com/VTUL/fcrepo4-ansible
• Jmeter v3.1, and R v3.4.1: Installation scripts for Ubuntu
  fcrepo4-jmeter: https://github.com/fcrepo4-labs/fcrepo4-jmeter
• Fcrepo performance analysis: https://github.com/fcrepo4-labs/fcrepo_perf_analysis
• Amazon Web Services (AWS): http://aws.amazon.com
Experiment Setup

- All experiments are conducted in AWS
- Deploy Fedora 4 using Fedora 4 Ansible
- Fedora 4 version: 4.7.3 and 4.7.4

<table>
<thead>
<tr>
<th>Test</th>
<th>Server</th>
<th>Client</th>
<th>Time</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>M3.xlarge (4 vCPUs, 15 GiB memory, 16TB)</td>
<td>M3.xlarge (4 vCPUs, 15 GiB memory)</td>
<td>5 days</td>
</tr>
<tr>
<td>2-5</td>
<td>M3.large (2 vCPUs, 7.5 GiB memory, 500GB)</td>
<td>T2.medium (2 vCPUs, 4 GiB memory)</td>
<td>7 days</td>
</tr>
</tbody>
</table>
Experiment workflow

Guide: How to Conduct Performance Tests and Share Results
Running Experiments in AWS
Experiment Result - Test 1

- Upload many large size (240-250MB) of files to Fedora 4
- Total file size: 4.7.3 (9.3T) 4.7.4 (10T)

X: PUT Perf Container
Y: Time elapsed
Experiment Result - Test 2

- Upload many small size of files to Fedora 4

Fedora 4.7.3

Fedora 4.7.4

X: PUT Perf Container
Y: Time elapsed
Experiment Result - Test 3

- Upload many medium files (10-100KB) to Fedora 4

Fedora 4.7.3

Fedora 4.7.4

X: PUT Perf Container
Y: Time elapsed
Experiment Result - Test 4

- Create many containers to Fedora 4

Fedora 4.7.3

Fedora 4.7.4

X: PUT Perf Container
Y: Time elapsed
Experiment Result - Test 5

- Create many containers with RDF bodies to Fedora 4

Fedora 4.7.3

Fedora 4.7.4

X: PUT Perf Container
Y: Time elapsed
<table>
<thead>
<tr>
<th>Test Description</th>
<th>Total Objects (v4.7.3)</th>
<th>Total Objects (v4.7.4)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Test 1 - Large</td>
<td>42,043</td>
<td>45,047</td>
</tr>
<tr>
<td>Test 2 - Small</td>
<td>815,065</td>
<td>771,750</td>
</tr>
<tr>
<td>Test 3 - Medium</td>
<td>836,123</td>
<td>924,265</td>
</tr>
<tr>
<td>Test 4 - Containers</td>
<td>930,959</td>
<td>959,011</td>
</tr>
<tr>
<td>Test 5 - Containers with RDF bodies</td>
<td>894,149</td>
<td>880,333</td>
</tr>
</tbody>
</table>
Conclusion

• Conduct performance test in a unified environment – AWS
• Software are open, customizable, and easy to use
• Repeatable experiment setup that can be used to compare the performance between different Fedora 4 versions
• Fedora 4.7.4 performs better than 4.7.3 in most tests
• You can conduct your own performance tests
• Your contributions are more than welcome!
Future Works

• Extend current performance tests
  – With JDBC Object Store
  – Multiple clients

• Create new performance tests
  – Hyrax
  – Solr / Jena Fuseki / Sesame / Camel …
  – Other real world scenarios…

• New use cases from the Samvera community
  – What kind of tests that would be interested to YOU?
Resources

- Fedora 4: [https://github.com/fcrepo4/fcrepo4](https://github.com/fcrepo4/fcrepo4)
- Fedora Tech: [https://groups.google.com/forum/#!forum/fedora-tech](https://groups.google.com/forum/#!forum/fedora-tech)
- Fedora 4 Performance and Scalability: [https://wiki.duraspace.org/display/FF/Performance+and+Scalability](https://wiki.duraspace.org/display/FF/Performance+and+Scalability)
- Fedora 4 Performance test How-to document: [https://wiki.duraspace.org/display/FF/How+to+Conduct+Performance+Tests+and+Share+Results](https://wiki.duraspace.org/display/FF/How+to+Conduct+Performance+Tests+and+Share+Results)
- Fedora 4 Jmeter: [https://github.com/fcrepo4-labs/fcrepo4-jmeter](https://github.com/fcrepo4-labs/fcrepo4-jmeter)
- Fedora 4 Ansible: [https://github.com/VTUL/fcrepo4-ansible](https://github.com/VTUL/fcrepo4-ansible)
- Fcrepo performance analysis: [https://github.com/fcrepo4-labs/fcrepo_perf_analysis](https://github.com/fcrepo4-labs/fcrepo_perf_analysis)
Q & A

Thank you!

Yinlin Chen
ylchen@vt.edu
Demo