OGSSim: Open Generic data Storage systems Simulation tool

Sebastien Gougeaud\textsuperscript{1}, Soraya Zertal\textsuperscript{1}, Jacques-Charles Lafoucriere\textsuperscript{2}, Philippe Deniel\textsuperscript{2}

\textsuperscript{1}Université de Versailles St-Quentin-en-Yvelines
PRiSM – CNRS/FRE-3709
\textsuperscript{2}CEA-DAM

26 August 2015
Plan

1. Introduction
2. OGSSim design & implementation
3. Experimentation & validation
4. Conclusion
Main existing simulators:
- DiskSim: HDD, synthetic trace generator, simple systems
- DiskSim extensions (as FlashSim): SSD
- SSDSim: SSD, advanced command handling (copyback, multiplane, ...)

Drawbacks:
- Difficulty of adding new computation models or technologies
- No support of complex systems (different disks or configurations)
Why OGSSim?

- Open: simple inclusion of new modules as:
  - device maintenance modules

- Generic: handling various and hierarchical configurations of different devices as:
  - HDD
  - SSD
  - NV-memory (in the future)
Global Overview

- Workload
- Pre-processing
- Volume driver
- Hardware configuration
- Performance evaluation
- Execution
- Device driver
- Trace
- Arch.
- Visualization
- Results

Request array
Hardware
Request index
Subrequest index
Times
Subrequest index
Hardware
Input files

Trace (RAW format)

<table>
<thead>
<tr>
<th>#timestamp</th>
<th>type</th>
<th>address</th>
<th>size</th>
</tr>
</thead>
<tbody>
<tr>
<td>8.96547</td>
<td>1</td>
<td>862817376</td>
<td>8</td>
</tr>
<tr>
<td>30.7717</td>
<td>1</td>
<td>38052016</td>
<td>8</td>
</tr>
<tr>
<td>112.142</td>
<td>0</td>
<td>1687375240</td>
<td>16</td>
</tr>
<tr>
<td>186.59</td>
<td>0</td>
<td>1301699720</td>
<td>16</td>
</tr>
<tr>
<td>193.491</td>
<td>0</td>
<td>953515296</td>
<td>16</td>
</tr>
<tr>
<td>213.821</td>
<td>0</td>
<td>184457704</td>
<td>16</td>
</tr>
<tr>
<td>223.786</td>
<td>1</td>
<td>1618577376</td>
<td>8</td>
</tr>
<tr>
<td>231.497</td>
<td>0</td>
<td>1923929216</td>
<td>16</td>
</tr>
<tr>
<td>255.712</td>
<td>0</td>
<td>1467238296</td>
<td>16</td>
</tr>
<tr>
<td>298.31</td>
<td>1</td>
<td>743947312</td>
<td>8</td>
</tr>
<tr>
<td>398.651</td>
<td>0</td>
<td>1571641704</td>
<td>16</td>
</tr>
</tbody>
</table>

Hardware configuration (XML format)

```xml
<architecture>
  <buses nbbuses="3">
    <bus name="B0" nbports="16" bandwidth="750" type="SCSI" />
  </buses>
  <system nbtiers="1" bus="B0">
    <tier nbvolumes="1" bus="B1">
      <volume nbdevices="10" bus="B2">
        <config type="RAID01" stripeunitsize="2048" />
        <device file="test_env/hdd_disk.xml" />
      </volume>
    </tier>
  </system>
</architecture>
```

OGSSim configuration (XML format)

```xml
<config>
  <general>
    <log mlvl="0" file="log/test_env_" />
  </general>
  <workload>
    <zmq intr="preproc" prot="tcp" addr="localhost" port="5555" />
    <subreq bsiz="50000" />
  </workload>
...</config>
```
Module description (1/3)

- **Workload**
  - Parses the trace file
  - Constructs the request array structure

- **Hardware configuration**
  - Parses the hardware configuration file
  - Constructs the hardware structure

- **Pre-processing**
  - Instantiates the volume driver modules
  - Launches the simulation process
  - Redirects the requests to the right volume driver
Module description (2/3)

- **Volume driver**
  - Instantiates the device driver modules
  - Decomposes the requests into subrequests depending on its layout model
  - Redirects the subrequests to the right device driver

- **Device driver**
  - Transmits the subrequests to the execution module
  - Updates the device state
  - Calls device maintenance routines
Module description (3/3)

- **Execution**
  - Determines the response time of the requests (service, transfer and waiting times)
  - Writes the results in an output file

- **Performance evaluation**
  - Retrieves information from the whole simulator
  - Constructs the visualization graphs or histograms
Experimentations

- Simulation time
- Spatial behavior
- Temporal behavior
Evaluation of the simulation time

Workload

- **S_trace1**: MPI IO Test, Los Alamos National Laboratory
- **S_trace2**: ALEGRA family of codes, Sandia National Laboratories
- **OLTP trace**: 67% of read requests, 4KB size, 1600 requests per second
Temporal behavior

- S_trace3 with three configurations:
  1. JBOD HDD
  2. RAID (5+2) HDD
  3. JBOD SSD / RAID 01 HDD

![Config. 1](image1)
![Config. 2](image2)
![Config. 3](image3)
Spatial behavior

- $S_{\text{trace1}}$ with three configurations
OGSSim, an Open and Generic Storage system Simulation tool which can:

- Handle various configurations and basic devices
- Study the whole system
- Analyze its performance, reliability, etc.

Future works

- Validation against real system
- Failure mode
- Bus queueing model
- Device mechanisms

Source code availability of OGSSim with basic features on October 2015
Thanks for your attention!