HammerCloud: benchmarking storages

CRISP Annual Meeting

Valentina Mancinelli

IT/SDC

02/06/2016

What is HammerCloud

HammerCloud is a framework for testing distributed systems.

Two types of tests:

- Functional (automated) testing: Are autoscheduled, which means configure once, run forever. Used usually to frequently submit few "ping" jobs for basic continuous validation.

- Stress (on-demand) testing: Are on-demand, which means configure once, run once. Used to submit benchmarks, evaluate changes to the infrastructure or configuration, evaluate SW changes, compare performances. Usually used to submits a large number of jobs concurrently during a slot of time.

HammerCloud

Composed by:

- Service to schedule, run, track status, gather results/statistics of tests

- Web interface to:
  - Display test status and test results
  - Configure new tests

Test monitoring

Overview of all the running jobs

View of the results of a test

IT-SDC

Test configuration

Current uses

- Used at CERN for Frequent validation and stress testing for different resources.
- Enabling VO- and site-administrators to run tests in an automated or on-demand manner.
- Used by 3 experiments (Atlas, CMS, LHCb)
- Many use cases

Benchmarking

- HammerCloud tests to submit MC simulation to crosscheck HEPSPEC (measure CPU performances)
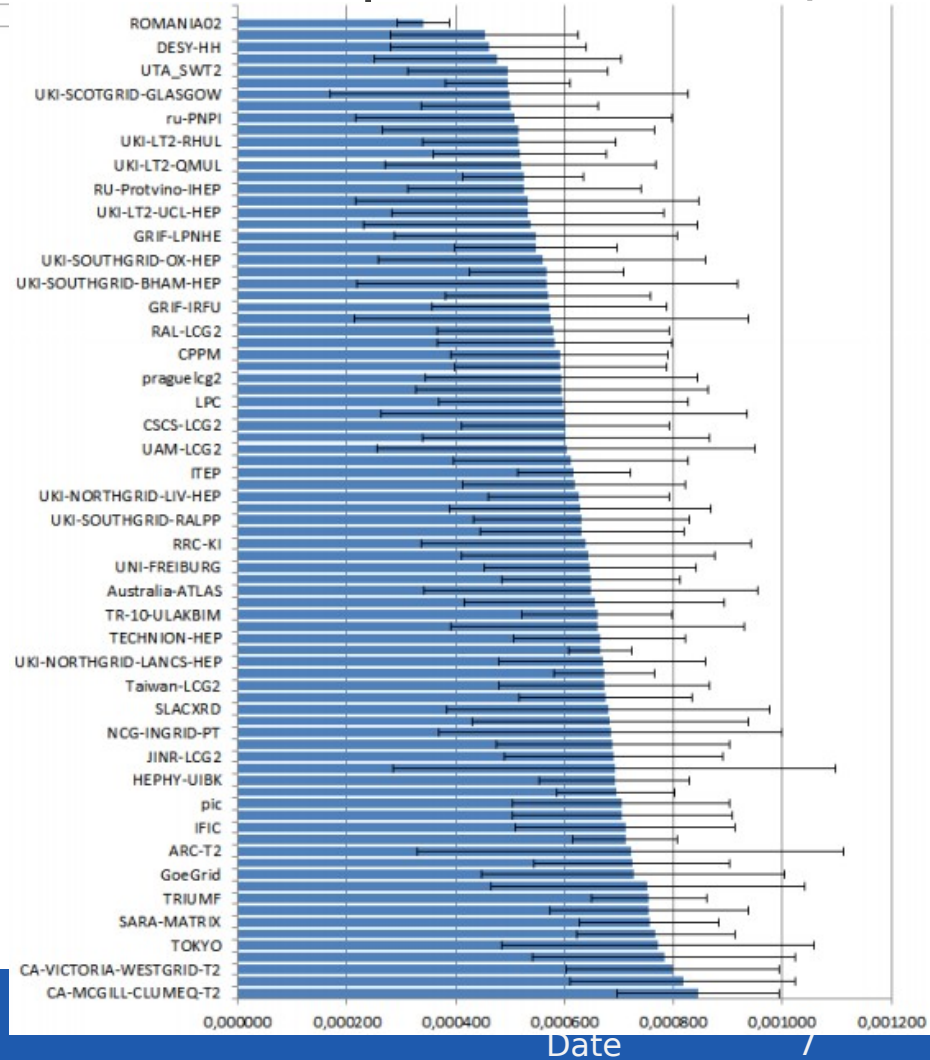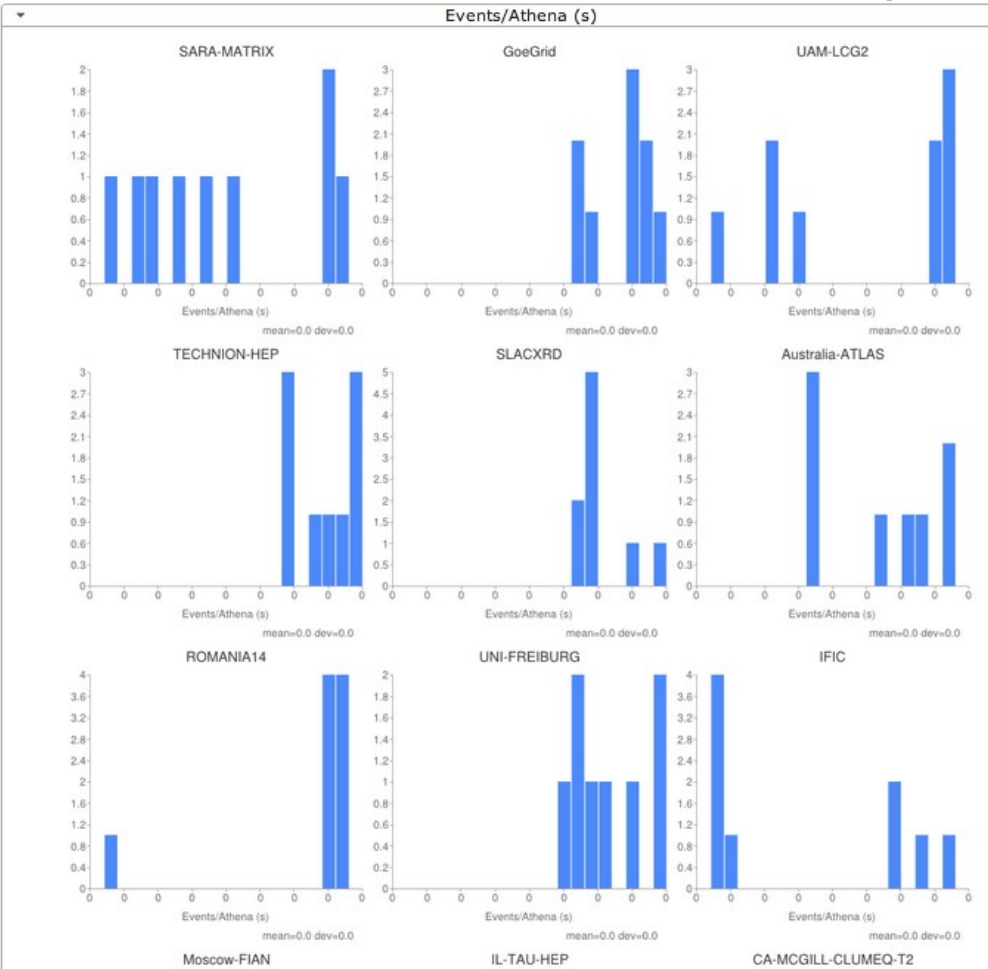
Storage testing

- Functional test to constantly monitor DPM performances

- DPM (Disk Pool Manager) is a grid enabled disk storage management system operated at more than 200 sites in the WLCG.

- Use HammerCloud to submit jobs to test data access performances:
  - Jobs use root (analysis application) to access storage, different protocols
  - 2 different read pattern: 1% and 100% of input file
  - Upload metrics to a database for analysis and visualization

state	id	host	clouds	start time (CET)	end time (CET)	total jobs
running	20036284	it-hammercloud-submit-atlas-03	T0_PANDA, UK_PANDA, FR_PANDA, 3 more...	21/5/2014 18:28	22/5/2014 20:14	141

Input type: PANDA
Output DS: user.gangarbt.hc20036284.*
Input DS Patterns: ROOTIOTests/datasetsToUse.txt
Ganga Job Template: ROOTIOTests/ROOTIOTests1667_panda.tpl
Athena User Area: ROOTIOTests/ROOTIOTests1667_svn.tar.gz
Athena Option file: ROOTIOTests/run-DPM.sh
Template: SDC Tests
View Test Directory (for debugging)

Overall.

c (131)
r (3)
s (4)
h (1)
f (2)

Overall.

CPU percentage
mean=36.5 dev=16.2

more plots »

Sites

Show 10 entries   Search:
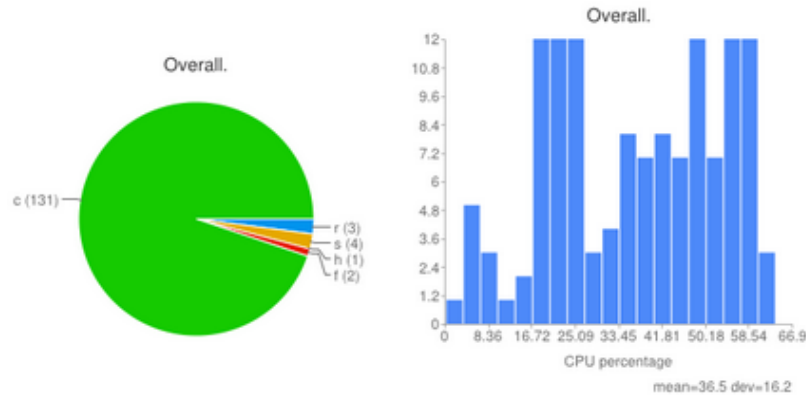
Site	S	R	C	F	Eff	T	Datasets	Queue	Max R	Resubmit	R. Force	Link
ANALY_TAIWAN_XROOTD_SL6	0	0	0	0	0.00	0	1	0	1	yes	no	»
ANALY_LAPP	1	0	5	0	1.00	6	1	0	1	yes	no	»
ANALY_INFN-FRASCATI	1	0	29	0	1.00	31	1	0	1	yes	no	»
ANALY_GLASGOW_SL6	0	2	27	1	0.96	30	1	0	1	yes	no	»
ANALY_ECDF_SL6	1	0	11	0	1.00	12	1	0	1	yes	no	»
ANALY_CERN_XROOTD	1	0	28	0	1.00	29	1	0	1	yes	no	»
ANALY_BNL_LONG	0	1	31	1	0.97	33	1	0	1	yes	no	»

DPM Root Read 1% TTC

WALL time [s] for DPM Root Read 1% TTC

	Value
BNL_LONG-DPMhttp-CERN_DPM_TRUNK	88.6
BNL_LONG-DPMxrootd-CERN_DPM_TRUNK	87.16
ECDF_SL6-DPMhttp-CERN_DPM_TRUNK	78.61
ECDF_SL6-DPMxrootd-CERN_DPM_TRUNK	72.93
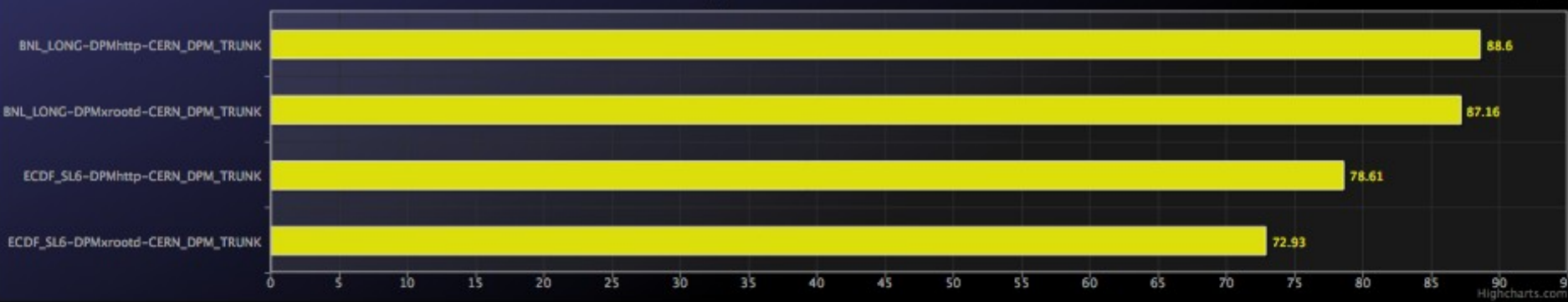
HammerCloud for CRISP

- Use HammerCloud as a tool to submit benchmarks for storage solution analysis

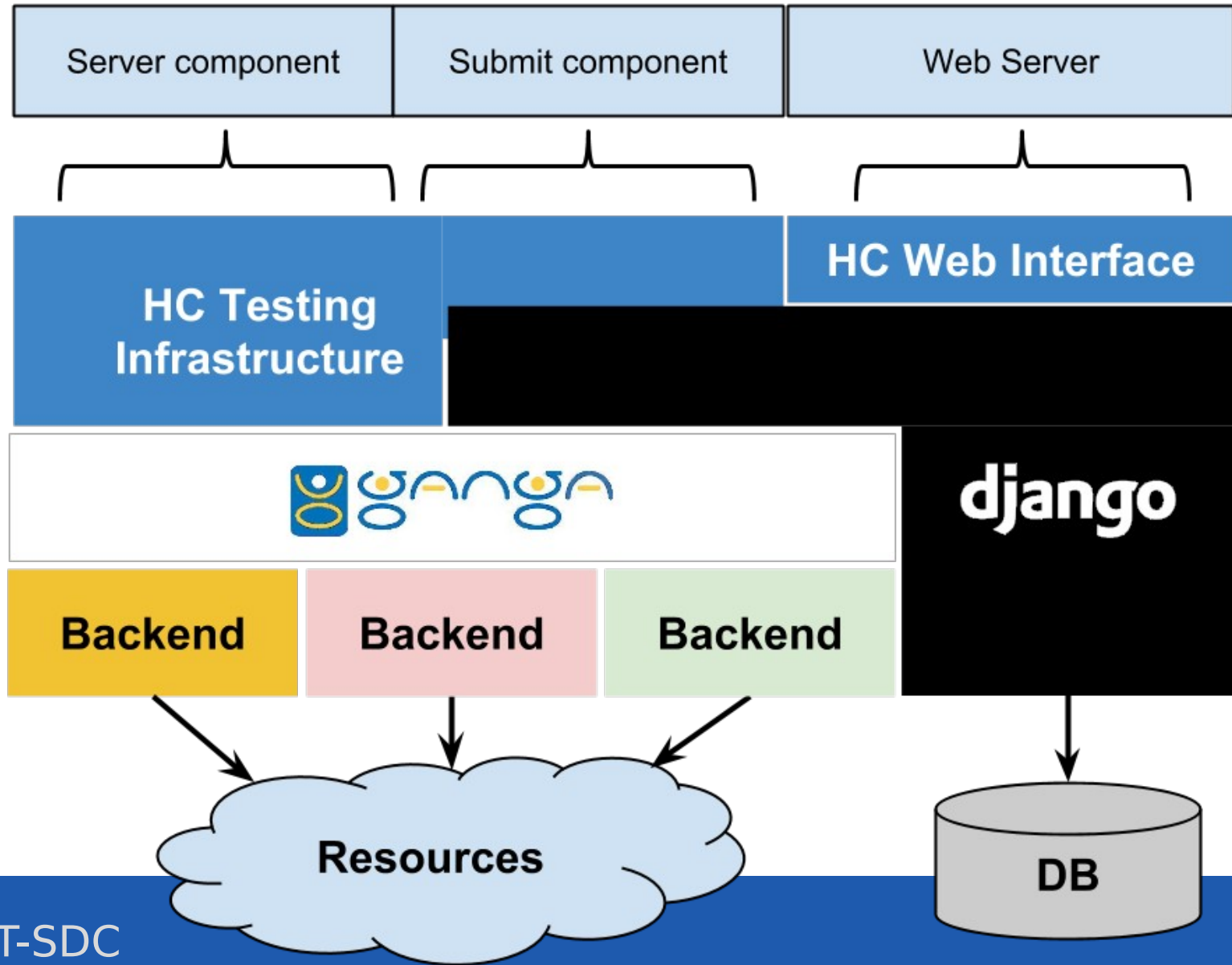- Develop and run a suite for quantitative performance **evaluation of data access interfaces**

- synthetically reproduce arbitrary access patterns through appropriate parametrisation

HammerCloud Components

Modular implementation

- Core that implements basic components

- Composed by applications that extend core model. Easily extensible and customizable

- Ganga (submission component) provides:

  - various backend to submit jobs
  - plugin system that allows to develop specific backend

Start using HammerCloud

Basic deployment:

- 1 host with web server, server component, submit component
- DB (hosted local or its own server)
- Run executable locally to access local/remote storage systems.

Then...

- HC is a distribuited system, easy to add new submission nodes to scale load
- Use different backend provided by Ganga to submit to various batch systems