ViGs: A Grid Simulation and Monitoring Tool for ATLAS Workflows

Aaron T. Thor[1], Gergely V. Záruba[1], David Levine[1], Kaushik De[2], Torre J. Wenaus[3]

[1]Department of Computer Science and Engineering, [2]Department of Physics
University of Texas at Arlington
[3]Brookhaven National Lab

Partially supported by US ATLAS and BNL

ATLAS

- The Large Hadron Collider (LHC) was built by the European Organization for Nuclear Research (CERN), and is the world's largest and highest-energy particle accelerator
- ATLAS (A Toroidal LHC ApparatuS) is one of the six particle detector experiments (ALICE, ATLAS, CMS, TOTEM, LHCb, and LHCf) constructed at the LHC

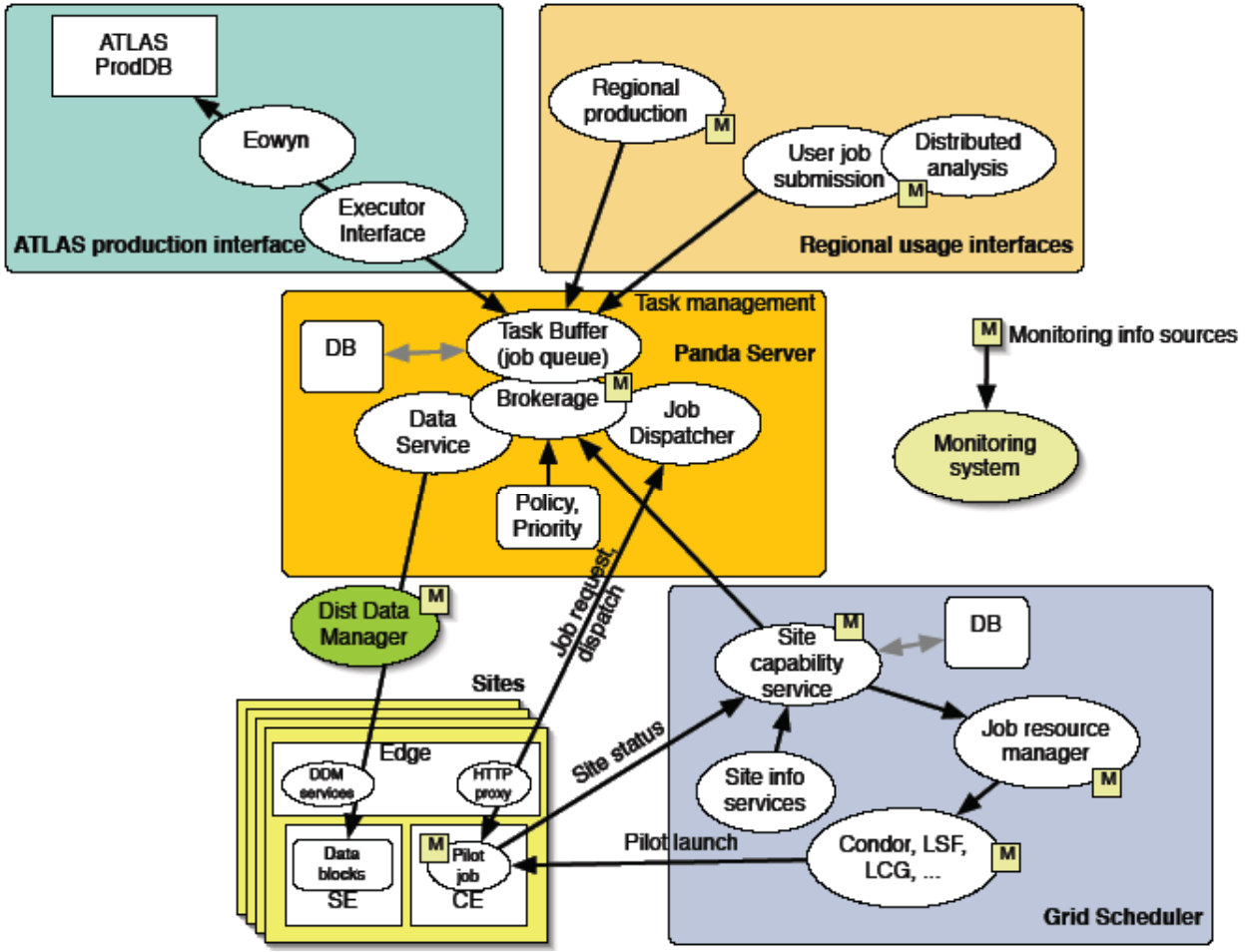
ATLAS

- *When the LHC becomes fully functional, generation of vast amount of data from experiments would soon follow in the near future*
- ATLAS jobs need to be scheduled and distributed to computing nodes located in dozens of clusters around the world (across US, Canada, Europe etc)

PanDA

- PANDA (Production ANd Distributed Analysis) is a software system designed and implemented under the direction of BNL with the goals of improved robustness, scalability, and better data / processing integration
- Purpose of PanDA is to handle ATLAS requirements for large scale scientific experimental event production, and the analysis processing of such events across geographically distributed locations.

PanDA



PanDA architecture

Adopted from: https://twiki.cern.ch/twiki/bin/view/Atlas/Panda

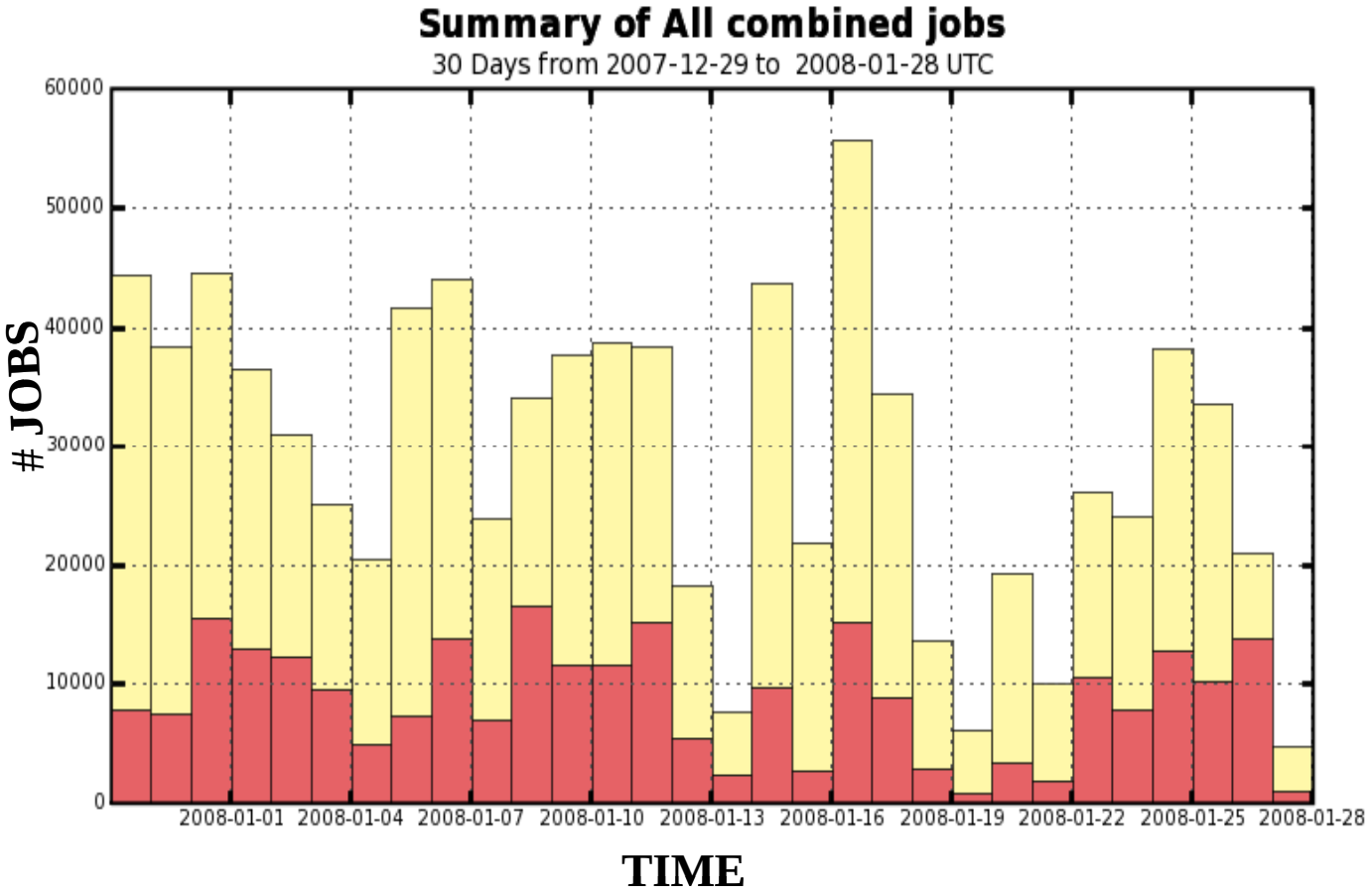
PanDA

Production job summary, last 12 hours (Details: errors, nodes)

Cloud Information	Nodes	Jobs	Latest	Pilots (3hrs)	defined	assigned	waiting	activated	running	holding	transferring	finished	failed tot trf other	
Overall Production	7908	81902	04-15 05:41	10807	0 / 0	2158 / 0	42 / 0	28922 / 0	14323 / 0	2731 / 0	14996 / 3464	16236 / 0	2534 / 19	14% 0% 13%
CA	909	9764	04-14 23:41	501	0	0	0	3692	1845	85	1008 / 0	3037	96	3% 0% 3%
DE	1219	14568	04-14 23:41	680	0	361	0	5391	2793	43	2020 / 0	3560	398	10% 0% 10%
ES	221	1980	04-14 23:40	32	0	0	0	1022	380	329	82 / 0	148	16	10% 0% 10%
FR	1802	15905	04-14 23:41	4886	0	462	0	3608	1971	1965	4121 / 1001	3514	263	7% 0% 7%
IT	201	2449	04-15 05:41	150	0	0	0	1275	87	6	458 / 63	548	75	12% 0% 12%
NL	266	7348	04-14 23:41	426	0	31	11	1425	565	11	3904 / 2368	205	1206	85% 0% 85%
UK	1716	11653	04-14 23:41	1248	0	1056	15	4796	2843	66	1179 / 31	1613	97	6% 2% 4%
US	1396	16435	04-14 23:41	2556	0	216	31	6990	3531	225	1972 / 0	3231	261	7% 1% 7%
TW	178	1800	04-14 23:40	328	0	32	0	723	308	1	252 / 1	380	103	21% 0% 21%

Panda Production job summary

PanDA

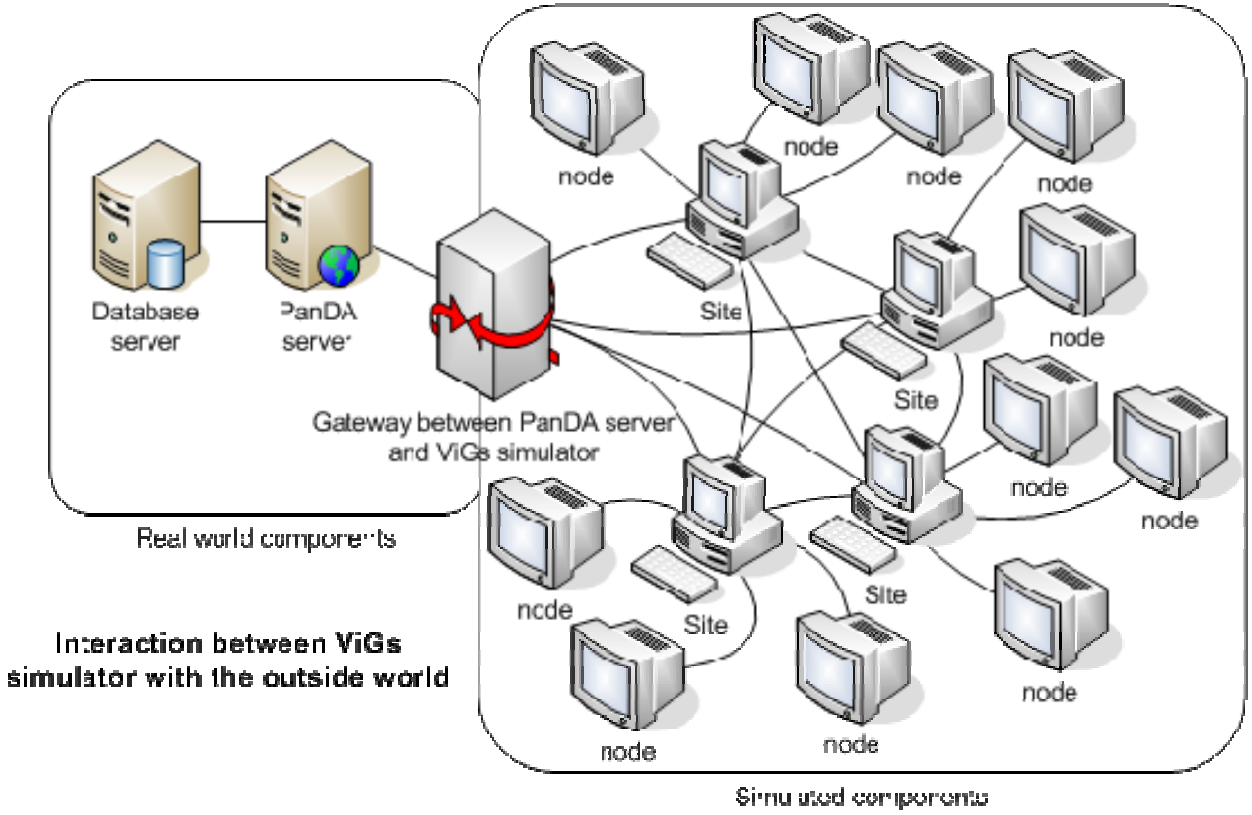


Summary of Panda finished and failed jobs over a one month period

ViGs

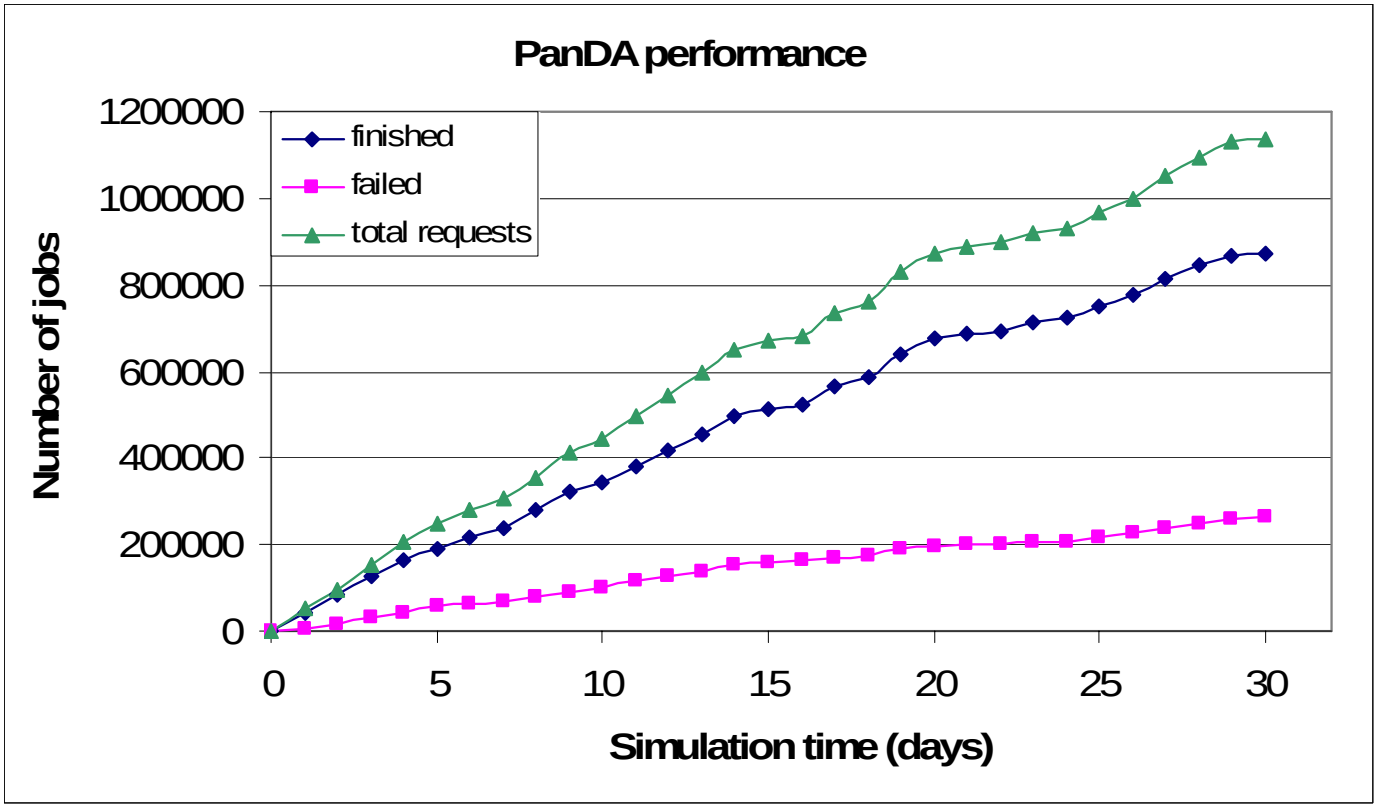
- In order to validate PanDA as well as exercise functionality, ViGs provides a simulated environment where actual ATLAS jobs are scheduled with simulated execution.
- ViGs is a hybrid simulation tool with an internal simulation scheduling engine and client simulator that can interact with real grid scheduling entities over the network (such as the PanDA system)

ViGs Overview



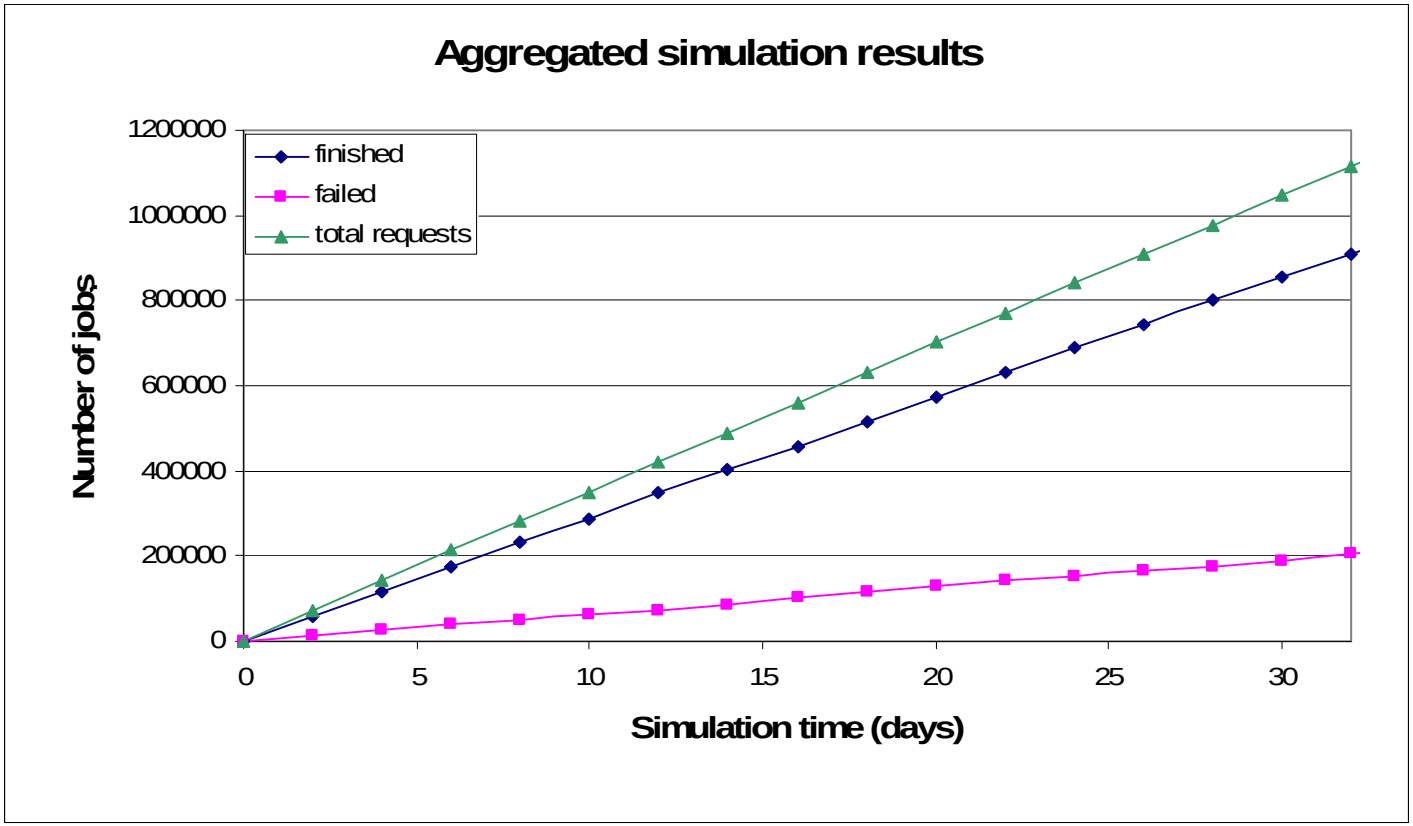
ViGs Interaction diagram

ViGs – PanDA sample



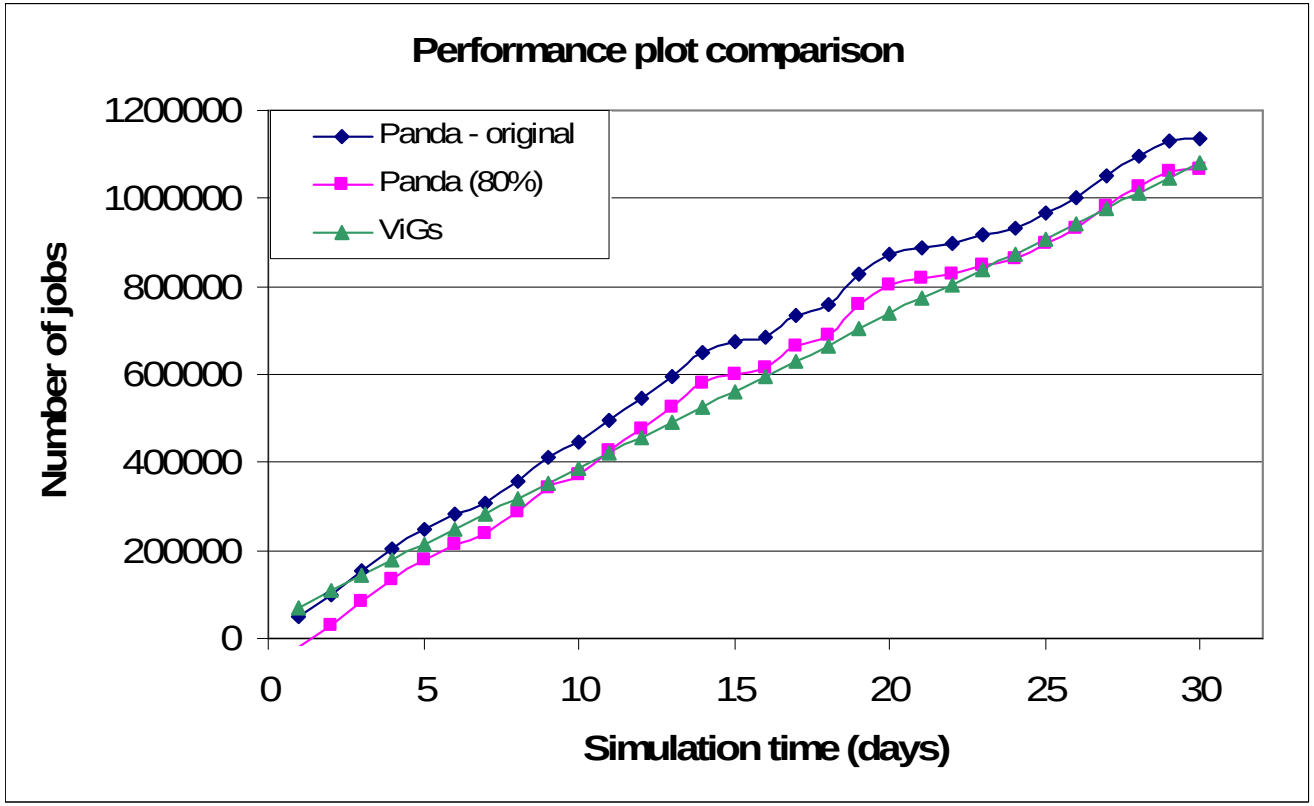
PanDA server one-month plot

ViGs – simulation sample



ViGs simulation results plot

ViGs – comparison with PanDA



PanDA vs ViGs plot

Conclusion

- ViGs is showing great promise as a platform for testing and studying numerous aspects of the PanDA Production system
- Eliminates the need to allocate resources for testing issues such as scalability and other performance issues
- Availability of ViGs enables development and fine tuning of the PanDA system

Appendices – Sample applications

- Performed evaluation and design validation of the PanDA system through the use of ViGs

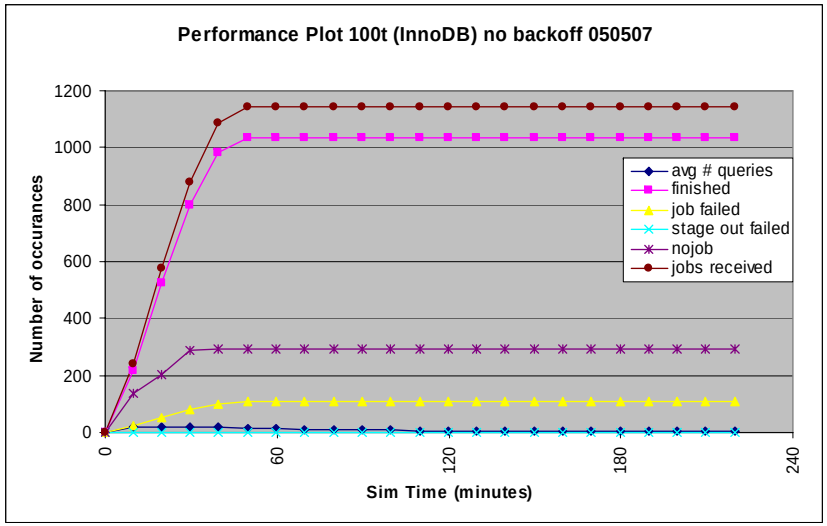


Figure 1. Performance results

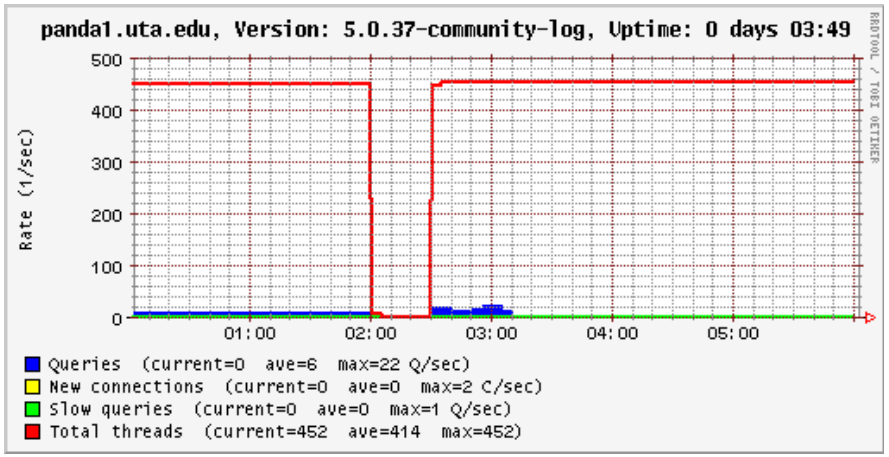
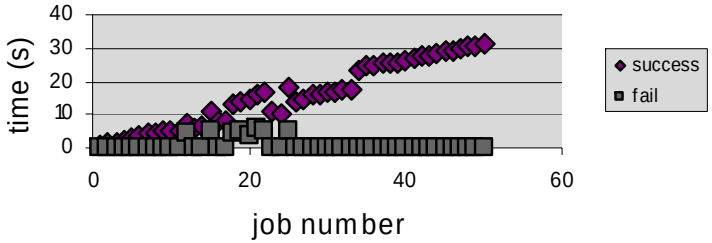


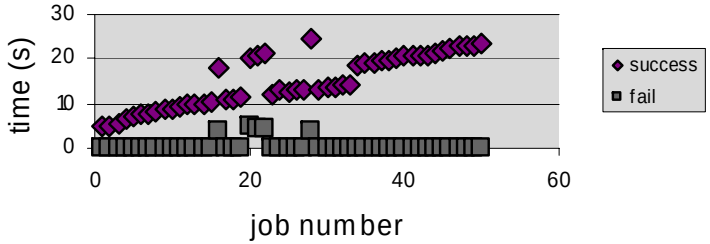
Figure 2. Corresponding DB analysis

Appendices – Sample applications

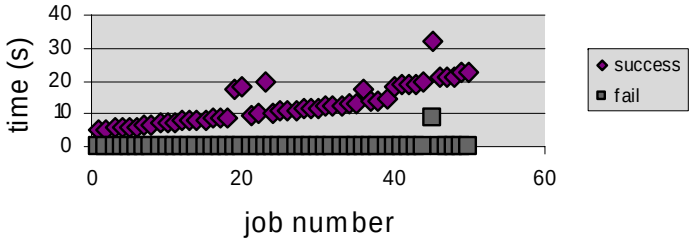
50 getJobs, max 50 http, 0.1s interval



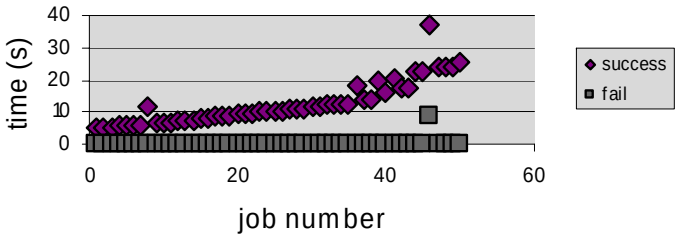
50 getJobs, max 50 http, 0.15s interval



50 getJobs, max 50 http, 0.2s interval



50 getJobs, max 50 http, 0.25s interval

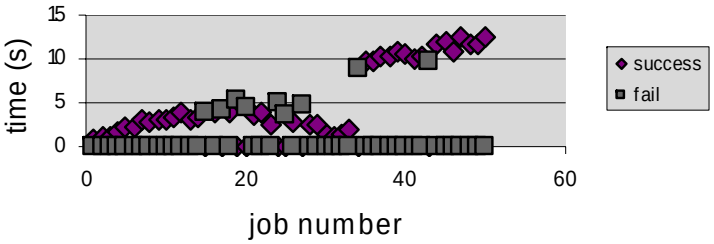


50 http connections, varying pilot job request intervals

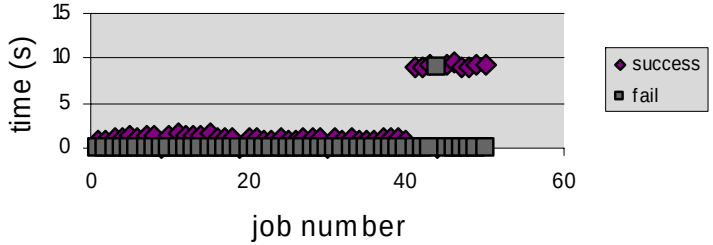
Appendices – Sample applications

50 getJobs, max 150 http, 0.1s interval

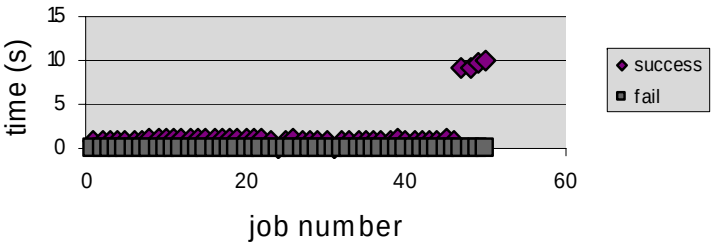
50 getJobs, max 150 http, 0.15s interval



50 getJobs, max 150 http, 0.2s interval



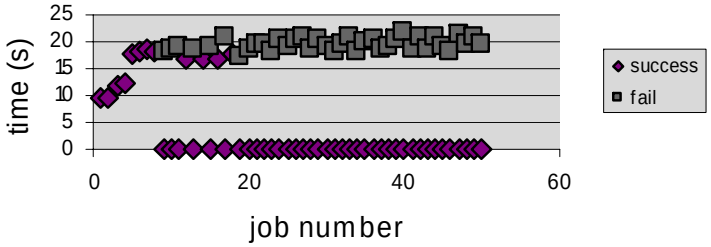
50 getJobs, max 150 http, 0.25s interval



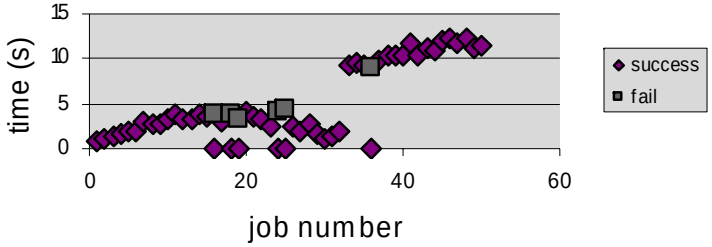
150 http connections, varying pilot job request intervals

Appendices – Sample applications

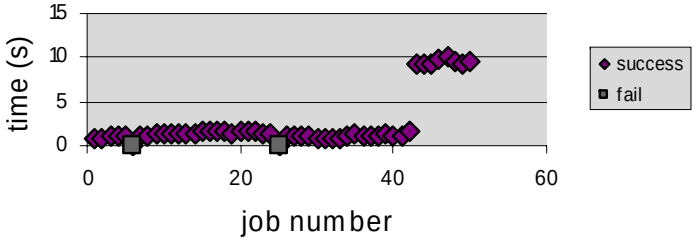
50 getJobs, max 256 http, 0.1s interval



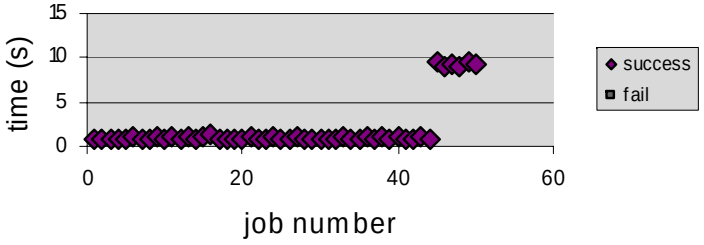
50 getJobs, max 256 http, 0.15s interval



50 getJobs, max 256 http, 0.2s interval



50 getJobs, max 256 http, 0.25s interval



256 http connections, varying pilot job request intervals