NAF Status Report

Untertitel der Präsentation

Yves Kemp
NUC, 11.06.2020

HELMHOLTZ SPITZENFORSCHUNG FÜR GROSSE HERAUSFORDERUNGEN

DESY.

Grid UI on EL7 systems

- History:
  - the Grid UI (voms-proxy-* etc) was set up from a CVMFS location at CERN
  - Problem: relying on a 2.7 Python
  - Conflicting with user setup of 3x Python
- Last month:
  - Grid UI setup remains like this on SL6 (since Python 2.7 anyway, and will disappear in 6 month)
  - Grid UI setup has been **removed on EL7 (WGS)**
  - `Note: Grid commands can be set up via 'source /cvmfs/grid.desy.de/etc/profile.d/grid-ui-env.sh'`
- Next week
  - Removing the automated Grid UI setup on **EL7 WNs** also

Covid-19 computing, the general and the special case

- General:
  - DESY contributing via Folding@Home and Rosetta@Home, on its Grid ressources in HH and ZN, partly via the WLCG combined effort of ATLAS and CMS. In addition, F@H and R@H are run as backfil in the HPC clusters Maxwell/HH and PAX/ZN. Zeuthen also provides some backfill on GPUs
  - The Maxwell cluster also hosts analysis of Corona virus X-ray data from Petra-III
  - ~~There is a news article in preparation to be published on www.desy.de~~ Done
- Special case: the NAF
  - Since GPUs are much more efficient at the current computational tasks, the NAF GPUs will be made available
  - Status: Dedicated Arc-CE has been set up, and integrated into NAF/BIRD. Currently undergoing final tweaking

Article: Accessible via www.desy.de

- https://www.desy.de/aktuelles/news_suche/index_ger.html?openDirectAnchor=1833&two_columns=0

2020/05/26

Back

Computer vs COVID-19

DESY's IT departments compute for corona research

Everyone knows the eerie visualisation of the coronavirus ... but what do its protein structures and protein bindings really look like? Which known drugs could possibly dock to these structures? Answers to these questions might considerably accelerate the development of therapies against COVID-19. DESY's IT department supports this cutting-edge research with several projects by providing computer power and know-how.

Download [490KB, 1772 x 1181]
DESY computer centre. Image: DESY, Heiner Müller-Elsner

A substantial part of the investigations on the coronavirus SARS-CoV-2 is now carried out digitally, with the help of computers. The interactions of different active substances with the proteins of the corona virus are simulated, which requires a lot of computing power. There are even projects that can be supported with your own computer: Rosetta@home and folding@home. The projects have in common that a big, complex problem such as the deciphering of proteins is broken down into small subtasks that are then distributed independently to the computers involved. The result of the calculation of a subtask is transmitted back to the central server, which then controls further calculations. Users can select a specific problem, and currently the most popular problem is the deciphering of the protein structures of the coronavirus.

Up next:

- The whole activity has gained visibility also within the directorate
  - Difficult to explain why DESY participates in several activities, and also goes under the umbrella of WLCG/CERN
- Wish to come up with another description, e.g. a „counter“ on a webpage with something like:

Covid-19 Compute @ DESY

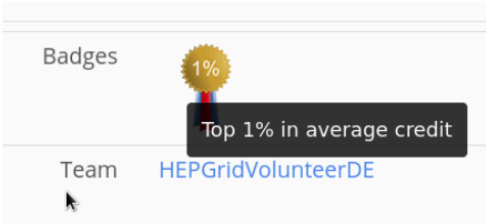
Contribution on 9.6.2020: NN CPU-Hours (equivalent to MM laptops) ... X% of the central DESY compute power

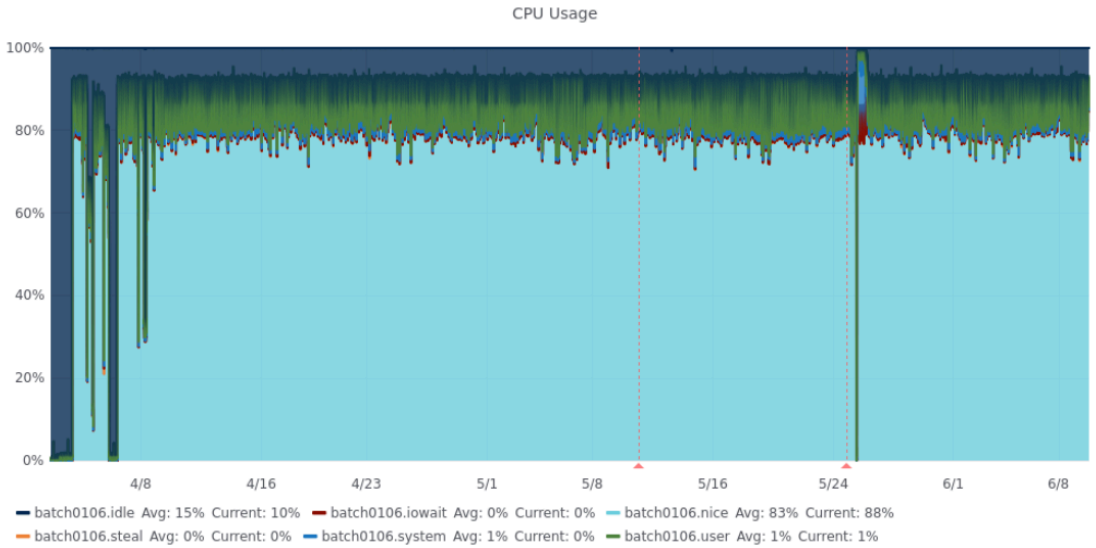
Contribution since 1.4.2020: XX CPU-Hours

rosetta@home @ DESY

(Thomas Hartmann)

- Running protein folding simulations via BOINC
  - build Singularity container on CVMFS for easier deployment
- dedicated out-of-warranty nodes with ~500 cores to Rosetta@home
  - DESY significant contributor
  - DESY-HH provided ~35000 CPUh since April





e.g., utilization of batch0106

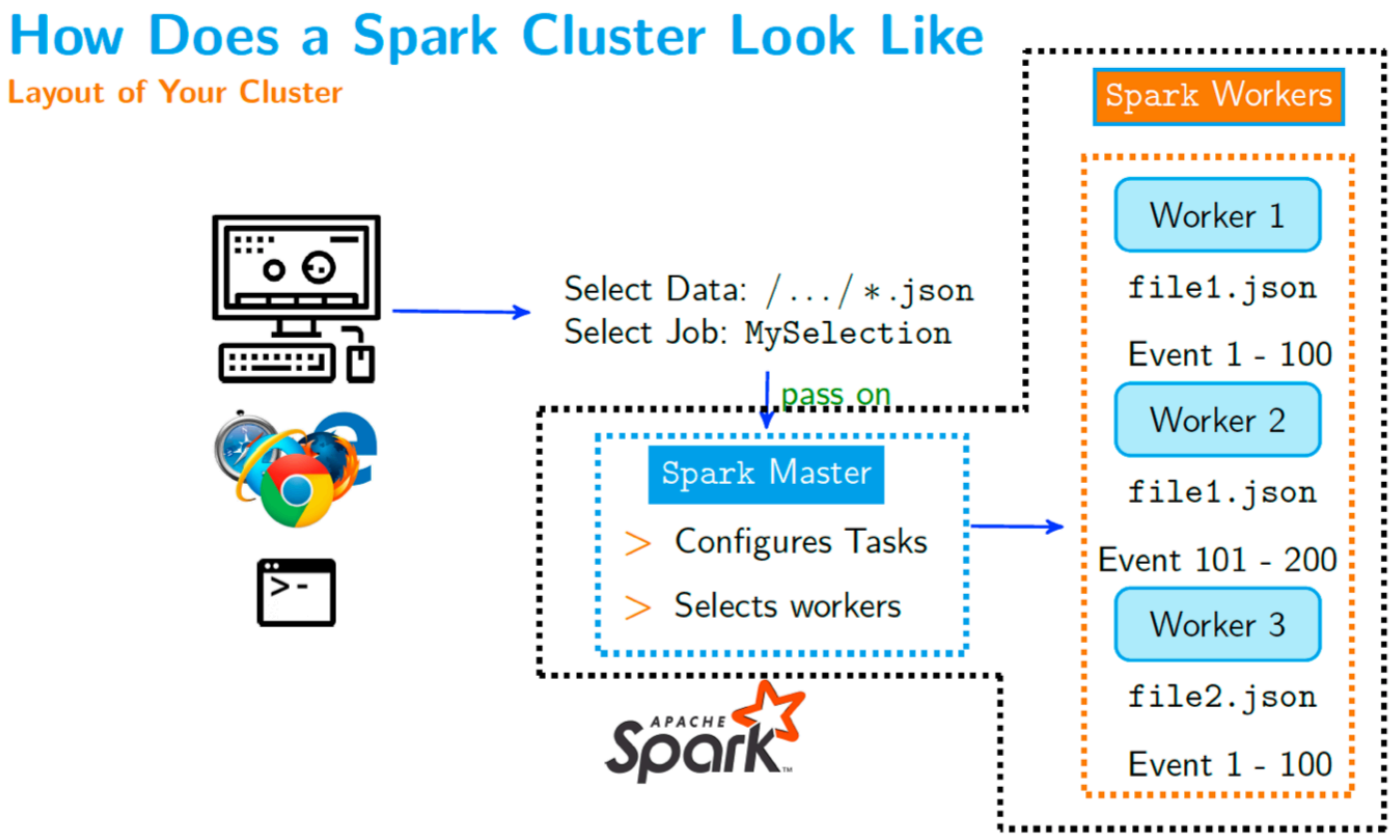
May: DESY & Friends on position 74

58	57	57	66	Chinese Dream	754,459,934	722,225	5,084,498	26,134,986	799,367	-
24	24	24	67	Microsoft	11,467,569,294	1,140,987	7,437,259	25,540,972	947,588	-
51	51	51	68	Raccoon Lovers	154,630,233,131	215,173	1,766,534	25,081,285	487,435	365+
5	5	5	69	BP6/VP6 User Group	2,375,593,756	668,715	5,093,925	22,637,808	722,628	365+
418	349	335	70	WALTER KLAES	78,239,584	569,638	5,762,800	20,715,334	753,046	-
108	104	103	71	Team Biss!	337,704,149	645,167	4,856,058	17,777,156	650,196	365+
80	77	77	72	MacRumors.com	478,118,751	523,950	4,134,116	16,851,733	562,781	365+
59	57	57	73	Cisco Systems	659,144,847	1,215,627	7,754,853	16,671,244	807,345	-
787	600	562	74	HEPGridVolunteerDE	42,680,791	668,422	5,415,168	16,472,948	647,768	-
3	3	3	75	PC Format	2,856,739,309	523,371	3,856,717	16,367,484	553,679	365+
93	93	92	76	Team-Goobee.org	371,124,010	533,330	2,962,128	16,183,675	470,273	-
42	42	42	77	BOINC World	16,513,166,813	721,814	3,700,514	16,134,380	538,563	-

Apache Spark on Condor

(Thomas Hartmann)

- Christian working on Apache Spark deployment
  - interactive data analysis/processing platform
  - Spark workers and heads in Singularity containers
  - starting additional workers as Condor jobs to dynamically extend compute power
  - parallelized I/O
  - interactive/parallel processing
  - Spark used for dCache log stat processing



Citing the DESY compute infrastructure

- Members of Helmholtz institutes must cite LK-II infrastructures used for their work
  - (that is at least what I understood ... I might be wrong on the details)
- Citation: A real journal, not only conference proceeding

Working on the reference:

- The NAF (and DESY Grid, and Maxwell, and DESY dCache infrastructure) do not have such a reference
- Plan: Expand the IDAF/beyond HEP paper for CHEP (combining Grid & NAF & Maxwell & associated storage) to something that can go into Springer Computing and Software for Big Science

And then getting people to use the citation ...