



Netherlands
Bioinformatics
Centre

NBIC Galaxy@HPC Cloud

Mattias de Hollander
David van Enckevort
Leon Mei
Marc van Driel
Rob Hooft



SURFsara HPC Cloud

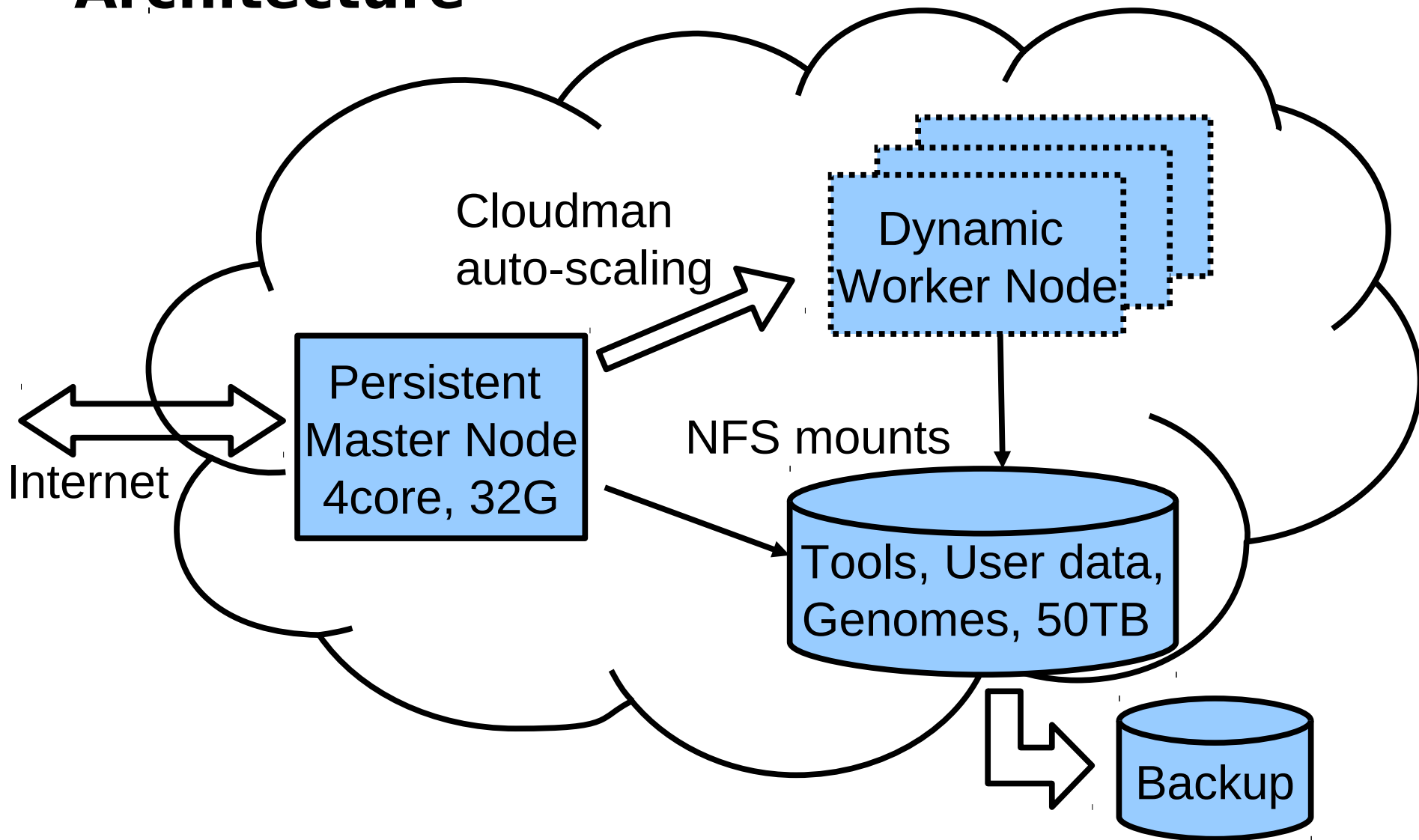
- National computing and networking center
- 19 nodes, 32 cores and 256 GB RAM each
 - Intel 2.13 GHz (Xeon-E7 "Westmere-EX")
- 400 TB storage in total
- Maximum 40 Gb access to storage (per node)
- Lightpath connection (1Gb->10Gb) between all Dutch universities & medical centers, BGI?

<https://www.surfsara.nl/systems/hpc-cloud>

Migration to Cloud

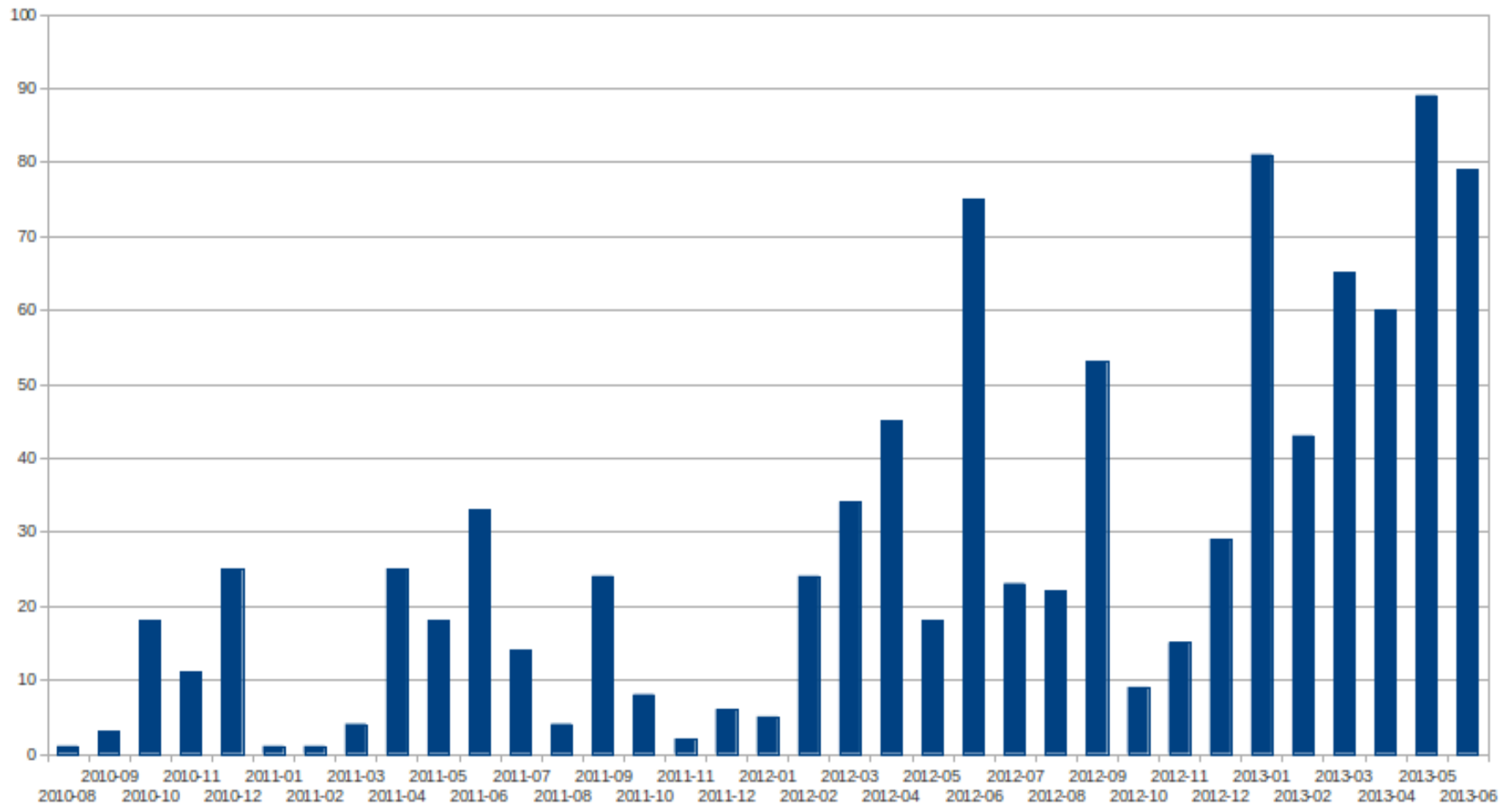
- Galaxy2.nbiceng.net
 - 4 CPU, 24G RAM, 1.5T HD, 100Mbit Internet
- Migration project July-September 2012
- Supported by BiGGrid, SURFsara, NBIC, NIOO
- Used as the base for other project specific Galaxy servers in the HPC cloud
 - CTMM-TraIT, CGD

Architecture



Usage

#new users per month



Portal To Other Resource

The screenshot shows the Galaxy / NBIC web portal interface. The browser window is titled "Galaxy - Mozilla Firefox" and the address bar shows "galaxy.nbic.nl". The main navigation bar includes "Galaxy / NBIC", "Analyze Data", "Workflow", "Shared Data", "Visualization", "Admin", "Help", and "User". A status bar at the top right indicates "Using 6.0 Gb".

The main content area is titled "Cluster Status" and displays a table with the following data:

Chromosome	Job ID	Status	Progress	Runtime (min)
3	7121414	Running	<div style="width: 23%; background-color: green; border: 1px solid gray;"></div> 23%	00:14

Below the table, the job status is summarized: "Job status (1 running, 0 queued, 0 finished, 1 total) (data is refreshed every 90 seconds)".

The left sidebar contains a "Tools" menu with various categories and links, including "VCF Tools", "NGS: Bedtools", "NGS Taskforce: Hubrecht - Alignment tool benchmarking", "NGS Taskforce: WUR denovo benchmarking", "NGS Taskforce: LUMC - GAPSS v2", "Haplotype sharing", "NGS Taskforce: LUMC - GAPSS v3", "NGS Taskforce: LUMC - deepSAGE", "NPC: msCompare", and "CTMM TraIT".

The right sidebar contains a "History" panel with a list of recent jobs:

- genallice meeting (7.1 Mb)
- 104: GRONCROSS Combined Output on data 102 and data 91 (Job is currently running)
- 103: GRONCROSS on data 102 and data 91 (html report) (Job is currently running)
- 102: Logfile from data 91 (2 lines, format: tabular, database: ?)

The browser's address bar at the bottom shows the URL: "galaxy.nbic.nl/tool_runner?tool_id=CROSS".

Some Hurdles

- Installation using Cloudman
 - Fabric scripts use hardcoded versions, some of them are outdated
- Too many layers so sometimes hard to find the error
 - Tool, Galaxy, VM or Cloud
- I/O Performance issue
- MySQL to PostgreSQL: py-mysql2pgsql