## horizontal line

HPC cluster *Gearshift*

2017-06-11 • hpc-3.0-beta

**─**

UMCG Research IT program

# Summary

This document describes the architecture for a new High Performance Computing (HPC) cluster named *Gearshift* and built for the UMCG research IT program. This will be version 3.x of the HPC architecture and *Gearshift* will replace one of the current research clusters named *Calculon* and *Boxy* (version 2.x) whose hardware will be EOL later this year. Later a second new cluster based on the version 3.x architecture will replace the second cluster to make sure we maintain fall-back capacity in case of disasters.

# Goals

1. High Performance Computing & Storage for all researchers @ UMCG.
2. Researchers @ UMCG can easily collaborate with researchers from elsewhere in (inter)national consortia.
3. At least 2 “self-contained” clusters in two different data centers to enable fall-back to the other site in case of (un)scheduled maintenance and to reduce the risk of losing 100% capacity.
4. Architecture facilitates best practice data management according to *FAIR* principles.
5. Architecture facilitates accounting of resources (for both volatile resources like core hours and non-volatile resources like disk space).
6. Architecture is auditable, compliant with UMCG IT guidelines and enables (ISO/NEN) certification.

[**Summary**](#_au51mny0sx6) **0**

[**Goals**](#_3at9u9s4e0vp) **0**

[**Specifications**](#_l6lglw3m1tnt) **1**

[Authentication](#_h1sciv8md2vv) 1

[**Milestones**](#_yyrhu7ml5bea) **1**

[Lorem ipsum](#_buwz1tcz7y35) 1

[Dolor sit amet](#_p2nityf5kx5q) 1

# Specifications

## Machine types and naming schemes

The naming scheme will use the same themes as the used for the current research clusters ***Boxy*** and ***Calculon***. Machines with a proper DNS entry and hence fully qualified domain name

* Will **not** be named after organizational units, because they tend change.
* Will **not** be named after specific functionality, because functionality tends to be added/removed and is not very specific resulting in too common names.
* Will be named according to a theme; see below.

Machines without a fully qualified domain name and which are used only “internally” within a cluster

* Will be named using combination of a two character prefix derived from the cluster name and the function of the machine.

Type of machines:

* User Interface (UI):
  + For logins by (end)users
  + Slurm tools/commands for submitting and monitoring of jobs
  + Read-only access to software, modules and reference data deployed with EasyBuild+Lmod in /apps/…
  + Access to (parts of) large shared parallel file systems (with root squash)
* Deploy Admin Interface (DAI):
  + For login by deploy admins only.
  + Read-write access to software, modules and reference data deployed with EasyBuild+Lmod in /apps/...
  + No slurm tools/commands installed, so no accidental job management.
  + No access to large shared parallel file systems
* Sys Admin Interface (SAI)
  + For sys admins only (like “winterpeen” for Calculon/Boxy clusters)
  + For monitoring like quota reports, slurm usage report and other cron jobs.
  + Access to (complete) large shared parallel file systems (without root squash)
* Proxy
  + For multi-hop SSH logins from “outside”
  + Nothings else: no job management, no mounts of shared file systems, no homes, no /apps, etc.
* Scheduler
* Job Database
* Accounts
  + LDAP with users entitled for Gearshift cluster <- IDVault <- COmanage
* Compute Nodes

Themes:

* Research cluster and primary UI: robot from the [Futurama scifi sitcom](http://futurama.wikia.com/wiki/Category:Robots)
* Admin Interfaces: root/carrot varieties
* Proxy for access from outside: rooms preceding another room

The new research cluster will be named after the Futurama robot character ***Gearshift****.* This will be also the name of the primary User Interface (UI) server. The Proxy will be named ***airlock***, the Deploy Admin Interface (DAI) ***sugarsnax*** and the Sys Admin Interface (SAI) ***imperator***.

Other machines that are part of the cluster and only accessible using internal network interfaces (schedulers, compute nodes, account servers, etc.) will use the two character prefix ***gs*** followed by a dash and the function of the machine.

For a complete overview of machine names see the [Gearshift Inventory](https://docs.google.com/spreadsheets/d/1dqoXJv9HMDewTDBvlr9LmZUXO8izaEcchSKM59lUJMY/edit#gid=0).

## Platform

The following software frameworks will be used:

* Deployment management: xCAT and Ansible Playbooks
* VM OS: CentOS Linux 7.3
* Openstack OS: Ubuntu 16.04 LTS
* Module system: Lmod 7.x
* (RPM) Package management by sys admins: spacewalk
* Bioinformatics package management by deploy admins: EasyBuild 3.x
* Virtualization: OpenStack Ocata
* Scheduler: SLURM 17.x

## Domain and data center

Gearshift is physically located @ DCE and in the ***\*.hpc.rug.nl*** network domain.

How does this affect the ability to interface with other infra like workspaces, UMCG Isilon bulk storage will be excluded from this project, iRODs based long term storage @ RUG CIT, backup services, etc.??

## Authentication & Authorization

For managing user accounts and access to groups + entitlements we’ll exclusively use a COmanage portal, which can be accessed with federated identity using “home institute” accounts via SURFconext (SAML protocol).

The ldap authentication will use the central loadbalanced LDAP service.

Once a user has

1. created an account in the COmanage self-service portal
2. and has been accepted as member of at least one group, which has been entitled for the *Gearshift* cluster

he/she can login using SSH:

* SSH logins from one machine of the UMCG research IT clusters to another in the same environment will use single factor auth with personal accounts and public-private key pairs.
* (Password based logins will be disabled disabled).
* SSH logins from outside the UMCG research IT clusters to one that is part of the UMCG research IT clusters require multi hop SSH via security hardened Proxy server(s), which act as “stepping stone” and which use two factor auth.
  + Transparent multi hop SSH (<http://wiki.gcc.rug.nl/wiki/TransparentMultiHopSSH>)
  + Dedicated proxy VM named: *airlock*
  + Two factor auth for access to Proxy with public key + SMS token

## Deploy Management

* Ansible playbooks + config templates in Gogs repo
* Gogs server hosted by RUG CIT: servername, reponame, …
* Workflow with personal online forks and pull request to merge code back into the *“blessed master”* branch.
* Push to *“blessed master”* disabled on server-side.
* Review of pull requests enforced: users cannot review their own pull request and review is required before a pull request can be merged.
* Deployment of bioinformatics software using EasyBuild easyconfigs
  + Depad group with power users can deploy with EasyBuild and without root access in /apps
  + Primary /apps read-write mounted on **local** storage system of a dedicated VM for deploy admins: ***sugarsnax***
  + Rsynced secondary /apps on prm file system mounted read-only on UI.

Rsynced secondary /apps on tmp file system mounted read-only on compute nodes.

## Monitoring

...

# Roadmap

|  |  |  |  |
| --- | --- | --- | --- |
| What | When | Who | Done |
| Make architecture with roadmap | 26-10-17 | WN/ER/PN |  |
| IP’s in idracs zetten | 16-10-17 | WN/ER |  |
| Network link to junipers @DCE |  | CIT |  |
| Complete PIA and have architecture reviewed by RUG privacy officer Arjen Deenen <a.r.deenen@rug.nl> |  | HJZ |  |
| Review of architecture by UMCG staff (Incl. PIA + BIA -> RI) |  | UMCG |  |
| Create git repo and add users/team | 19-10-17 | ER |  |
| Create Ansible playbooks to deploy openstack cluster | 19-10-17 | WN/ER |  |
| Create Ansible playbooks to deploy compute cluster | 2-11 | WN/ER |  |
| Create/reserve: VLANs, IP plan, DNS entries | 26-10 | WN |  |
| Configure Isilon storage as tmp | ? | WN/ER/UMCG |  |
| Create playbooks for various VMs (UI, proxy, SAI, DAI, scheduler.) | 22-12 | WN/ER |  |
| Create enrollment/authorization workflow in COManage | 19-10 | PN |  |
| Connect COmanage with research IDV |  | IAM |  |
| Create mount configs for hpc-environment script (or replace it’s functionality with something else) | ? | CIT/UMCG |  |
| Create Ansible playbooks to deploy bioinformatics software with EasyBuild | 26-10 | PN |  |
| Roll out the complete cluster automated |  | All |  |

# Supplements