Soleil Computing and deployment

IT infrastructure and containerization

Alain BUTEAU : Leader of I.T infrastructures group
Patrick MADELA : CI/CD and dev. engineer

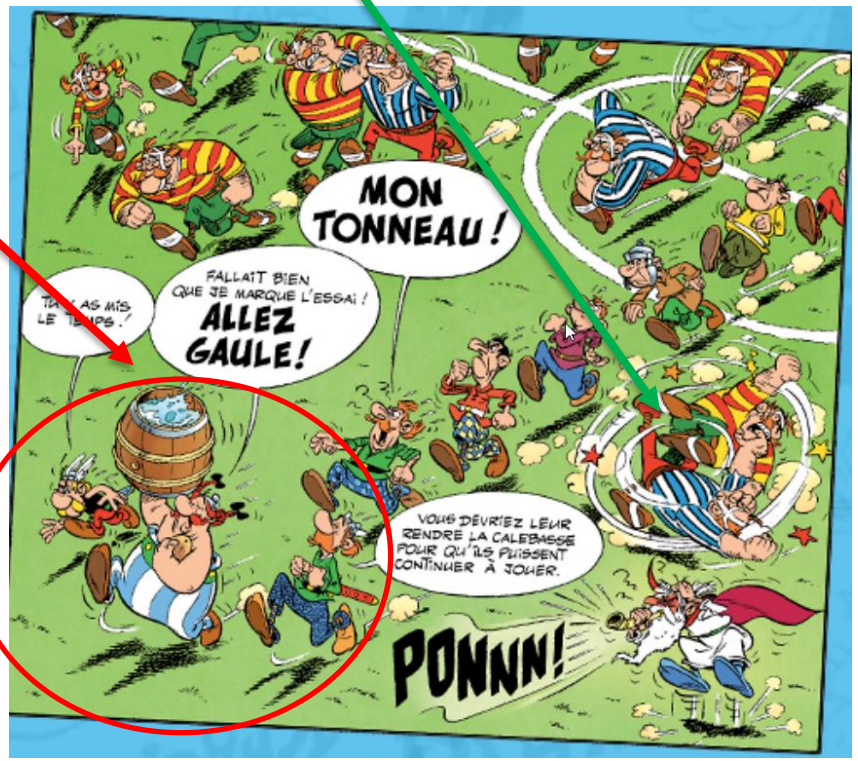
SOLEIL II
La science éclaire l'avenir
Science lights up the future

Summary

The I.T infrastructure at SOLEIL a couple of years ago

« Happy » users

I.T Infra team



Data center architecture

DataCenter Current Architecture

- SOLEIL has 2 datacenters : RGI1 and RGI2
  - RGI1 in office building
  - RGI2 in synchrotron building
- In 2022 we started an audit with an external company (called APL)
  - **RGI1 :**
    - Modernization started in 2019/2020 with a “cold corridor” containment done
    - Cooling capacity is **76KW** “cold”
      - An extra redundancy cooling system has been specified . Call for tender is on going
    - 16 racks extensible to 24 racks
  - RGI2
    - No possibility to have a “cold corridor” (because of technical sub-floor)
    - Cooling capacity in “degraded mode” is the limitation (**49KW**)
    - Installation of “hot corridor” completed . Next step is to populate the crates

	Situation actuelle	
	RGI 1	RGI 2
ÉNERGIE		
P ondulée max (kW)	80	320
P ondulée consommée (kW)	20	15
P normale consommée (kW)	15	13
P max théorique de la salle (kW)	**80**	**320**
P consommée de la salle (kW)	35	28
P élec disponible (kW)	45	292
Niveau de résilience	2N	2N
Taux de charge élec (%)	44%	9%
Seuil prudentiel (90%) en kW	72	288
Seuil critique (95%) en kW	76	76
REFROIDISSEMENT		
P Clim max en DD (kWf)	36	49
P Clim sur EG (kWf)	0	120
Apports calorifiques actuels	6	2
P max théorique de la salle (kWf)	**30**	**47**
P consommée de la salle (kWf)	35	28
P froid disponible (kW)	-5	19
Niveau de résilience	N+1	N+1
Taux de charge clim (%)	118%	60%
Seuil prudentiel (90%) en kWf	27	42
Seuil critique (95%) en kWf	28	45

Perte de la résilience froid

RGI1

RGI2 (before)

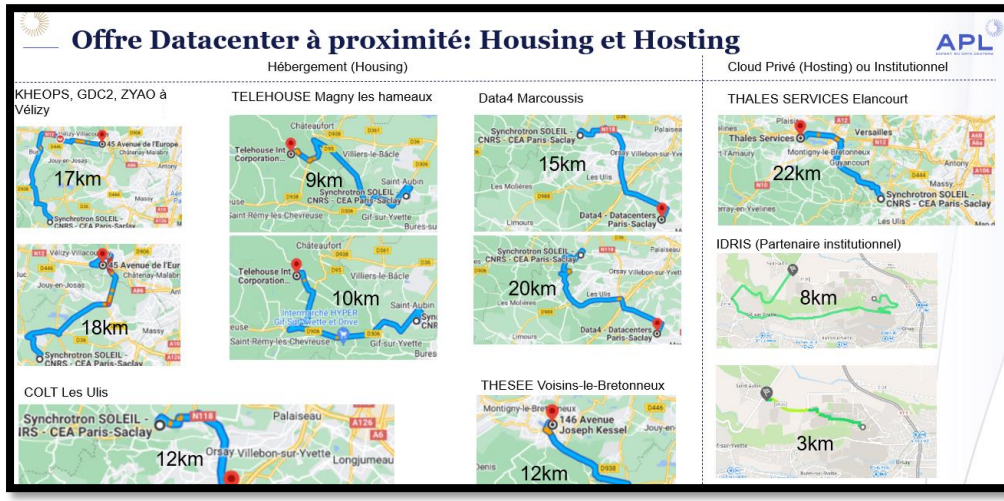
RGI2 (after)

SOLEIL II
La science éclaire l'avenir
Science lights up the future

DataCenter : The futur

- The strategy is to have a hybrid approach
  - Mix of “On premise” and “Cloud”

- Depending on the projection for the capacity planning (storage and HPC nodes) several options will have to be studied
  - Option 1 Cloud hosting either with commercial companies and/or “public” research entities
  - Option 2 : Setting up a 3rd on premise datacenter
    - Either in an existing SOLEIL building
    - Either with a container approach

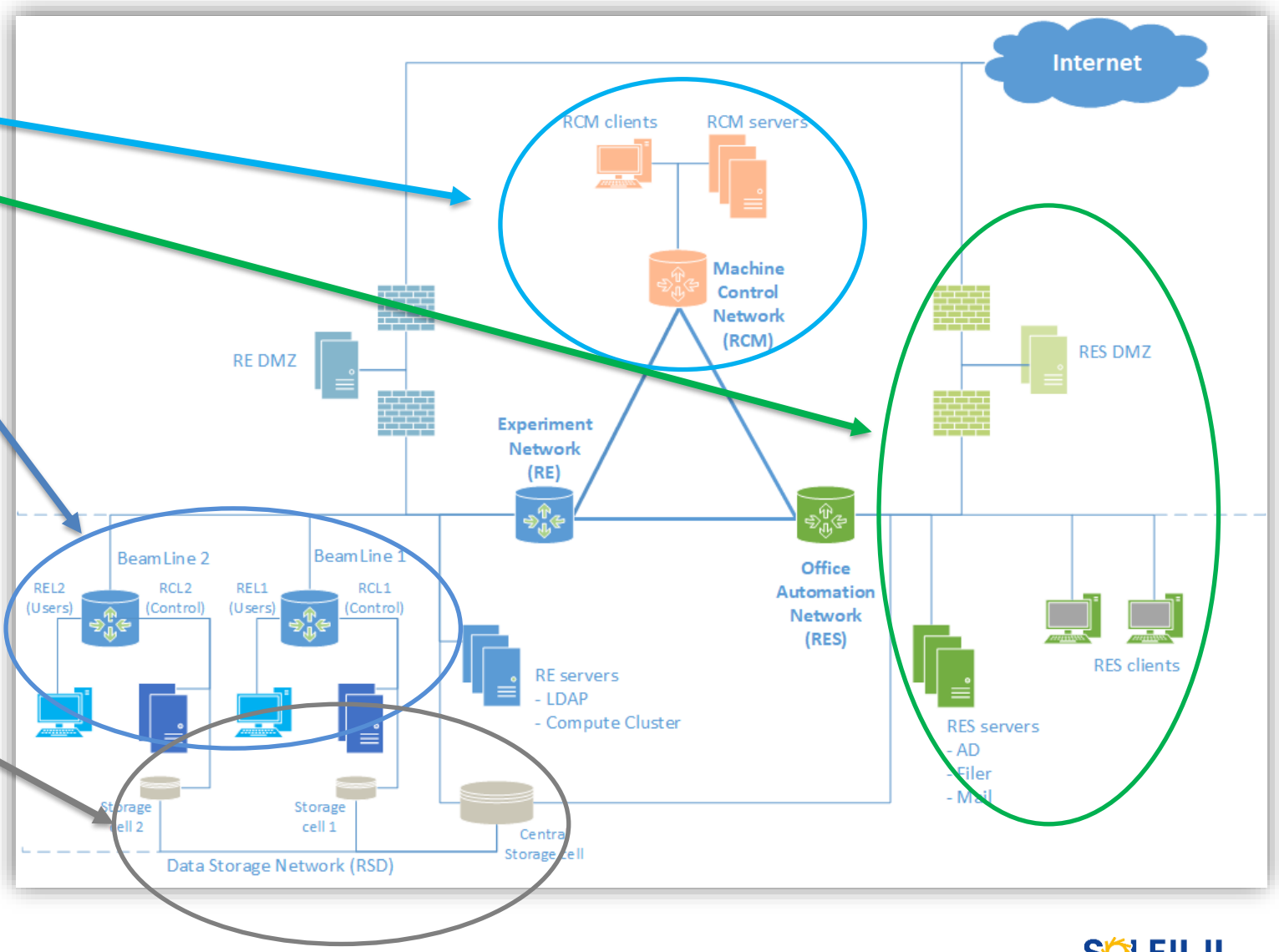


- Remark :
  - Having an “On premise” 3rd datacenter also has operational advantages for High Availability mechanisms
    - A third point is mandatory for arbiter mechanisms to avoid split brain
    - It may have financial advantages as the loss of 1 “room” means only losing 33% of the compute capacity (instead of 50%) today

Network architecture

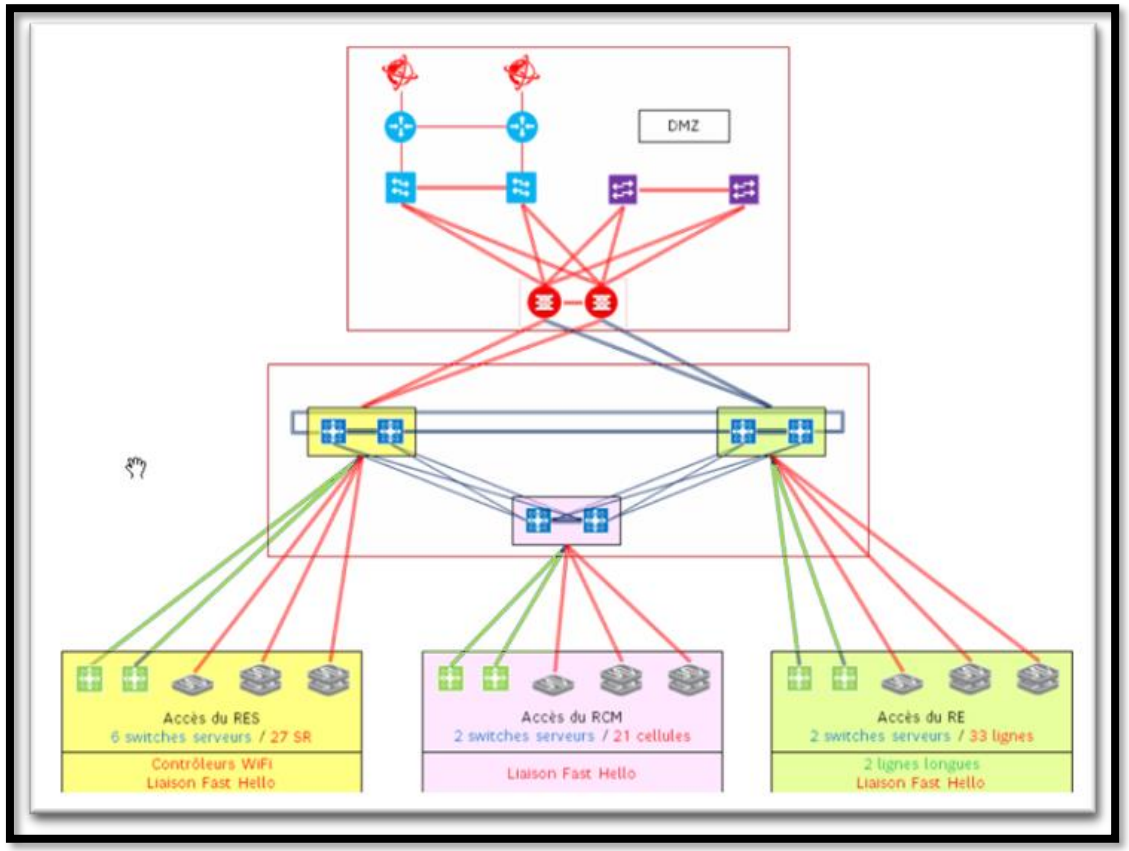
Network architecture: Current situation

- 3 majors network zones :
  - RCM : Accelerator network
  - RES : enterprise network
  - RE : Beamlines network
    - 2 VLANs per beamline
      - 1 for Controls RCLx
      - 1 for external users RELx
- 1 Dedicated Storage Network



Network architecture: Current situation

- Renewal of the core switches architecture has been done in January 2023 to have a backbone distribution of 10Gb up to 100Gb/s on all links:
  - **100 Gbs links (inter core switches)**
  - **1 or 10Gb/s (to BL/Machine local switches)**
  - **10 or 40Gbs (to datacenter TOR switches)**
- Today
  - Beamlines switches connections to backbone is 10GB/s
  - Machine : All distribution switches are still connected to backbone on 1GB/s links
- Next Steps
  - Design new VLANs (2025/2026/…..)
    - For security reason (“micro-segmentation”)
    - For operational reasons (like admins VLANS for subsystems)
  - Renew the distribution switches on Machine Cells or Beamlines (SOLEIL-II)
    - Depending on SOLEIL-II requirements

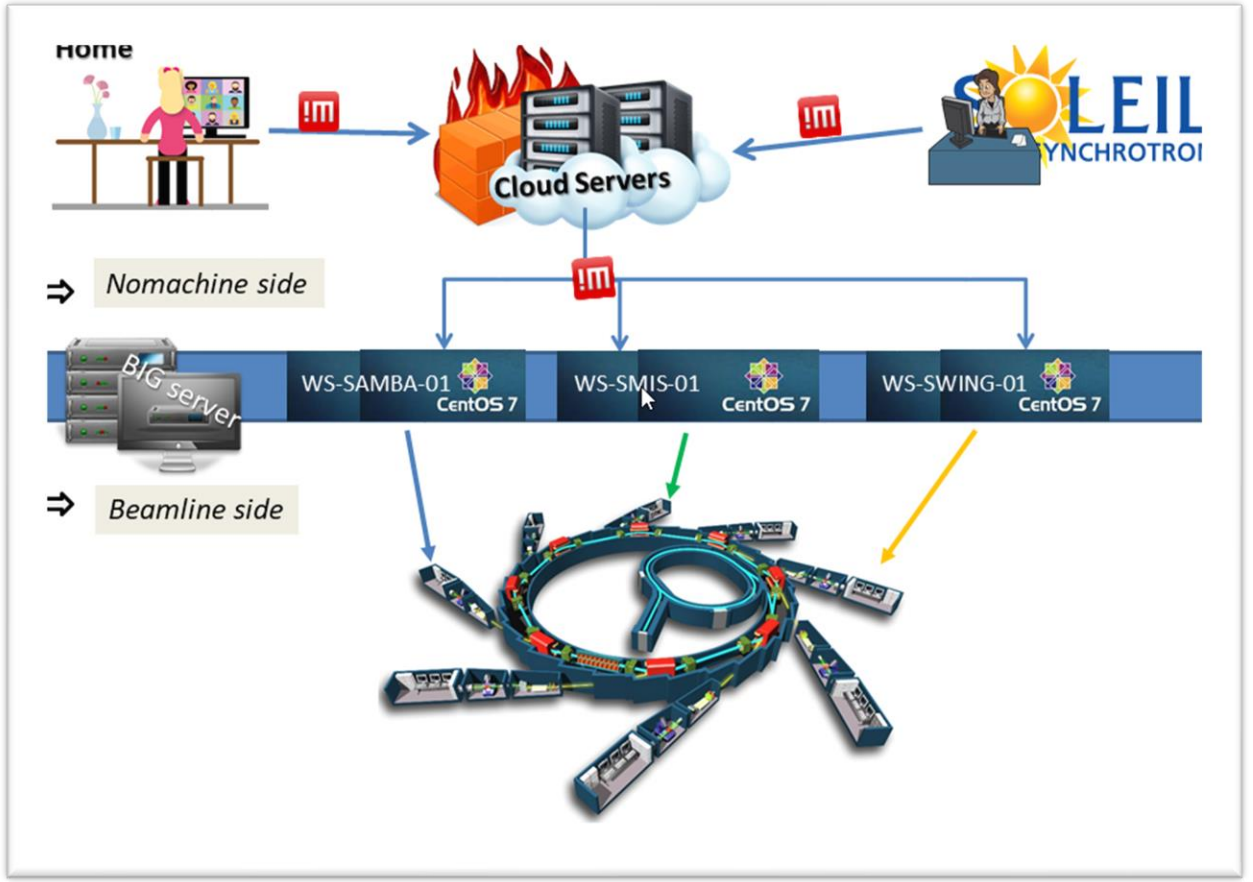


Remote access and CyberSecurity

Remote access

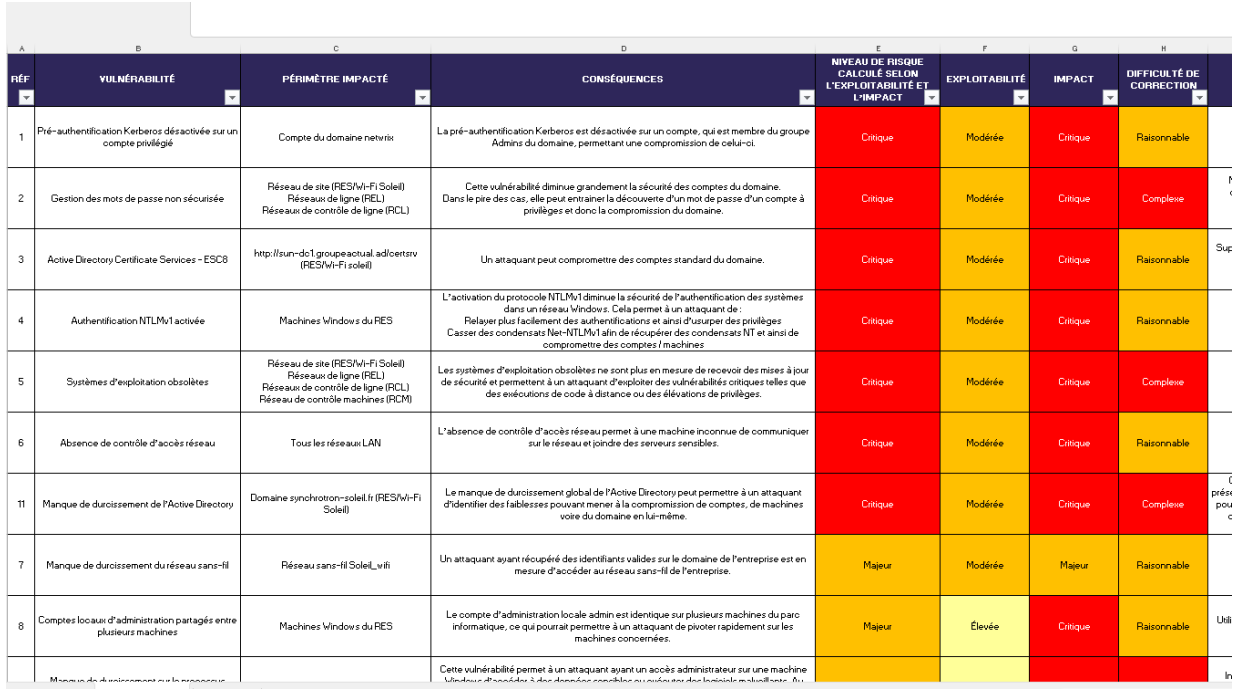
- Current situation
  - 3 different entry points depending on user status
    - For beamlines external users and SOLEIL beamline staff through the **NoMachine** solution
    - For SOLEIL technical teams (I.T and accelerators) access through another **NoMachine** infrastructure **AND** through **VPN**
    - For external suppliers : Access through the **WALLIX bastion**

- Future plans
  - Use the VISA solution for remote access (see “HPC/Data services”) for beamline users and staff access to Beamline



CyberSecurity

- Current situation
  - An audit by an external company has been done at the end of 2023
  - External pentest of the DMZ interface
- The actions plans mostly focuses on “enterprise I.T”
- Actions already started/completed includes
  - ActiveDirectory weaknesses corrections
  - Reinforcement of password policies for users/admin/services account
  - Enhance user account management (Onboarding/Offboarding processes)
  - Review of IT admins rights
  - Decommissioning of DMZ servers
  - Uniformization of DMZ reverse proxy architecture
- Short term planned actions (< 2024)
  - Setup MFA for SOLEIL enterprise applications
  - Review of network zones isolation
- Mid term planned actions (2025)
  - Get another external audit to check we progressed !! (and go deeper in vulnerabilities)
  - Deploy MFA for SOLEIL scientific applications (DUO, VISA, ..)
  - Implementation of WebAccessFirewall in the reverse proxy layer
  - Training program for SOLEIL employees AND IT people

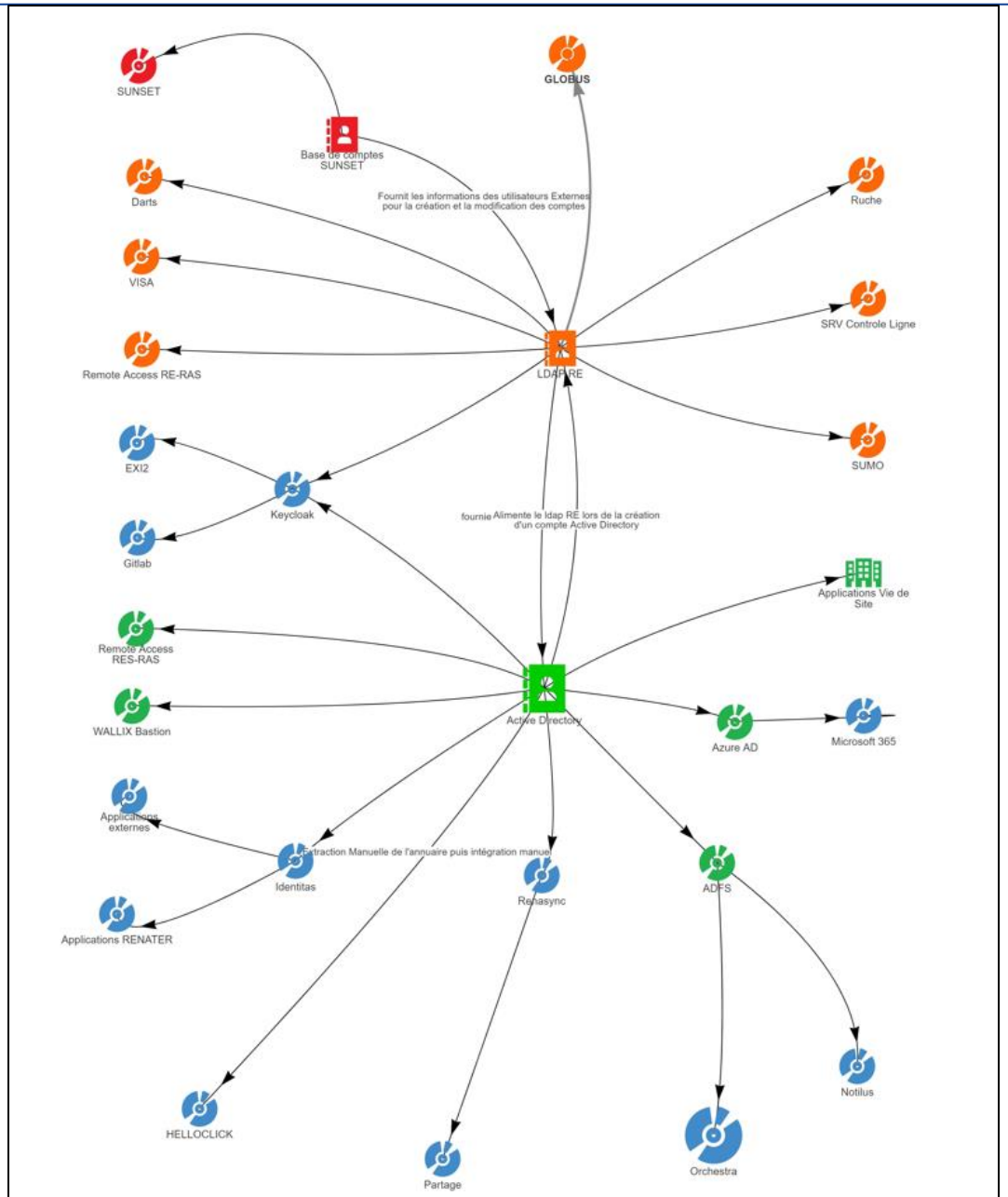


RÉF	VULNÉRABILITÉ	PÉRIMÈTRE IMPACTÉ	CONSÉQUENCES	NIVEAU DE RISQUE CALCULÉ SELON L'EXPLOITABILITÉ ET L'IMPACT	EXPLOITABILITÉ	IMPACT	DIFFICULTÉ DE CORRECTION
1	Pré-authentification Kerberos désactivée sur un compte privilégié	Compte du domaine netwrix	La pré-authentification Kerberos est désactivée sur un compte, qui est membre du groupe Admins du domaine, permettant une compromission de celui-ci.	Critique	Modérée	Critique	Raisonnable
2	Gestion des mots de passe non sécurisée	Réseau de site (RES/Wi-Fi Soleil) Réseaux de ligne (REL) Réseaux de contrôle de ligne (RCL)	Cette vulnérabilité diminue grandement la sécurité des comptes du domaine. Dans le pire des cas, elle peut entrainer la découverte d'un mot de passe d'un compte à privilèges et donc la compromission du domaine.	Critique	Modérée	Critique	Complexe
3	Active Directory Certificate Services - ESC8	http://sun-dc1.groupeactual.adl/certsrv (RES/Wi-Fi soleil)	Un attaquant peut compromettre des comptes standard du domaine.	Critique	Modérée	Critique	Raisonnable
4	Authentification NTLMv1 activée	Machines Windows du RES	L'activation du protocole NTLMv1 diminue la sécurité de l'authentification des systèmes dans un réseau Windows. Cela permet à un attaquant de : Relayer plus facilement des authentifications et ainsi d'usurper des privilèges Casser des condensats Net-NTLMv1 afin de récupérer des condensats NT et ainsi de compromettre des comptes / machines	Critique	Modérée	Critique	Raisonnable
5	Systèmes d'exploitation obsolètes	Réseau de site (RES/Wi-Fi Soleil) Réseaux de ligne (REL) Réseaux de contrôle de ligne (RCL) Réseau de contrôle machines (RCM)	Les systèmes d'exploitation obsolètes ne sont plus en mesure de recevoir des mises à jour de sécurité et permettent à un attaquant d'exploiter des vulnérabilités critiques telles que des exécutions de code à distance ou des élévations de privilèges.	Critique	Modérée	Critique	Complexe
6	Absence de contrôle d'accès réseau	Tous les réseaux LAN	L'absence de contrôle d'accès réseau permet à une machine inconnue de communiquer sur le réseau et joindre des serveurs sensibles.	Critique	Modérée	Critique	Raisonnable
11	Manque de durcissement de l'Active Directory	Domaine synchrotron-soleil.fr (RES/Wi-Fi Soleil)	Le manque de durcissement global de l'Active Directory peut permettre à un attaquant d'identifier des faiblesses pouvant mener à la compromission de comptes, de machines voire du domaine en lui-même.	Critique	Modérée	Critique	Complexe
7	Manque de durcissement du réseau sans-fil	Réseau sans-fil Soleil_wifi	Un attaquant ayant récupéré des identifiants valides sur le domaine de l'entreprise est en mesure d'accéder au réseau sans-fil de l'entreprise.	Majeur	Modérée	Majeur	Raisonnable
8	Comptes locaux d'administration partagés entre plusieurs machines	Machines Windows du RES	Le compte d'administration locale admin est identique sur plusieurs machines du parc informatique, ce qui pourrait permettre à un attaquant de pivoter rapidement sur les machines concernées.	Majeur	Élevée	Critique	Raisonnable
	Manque de durcissement sur le processus		Cette vulnérabilité permet à un attaquant ayant un accès administrateur sur une machine Windows d'accéder à des données sensibles ou exécuter des logiciels malveillants. Au				

Identity management

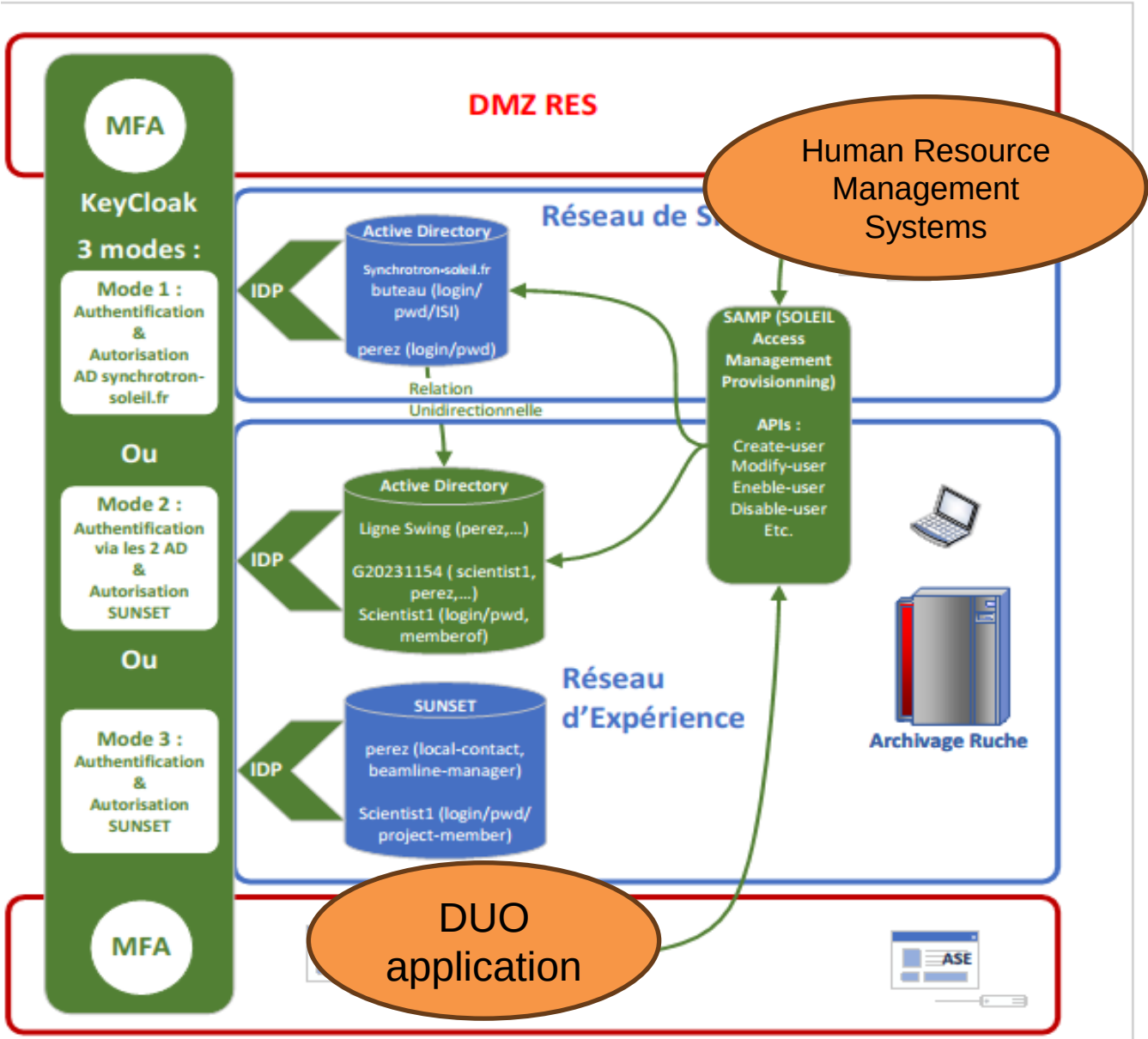
Identity Management : current status

- We have 2 directories :
  - 1 **ActiveDirectory** for SOLEIL staff
  - 1 **LDAP** for both SOLEIL staff (beamline and engineering staff) and external users
- LDAP suffers from sever defaults and poor data quality
  - Leak passwords, organizational structure not coherent with SOLEIL structure
  - Technical complexity for group management
  - Etc .
- ActiveDirectory data quality has been enhanced in 2023/2024
  - Tens of reviews
  - Clarification of onboarding process



Identity Management : Future plans

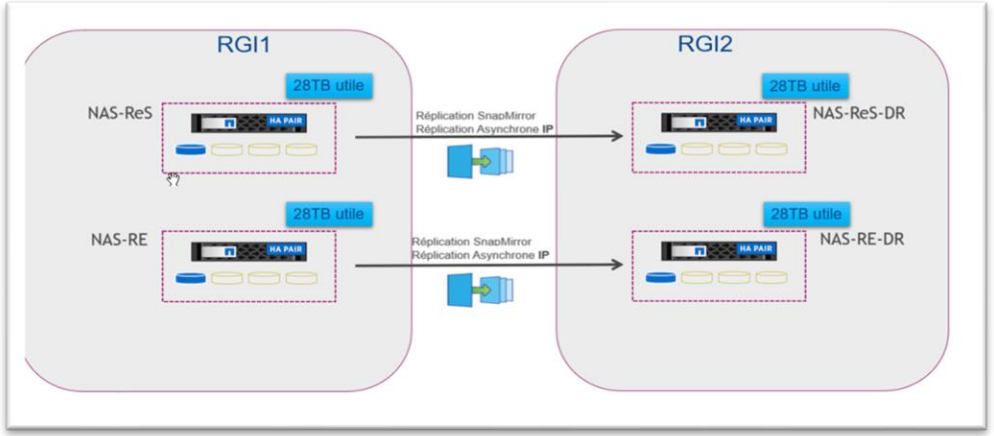
- We are targeting a unique directory based on our AD
  - Extended to “external users”
- Full Automation of Onboarding/Outboarding processes
  - Through webservices for :
    - user creation
    - user modification
    - user deletion
    - Etc ..
- Keycloak is deployed on top of this ActiveDirectory
  - For OpenID connect applications
  - To provide MFA mechanisms



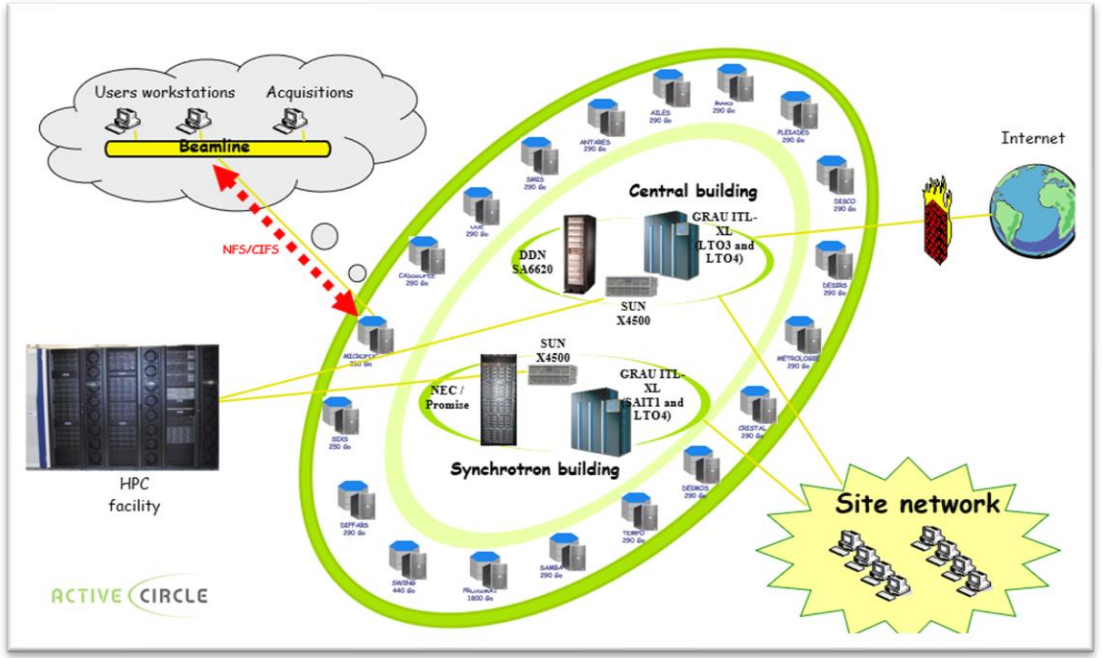
Storage architecture

Storage architecture : Current situation

- Netapp NAS used for :
  - Enterprise storage
  - Control systems (Machine and Beamlines) applications and Proxmox VM images
  - 4 NetApp nodes on 2 clusters
    - distributed on the 3 networks : RES , RE/RCM
    - Data are “siloed” thanks to the Netapp SVM (Storage Virtual Machine) mechanism

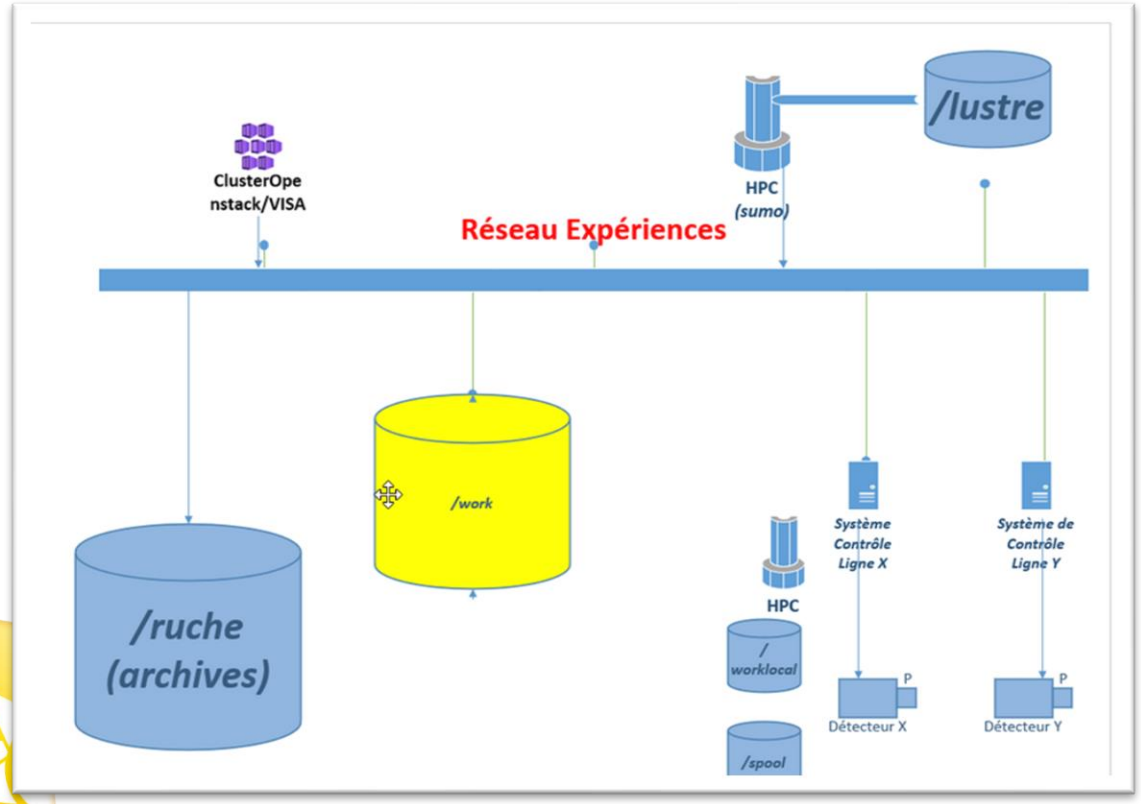


- For experimental data : a ”HSM like” solution based on the ActiveCircle product and distributed on
  - Each beamline has a HA cluster of 2 local storage nodes
  - 2 “large” (4PB and 6PB) capacity storage in the 2 datacenters
  - 1 LTO8 tape library with 550 slots and 10 drives
  - Through the FITS project we have a partial cloud copy :
    - 1 “online” copy hosted on the IDRIS datacenter
    - 1 “offline” copy hosted in Lyon on the IN2P3 infrastructure



Storage architecture : Future plans

- For experimental data : “Hot Data”
  - Provide a centralized high performance storage space (/work) for demaning “real time” calculations (from HPC or VISA)
  - We are foreseeing an IBM ESS 3500 GPFS based solution



- For experimental data : warm and cold data will be separated
  - Location on SOLEIL libraries and/or FITS IN2P3 infrastructures
- Datatransfer from “Hot data zone” ←→ “Warm/cold data zone”
  - Will be done through a “backup like solution” :
    - for instance with products such as bacula
  - Or a “data management” solution
    - Such as arcitecta product
  - Which means that data movement will not be automatic anymore like in our HSM solution

HPC services

HPC clusters : current status

- Current Computing infrastructure
  - **For simulations**
    - Use of CCRT (a French national supercomputing facility)
  - **For internal SOLEIL use and interactive calculations**
    - A small (and old/ 8 years!) NEC cluster :
      - 1 front end for interactive calculations
      - 13 (CPU/GPU) nodes accessible through SLURM
    - Local “mini HPC” resources on tomography beamlines
  - Deployment of a separated OpenStack infrastructure to provide a “Data Analysis” portal thanks to VISA application

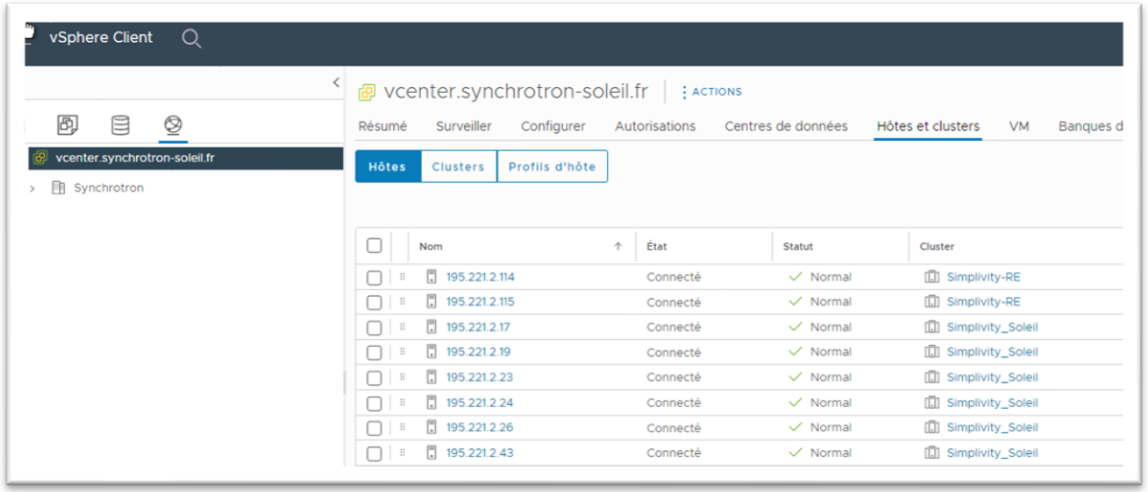
HPC clusters : future plans

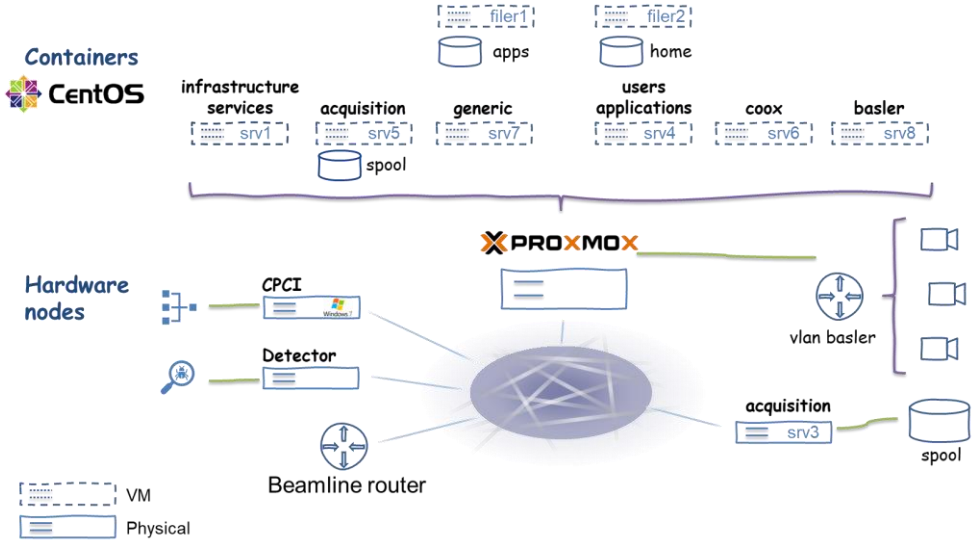
- Computing infrastructure
  - **For simulations**
    - Continue the partnership with CCRT
    - Ask for CPU time to the IDRIS and IN2P3 Supercomputing facility
  - **Renew the local HPC cluster**
    - And focus its usage through SLURM
    - Target architecture has been defined
    - Production phase targeted end of Q2 2024
  - **Use the VISA solution for :**
    - HPC on-demand
    - To manage interactive calculations (example MATLAB)
    - As a way to submit SLURM jobs to the “SOLEIL HPC cluster”
  - Work with the VISA collaboration partners to share scientific applications through Singularity containers

Virtualisation services

Virtualization

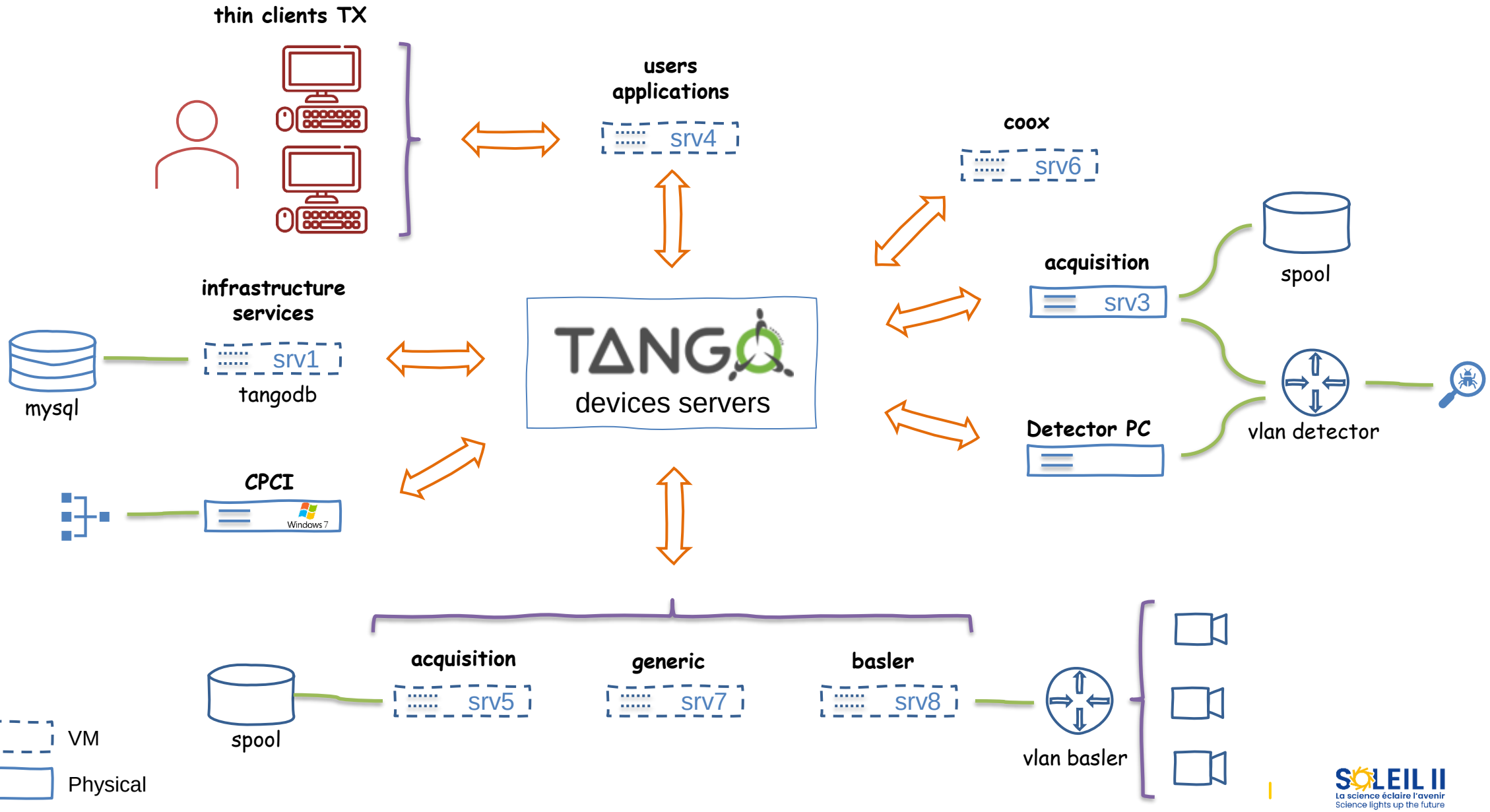
- VMWare/Simplivity for all SOLEIL
  - Based on the hyperconverged Simplivity solution
    - 8 nodes hosting hundred of VM on 2 HA clusters
    - Upgrade is on going to have 8 nodes in production Q2 2023
- Proxmox
  - Mostly used for control system network
    - Accelerators, beamlines and labs
- Openstack infrastructure in deployment
  - For VISA as a first “Use Case”
- Docker Swarm on VMWare/Simplivity
  - For web applications
- Kubernetes : no clear plans yet





Beamline control detailed architecture

Beamline control network

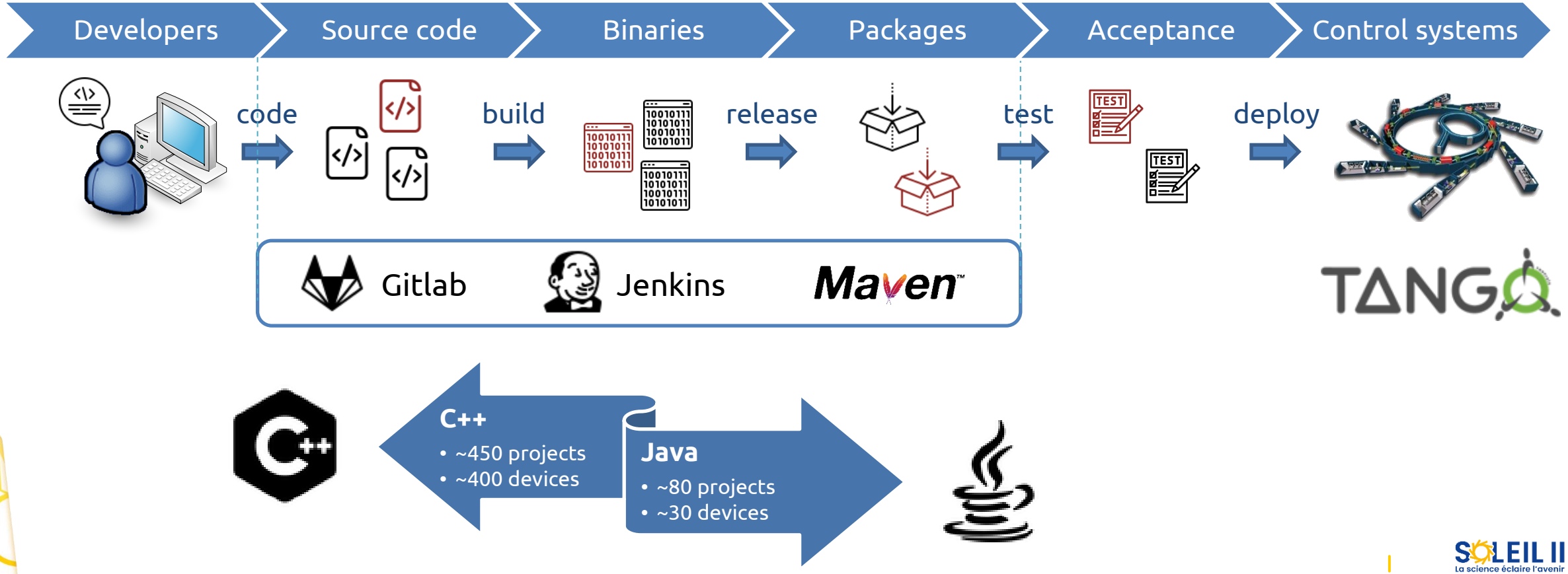


Summary

CI/CD

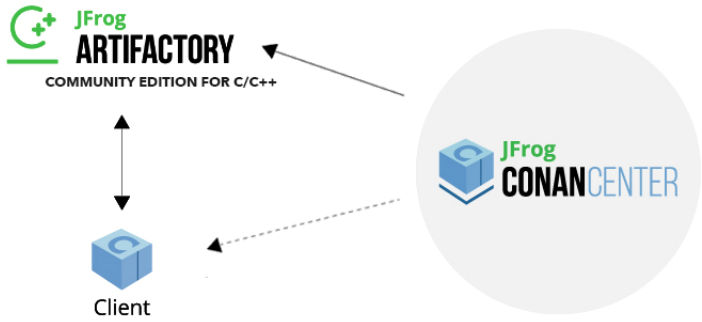
CI/CD

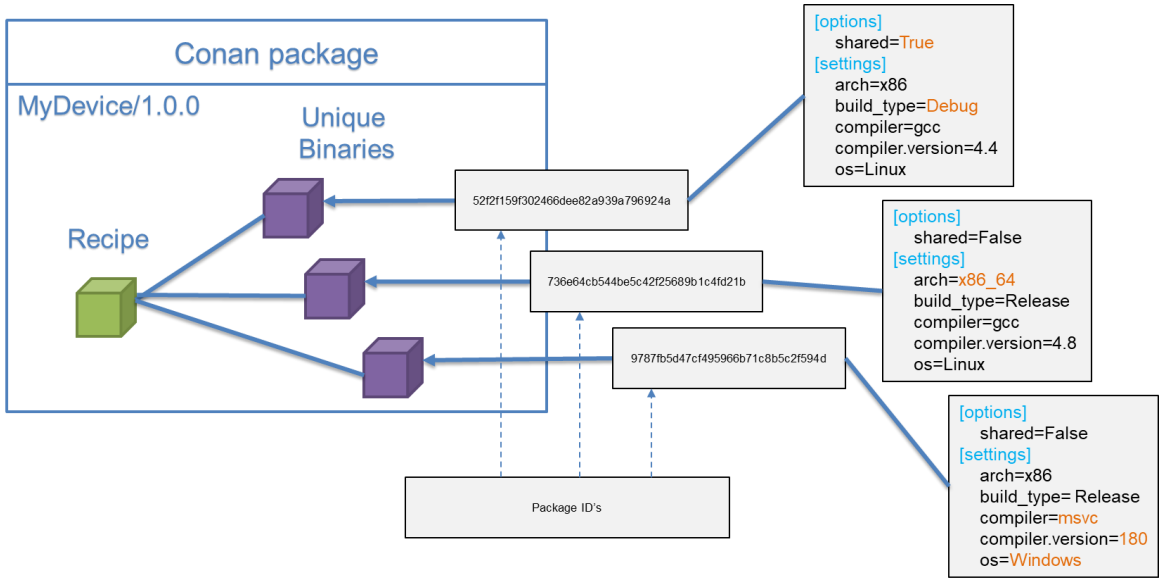
- ✓ CI/CD to build Tango device servers
- ✓ Deployment at each technical shutdown
- ✘ Components of our C++ factory are outdated
- ✘ Maven is not common for C++ development



New build approach for C++ with Conan and Cmake

- Conan = **Package and dependency manager** for C/C++
  - **Multi-platforms** and **multi-binaries** packages
  - **Abstract build system** for any other build system
  - Packages can be used from **any build system**: CMake, MSBuild, make …
  - **Public central repository** for the most popular open-source C/C++ libraries
  - Ideal solution for **C/C++ continuous integration** workflows





CI/CD plans and progress

In progress

- Finalize the replacement of **Maven** with **Conan** for C++
- Migrate to **64-bit** binaries as the standard for C++
- Support **C++11 and above**
- Migrate to the **latest LTS** versions of **OpenJDK** for Java
- Update software factory components

Upcoming

- Support of **future platforms** and **newest standards**
- Integrate **code analysis** and **testing tools**
- Extend CI/CD to other developments: **python, embedded software**
- Deploy an internal **Conda repository** to host private packages and mirror Conda-Forge, using solutions like Quetz, Artifactory, or ...
- More automation in **deployments**

Open questions

- Strategy on cloud outsourcing ?
- Strategy on virtualization and containerization ?
- Feedback about container orchestration ? About Kubernetes ?
- CI/CD practices ?
  - For C++ / Java / Python ?
  - For embedded software ?
  - For deployment ?
  - For repositories ?