The IT Infrastructure of the LUX-ZEPLIN Experiment at SURF

CHEP 2023

Norfolk, VA

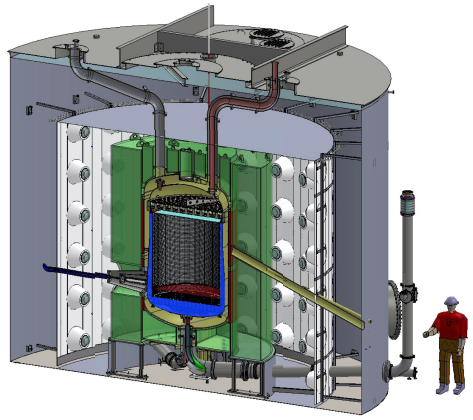
Steffen Luitz (SLAC National Accelerator Laboratory)

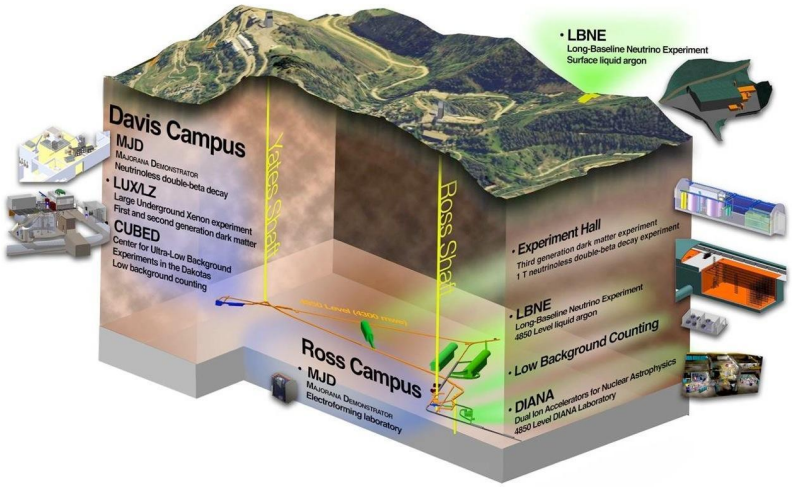
luitz@slac.stanford.edu

SLAC NATIONAL ACCELERATOR LABORATORY

The LZ Dark Matter Experiment

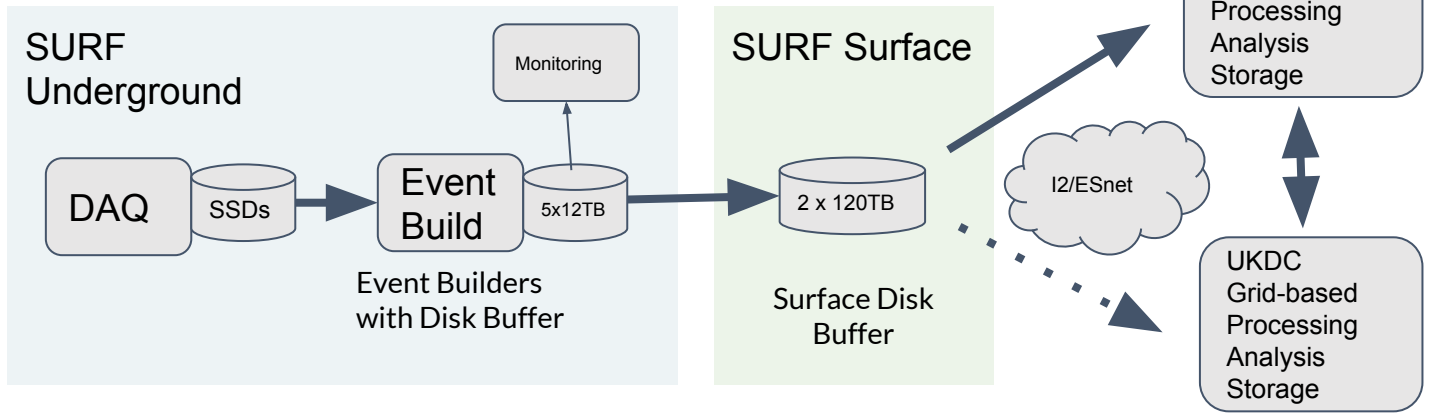
LZ detector





NIM A 163047 (2020)

- Liquid Xenon Time Projection Chamber (TPC)
  - 7 tonnes active LXe
  - Located 4850ft (~1mile) underground at Sanford Underground Research Facility (SURF) in Lead, South Dakota, USA
  - Currently operating and taking data
- Data rates
  - up to ~ 40MByte/s during routine Dark Matter search
  - up to ~ 350 MByte/s during calibration
- Dataflow deeply buffered
  - hold 2 months of Dark Matter search data at SURF
- ~200 active user accounts in system (shifters, experts, etc.)



Designing the LZ Online IT system

IT Environment at SURF

- Small lab in remote location, small IT org
- Very limited IT support for experiments
  - Firewall, VPN, basic user management
  - no services
- Site/underground network
  - hostile mine environment
  - too slow at design time, upgraded since
- Internet connectivity good
  - 10Gbit/s Internet2
  - 2 commercial ISP connections

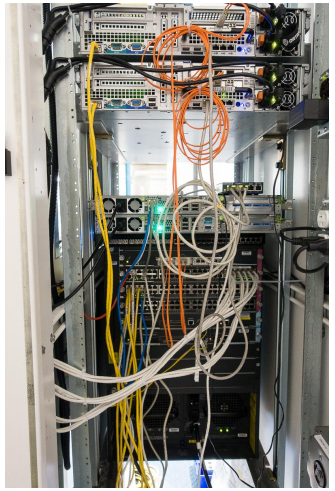
LZ requirements

- 10Gbit/s underground-surface-WAN
- Highly reliable access to underground
  - slow control & detector operations
- >> 100 physical devices on network
- ~ 500 user accounts over lifetime,
- “secure”, including remote access
- “easy to manage”, “cost-considerate”

Building our own

- Opportunity to **design** IT system
- Network
- Compute
  - Extensive use of virtualization
  - Centrally managed
    - Authentication, Directory
    - Configuration management
- Prototyping at SLAC to demonstrate performance and manageability
  - supporting LZ detector R&D and QA/QC

LZ System Test facility at SLAC



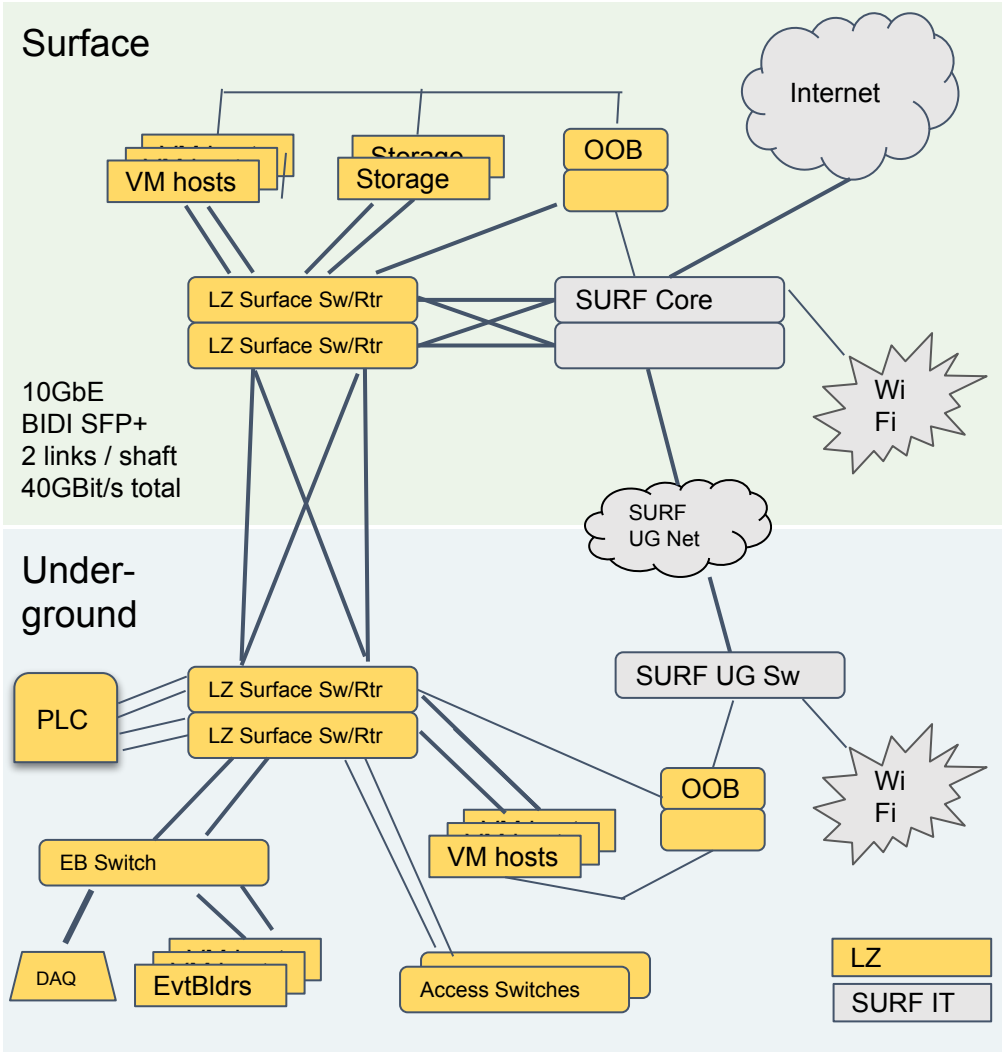
Network

Physical

- Direct connection of LZ underground and surface
- 4 strands of “dark fiber” from SURF (+2 for GPS)
- 2 Netgear m4300 stacks (underground, surface)
  - 10Gbit/s, Layer 3, low cost, expansive features, redundant
  - 4 x 10Gbit/s, point-to-point, fully redundant between stacks + to SURF
  - 2 fiber pairs with different paths (via Yates and Ross mine shafts)
- All critical devices (VM hosts, PLC, etc.): connections to their switch stack
- Separate Emergency/Out-Of-Band (OOB) network.
- 10 access switches underground (48 ports, 1Gbit/s each) for devices

Logical

- Layer-2 (VLAN) and Layer-3 IPv4 design
  - VLANs/subnets and Layer 3 ACLs create zones (security/fault isolation)
  - static routing, redundancy managed by stacking and Layer-2 (LACP)
- Connecting to SURF network via redundant vyOS Firewalls (VMs)
  - NAT to map internal services to (screened) public IPs
    - we use about 20 public IP addresses in a /27 network
    - once allowed by SURF firewalls, ports out on the open Internet
- OOB allows remote access with VM clusters (and LZ external firewalls) down
  - can use SURF network as alternative path for UG access



Underground and Surface Clusters

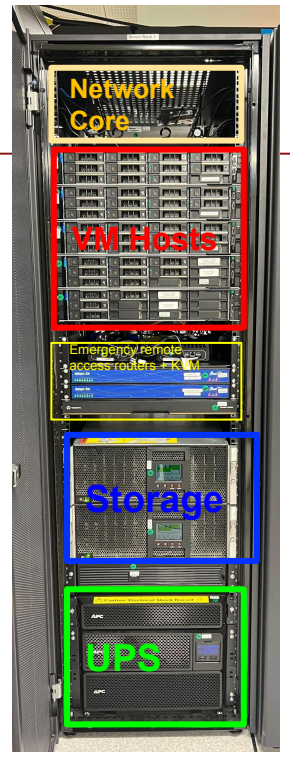
Separate; underground cluster has no operational dependencies on surface

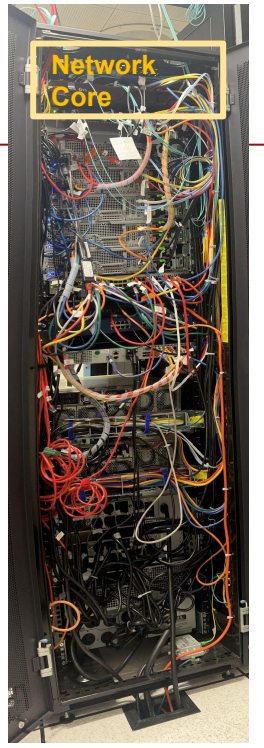
Hardware

- 2 VM host clusters (underground and surface), each:
  - 3+ x Dell R530/R540, 256 GByte RAM, 5 x 4TB SSD
  - each server 2 x 10 Gbit/s Ethernet to switch stack
- 2 Supermicro JBOD servers at surface (128 GByte RAM, 40 x 6TB HDD each)
  - ZFS RAIDz2 (some storage set aside for backups and local data)
- Fully remotely manageable
  - built-in service processors, external KVM, power control via PDUs
- All equipment dual power (1 leg on UPS, one leg on house power, generator-backed in both locations)

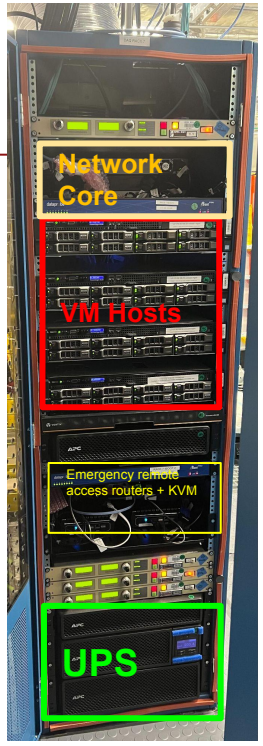
Virtualization Software (open source)

- Proxmox VE
  - We use full VMs, not containers
  - Cluster storage on Ceph
  - Clusters can each operate with 2 out of 3 nodes
    - Seamless routine maintenance using hot VM migration
- VM image backup with Proxmox Backup Server
- Running ca. 100 virtual machines total

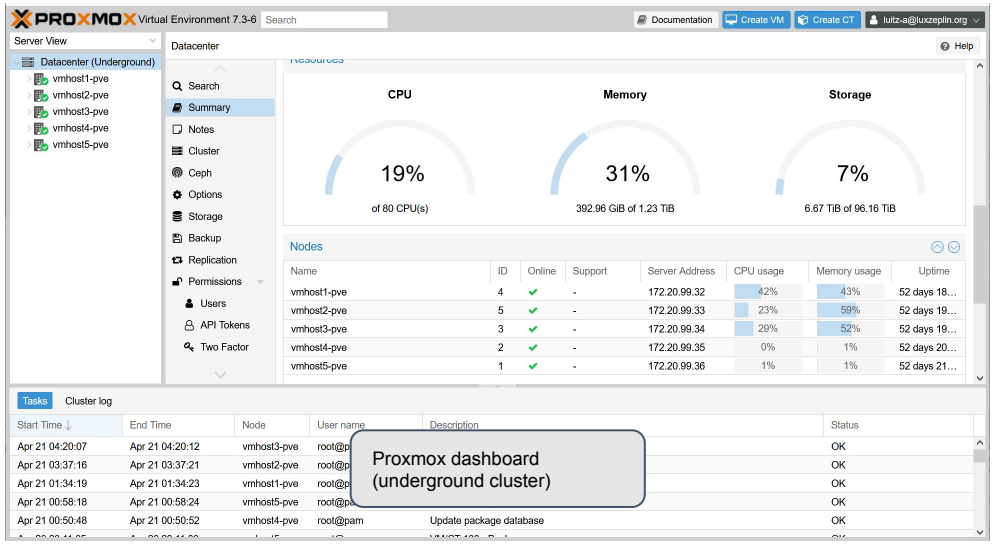




Surface Rack, front and back



Underground Infrastructure & Network Rack



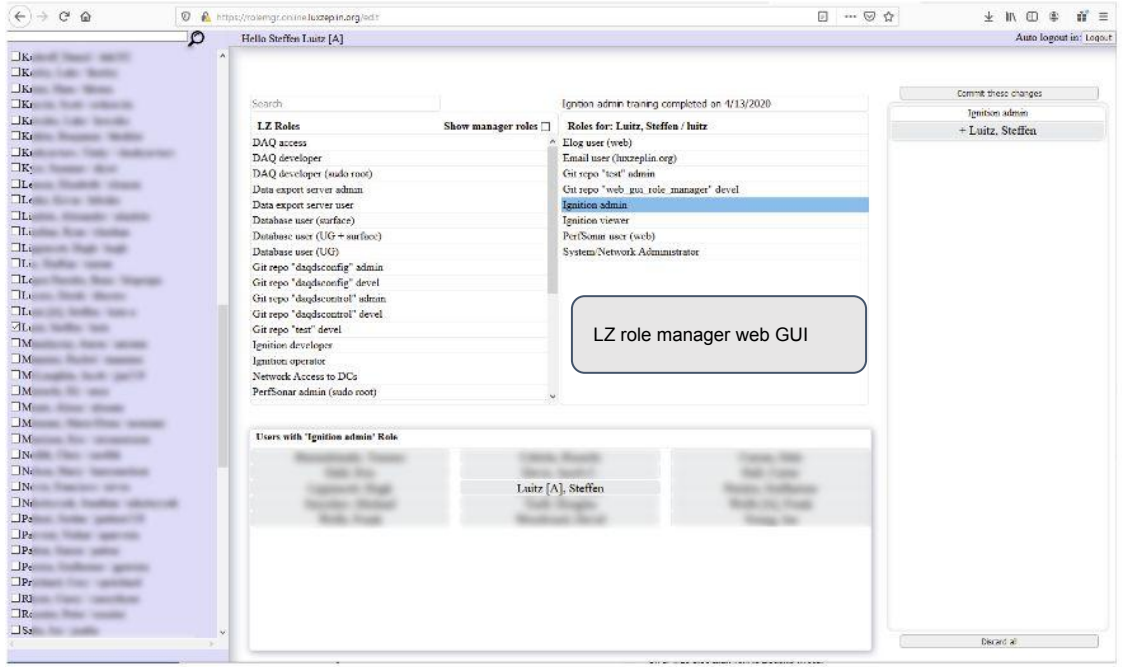
Proxmox dashboard (underground cluster)

S. Luitz

System Architecture, Management, and Services

Infrastructure and Management

- CentOS-7
- Samba 4 Active Directory with sssd on clients
  - Kerberos, LDAP and DNS with replication
  - 2 domain controllers (VM) each underground and surface
- Saltstack for configuration management
  - authoritative configuration in YAML files
  - includes network and firewalls
- Storage
  - VM images on Ceph
  - Extensive use of ZFS (zfs on linux)
    - snapshots, send/receive file systems
- https://luxzeplin.org domain for access
- Multiple MySQL 8 database server VMs
    - largest is slow control history with 2.1TB
  - replication+snapshots of replicas for downtime-less backups
- Central syslog collection (searchable)
- Backups (leveraging ZFS)
  - snapshots, local copies, off-site to University of Rochester



Role Management

- Developed role management system to manage fine-grained LDAP permissions (group memberships)
  - permissions granted/revoked based on 'role'
  - e.g. 'High Voltage Operator' role assigns all necessary access and slow control permissions to operate HV.
  - management via web interface, can be delegated
  - group and role definitions in YAML files

Remote Access, Security

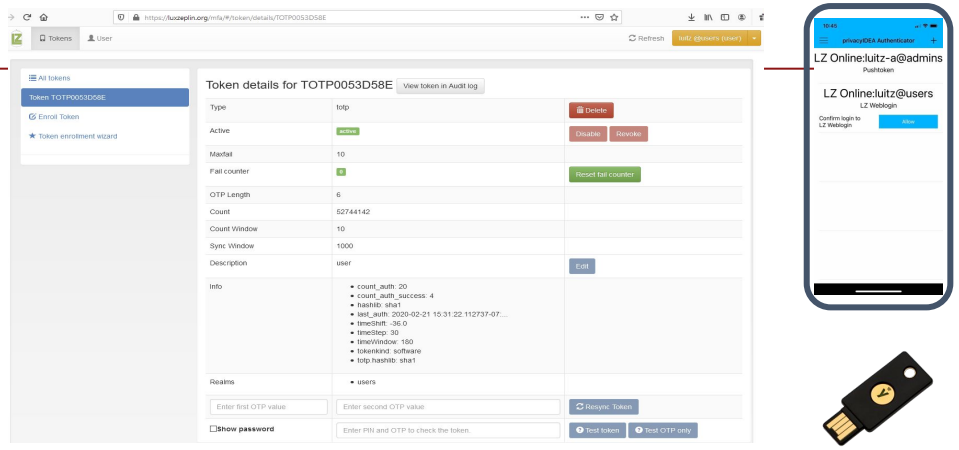
Remote Access

- 2-factor authentication for **all** remote access
  - Bring-your-own-token (TOTP, TAN, Yubikey, App)
  - PrivacyIDEA, open source, locally hosted.
- Reverse web proxy with SAML login for internal web apps
  - all web apps under single tree (luxzeplin.org)
  - permissions from LDAP through SAML
- OpenVPN
  - automated certificate management
  - fine-grained IP target address restrictions (from LDAP)
  - Education institution license of Viscosity client
- SSH
  - fine-grained IP target address restrictions (LDAP)

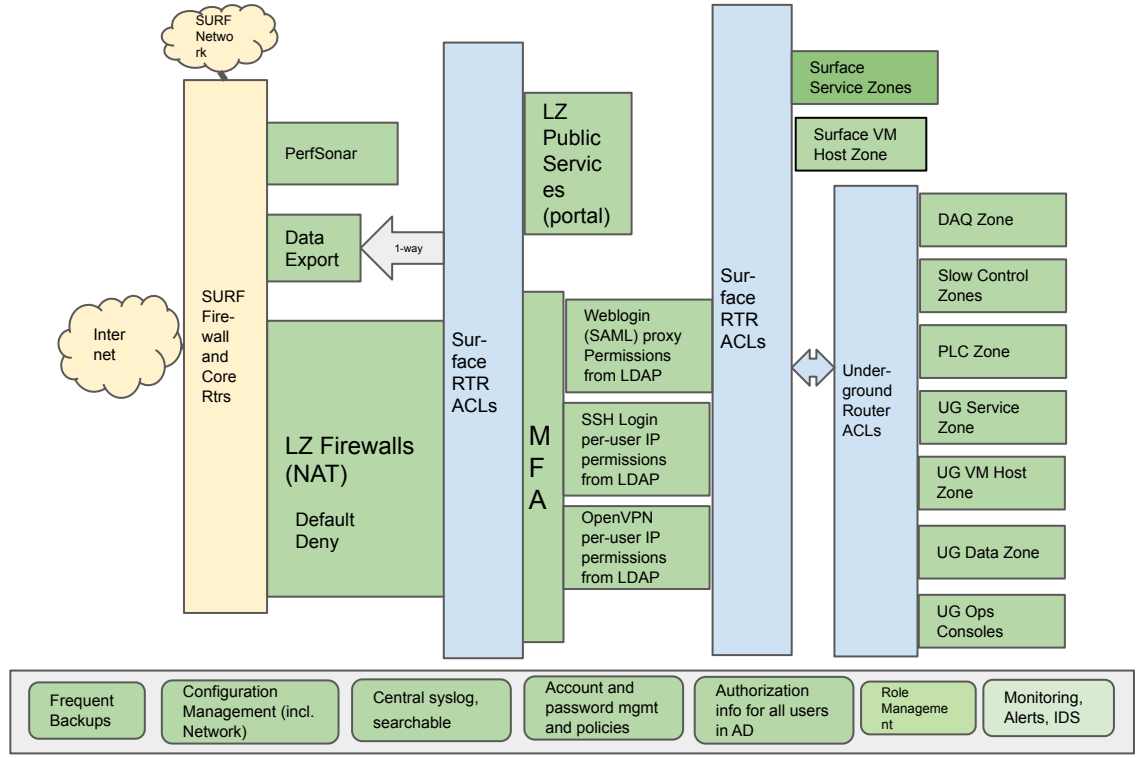
Security

- Defense in depth designed into the system
- Free/open source tools

PrivacyIDEA portal, mobile app, Yubikey



Security Architecture / Features



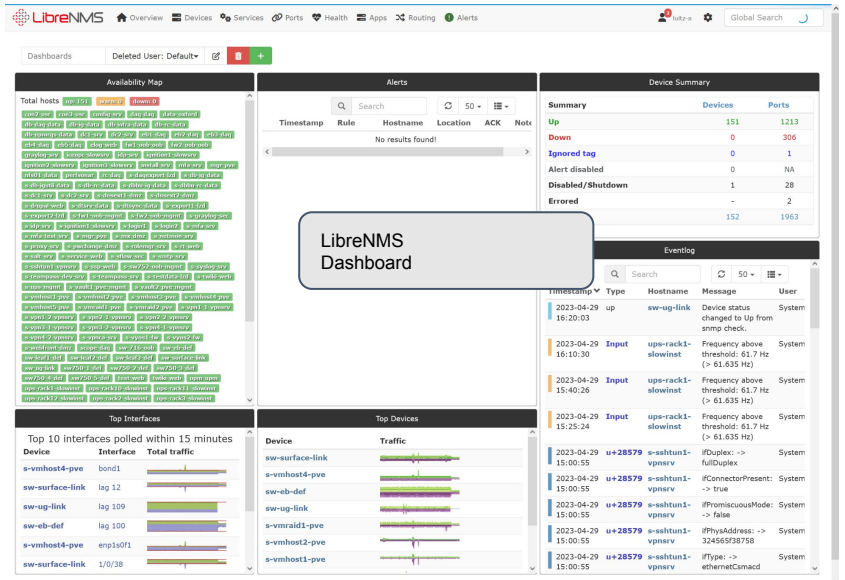
Operational Experience

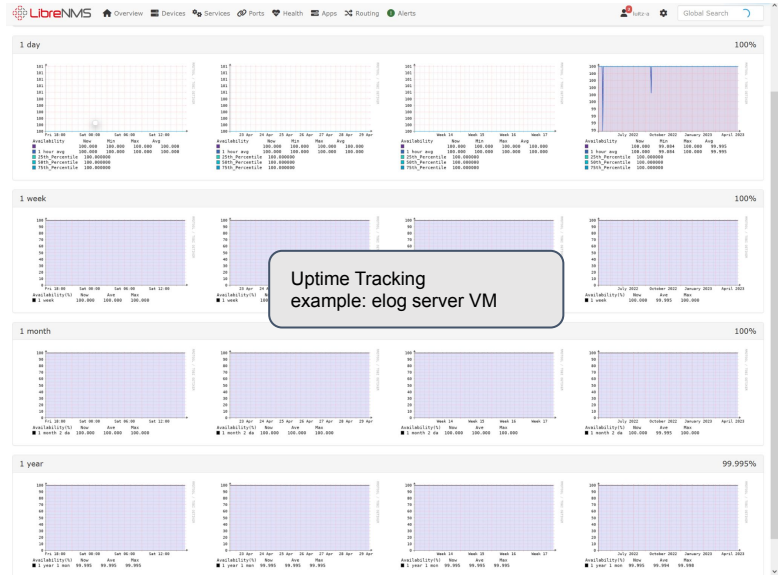
Staged installation, Rochester to SURF

- Setting up production system at University of Rochester
  - first application was electronic logbook to support LZ construction activities
  - most system setup and configuration via remote management from California,
  - resulted in intrinsically remotely operable system
- shipping pre-configured network/clusters to SURF, installation at SURF in 2018/2019
- operating very reliably since
- Full/reliable remote access and management essential during Covid-19

Availability

- Network/host monitoring system tracks downtime
- May 2022-now (1 year):
  - VM typical availability 99.995% (‘4 ½ nines’, ~30min downtime/year)
  - Physical host availability 99.95%-99.995% (slower boot times)
  - dominated by 2 scheduled LZ network outages (switch firmware upgrades)





Summary

- **LZ network and computing at SURF is working reliably**
- **Experiment is taking data**
- **Personal perspective**
  - In my 30 year career building and improving DAQ and Online systems, IT was always something already existing.
    - Organically grown, difficult to improve
    - In particular system management and Information Security aspects, retrofits usually collide with user expectations and convenience
  - I was always wondering if designing something from scratch is better
    - Both users and administrators
  - The answer is Yes!

LUX-ZEPLIN (LZ) collaboration

- Black Hills State University
- Brookhaven National Laboratory
- Brown University
- Center for Underground Physics
- Edinburgh University
- Fermi National Accelerator Lab.
- Imperial College London
- King's College London
- Lawrence Berkeley National Lab.
- Lawrence Livermore National Lab.
- LIP Coimbra
- Northwestern University
- Pennsylvania State University
- Royal Holloway University of London
- SLAC National Accelerator Lab.
- South Dakota School of Mines & Tech
- South Dakota Science & Technology Authority
- STFC Rutherford Appleton Lab.
- Texas A&M University
- University of Albany, SUNY
- University of Alabama
- University of Bristol
- University College London
- University of California Berkeley
- University of California Davis
- University of California Los Angeles
- University of California Santa Barbara
- University of Liverpool
- University of Maryland
- University of Massachusetts, Amherst
- University of Michigan
- University of Oxford
- University of Rochester
- University of Sheffield
- University of Sydney
- University of Wisconsin, Madison

US UK Portugal Korea Australia

SLAC S. Luitz

Science and Technology Facilities Council

Sanford Underground Research Facility
South Dakota Science and Technology Authority

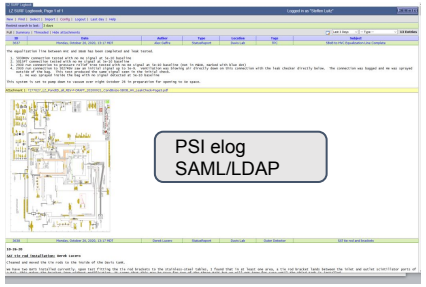
FCT Fundação para a Ciência e a Tecnologia
MINISTÉRIO DA EDUCAÇÃO E CIÊNCIA

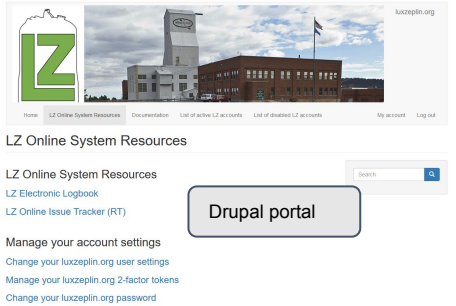
ibs Institute for Basic Science

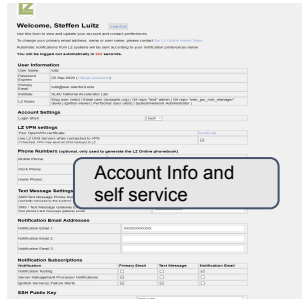
Backup Slides

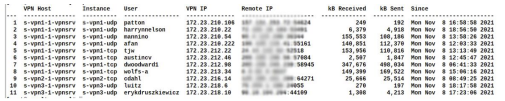
List of Tools

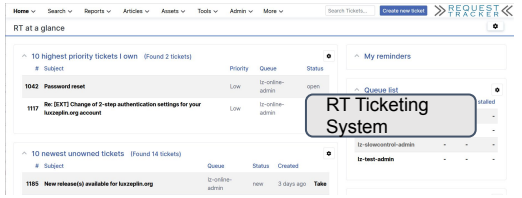
- Open Source
  - CentOS-7, ZFS on Linux (OS)
  - Saltstack (configuration management)
  - Proxmox Virtual Environment (Virtual Data Center)
  - Proxmox Backup Server (VM image backups)
  - Sanoid/Syncoid for ZFS (snapshot+remote backups)
  - Samba 4 AD (Directory, Kerberos, LDAP, DNS)
  - MySQL (database)
  - PrivacyIDEA (2-factor authentication)
  - vyOS (router/firewall linux distribution)
  - OpenVPN (VPN)
  - simplesamlphp (SAML IdP and SP)
  - PSI elog (electronic logbook)
  - Drupal (web portal)
  - LibreNMS and elastiflow (network monitoring)
  - SSP (password portal)
  - BestPractical RT (ticket tracker)
  - Graylog (syslog indexing and access)
  - Security Onion (intrusion detection)
  - PerfSONAR (Internet performance measuring/monitoring)
- Commercial (unlimited users licenses)
  - Ignition for Slow Control
  - Viscosity VPN client (educational site license)

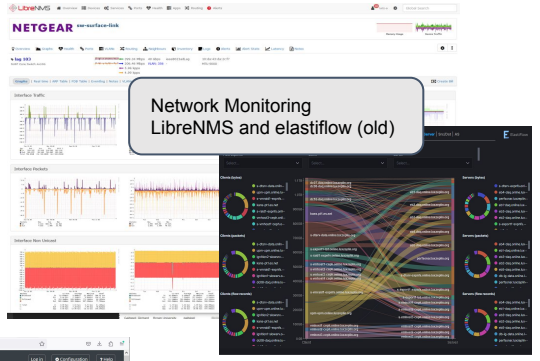


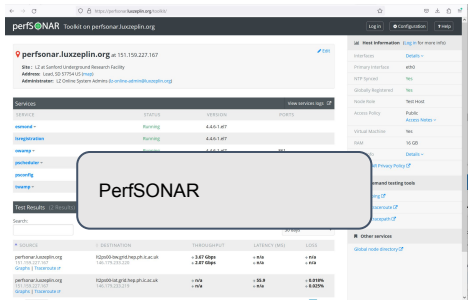












LZ Network Layer 3 sketch