

Scalable Systems Laboratory

Rob Gardner
Enrico Fermi Institute
University of Chicago



CHEP 2019
November 5, 2019





MONARC

Models Of Networked Analysis

At Regional Centers

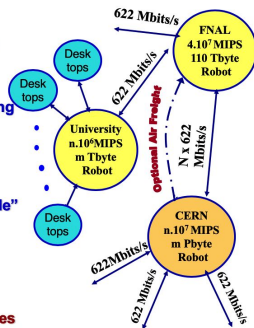
Caltech, CERN, FNAL, Heidelberg, INFN, Helsinki, KEK, Lyon, Marseilles, Munich, Orsay, Oxford, RAL, Tufts, ...

GOALS

- ➔ Specify the main parameters characterizing the Model's performance: throughputs, latencies
- ➔ Determine classes of Computing Models feasible for LHC (matched to network capacity and data handling resources)
- ➔ Develop "Baseline Models" in the "feasible" category
- ➔ Verify resource requirement baselines: (computing, data handling, networks)

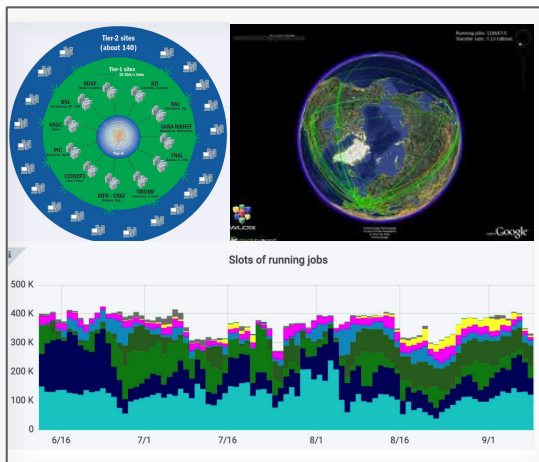
COROLLARIES:

- ➔ Define the Analysis Process
- ➔ Define Regional Center Architectures
- ➔ Provide Guidelines for the final Models



Modelling Regional Centers with MONARC simulation tools

CHEP2000 07 Feb 2000



1999

plans...

2019

production
workhorse

2022+

flexible facilities
for innovation

Evolving our Facilities

*"Evolutionary change towards HL-LHC is required, as the experiments will continue to use the current system. Mapping out a **path for migration** then requires a fuller understanding of the costs and benefits of the proposed changes. **A model is needed** in which the benefits of such changes can be **evaluated**, taking into account **hardware and human costs**, as well as the impact on software and workload performance that in turn leads to **physics impact.**"*

<https://arxiv.org/abs/1712.06982>



→ We need *Facility R&D*
to match our
Software R&D

NEXT
GEN
TIER2

A fundamental **redesign** of our Tier2 complex is needed to provide **flexibility** to innovate services (CI/CD pipelines), **reproducibility** to ensure reliability, and **security** to sustain new operations models

E.g. DOMA (analysis) facility elements

- service platform
- clustered storage
- analysis processing
- fast WAN for delivery
- fast LAN for access
- edge caches
- interactive services
- accelerators for ML and DOMA functions

Resources
for scalable
innovation
needed!

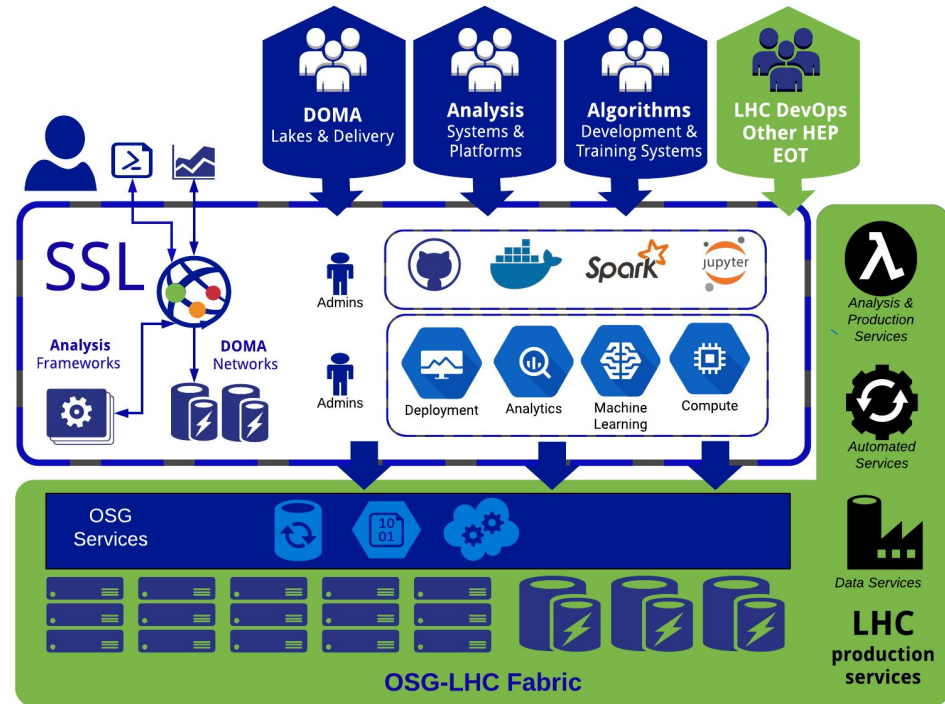


SSL as a Path to Production

Provisioning of software **environments** and development tools.

Distributed platforms materialized with tools like containerized edge services.

Integration point with the OSG, WLCG and LHC experiment services (data, analysis).



Orchestrating services in the SSL

- Need **flexible infrastructure** for supporting the workloads we expect
- **Dynamically reconfigure services** on existing hardware
 - e.g. HTC batch vs interactive frameworks
- Can we **leverage** cloud-native momentum?
- Want to **"glue" clusters together to achieve scale**, but abstract away infrastructure differences
 - Clear a smooth road for the developers
- Discussion to build a **multi-site K8s substrate**

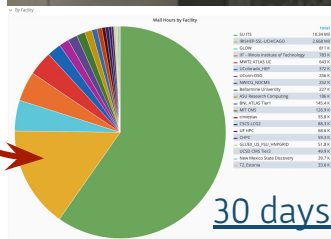


kubernetes

First piece of the substrate

- Repurposed R&D cluster
 - Vintage but nice (3k cores)
 - 48 cores, 256 GB RAM, 2x800GB SSD, 10Gbps networking per node
 - 2x40g to campus 100g SciDMZ
 - Kubernetes for flexibility for services and compute
- Deployed REANA application
- Deploy HTCondor workers and backfill with OSG Connect

Federated ID access (institutional, CERN account), edge services hosting, Unix account provisioning, OSG-LHC software environment



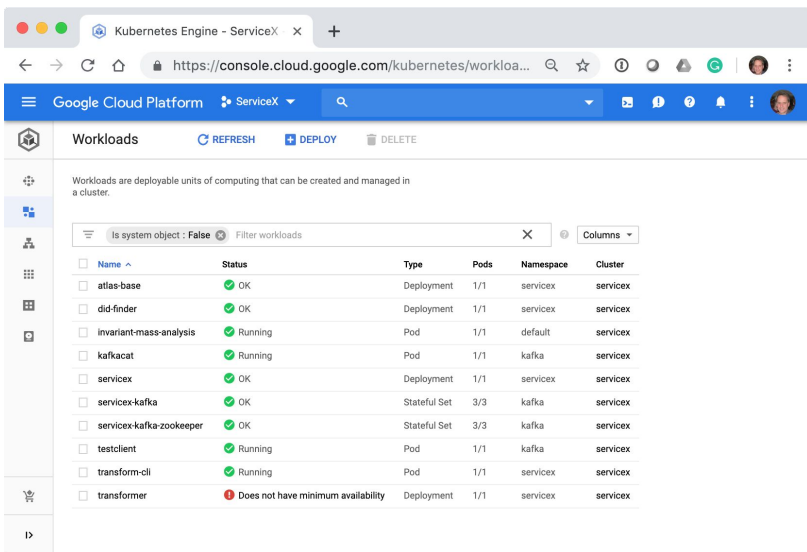
reana

[30 days OSG opportunistic \(link\)](#)

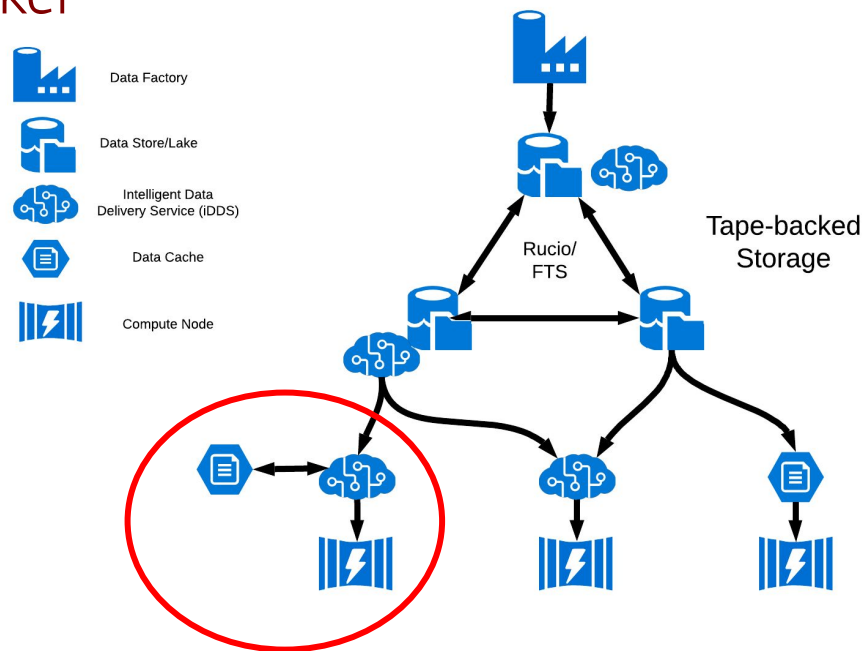
Using SSL for DOMA R&D

DOMA iDDS development on Google Cloud
reproduced on IRIS-HEP SSL with Docker
containers and Kubernetes

c.f. B. Galewsky's talk in
DOMA track 11:00 Monday



| Name | Status | Type | Pods | Namespace | Cluster |
|--------------------------|------------------------------------|--------------|------|-----------|----------|
| atlas-base | OK | Deployment | 1/1 | servicex | servicex |
| did-finder | OK | Deployment | 1/1 | servicex | servicex |
| invariant-mass-analysis | Running | Pod | 1/1 | default | servicex |
| kafkacat | Running | Pod | 1/1 | kafka | servicex |
| servicex | OK | Deployment | 1/1 | servicex | servicex |
| servicex-kafka | OK | Stateful Set | 3/3 | kafka | servicex |
| servicex-kafka-zookeeper | OK | Stateful Set | 3/3 | kafka | servicex |
| testclient | Running | Pod | 1/1 | kafka | servicex |
| transform-cli | Running | Pod | 1/1 | servicex | servicex |
| transformer | Does not have minimum availability | Deployment | 1/1 | servicex | servicex |



Now the wet blanket (or opportunities?)



- Multi-tenancy is hard
 - Can the HEP community contribute?
- Federation is hard – focus in K8s community has been multi-cluster, not multi-owner
- Container security becomes more important
 - Community image scanning service?
- Co-scheduling services & job workloads
 - Related: managing user priority & preemption



Lets crawl forward!

- Resource Providers (WLCG sites– we’re looking at you!)
 - ◆ Contributing to multi-prem, scalable platforms – how?
 - ◆ Especially interested in federation experiments, evaluations
- Users & Developers
 - ◆ SSL would like to “align” with you -- service delivery methods & services
 - ◆ Training events similar to CoDaS-HEP coordinated by IRIS-HEP [TEO team](#)
- Organizing communities
 - ◆ [SSL-TEAM forum](#) – monthly calls focused on building out federated kubernetes platform
 - ◆ What can we do more broadly with other enthusiasts in WLCG?

Conclusions

- While the grid model and software will evolve, certain principles remain in our domain:
 - distributed trust, resource aggregation & sharing
- New principles emerging:
 - declarative, reproducible infrastructure (as code)
 - federation operations (towards "NoOps")
 - connecting services in more flexible ways – e.g. service meshes
- We need to evolve facility infrastructure to support software innovation
 - k8s "substrate" seen as promising path to build multi-site scalable systems

K8s Substrate BoF
Thursday, 3:30pm, RB

Kubernetes pre-GDB, Dec 10

Thank You