CfA Nexus Science Platform

Omar Laurino - Center for Astrophysics | Harvard & Smithsonian

May 9, 2023 - IVOA Bologna

CENTER FOR **ASTROPHYSICS**
HARVARD & SMITHSONIAN

The Landscape

The CfA operates a range of telescopes at different wavelengths (among others: Chandra, Magellan, MMT, SMA, DASH, EHT, Whipple, VERITAS)

Key atomic and molecular databases (HITRAN, ATOMDB)

Theoretical simulation data

NASA Astrophysics Data System (ADS)

CENTER FOR **ASTROPHYSICS**

HARVARD & SMITHSONIAN

The CfA Nexus

" A science platform that will allow scientists to access, visualize and analyze multi-wavelength CfA data, link them to cloud computing, and to the literature via ADS. "

" make new discoveries possible by facilitating archival research "

CENTER FOR **ASTROPHYSICS**

HARVARD & SMITHSONIAN

The Architectural Philosophy

"Create the infrastructure connective tissue between data, tools, and services: the Nexus."

"Long-lasting and secure, extensible, and allow for modularity and different user views."

"The CfA-Nexus is not a monolith, rather the connective tissue around the CfA Data Centers, that will continue to curate their data."

CENTER FOR **ASTROPHYSICS**
HARVARD & SMITHSONIAN

Infrastructure Research & Development

"Collect scientific requirements, evaluate the required infrastructure, and install core functionalities that will enable an early demonstration of the platform with selected CfA datasets."

CENTER FOR **ASTROPHYSICS**
HARVARD & SMITHSONIAN

♻️ Reduce, Reuse, Recycle ♻️

Rubin Science Platform (RSP)

Built on top of Kubernetes, Helm, and Argo CD. VO is front and center.

Composed of several layers:

Application Infrastructure: Foundational services for RSP

Rubin Science Platform Core: Primary applications and services

RSP+: Additional applications and services extending RSP functionality

Can we reuse RSP components and adapt them for the Nexus requirements?

CENTER FOR **ASTROPHYSICS**

HARVARD & SMITHSONIAN

Lessons Learned

Phalanx repository as single source of GitOps truth.

Flat umbrella application structure (reflected in scripts).

Chicken and egg problems (e.g. commit LoadBalancer IP to version control).

Prototyping and developing not straightforward. (Especially Jupyter Lab.)

CENTER FOR **ASTROPHYSICS**
HARVARD & SMITHSONIAN

Infrastructure as Code

I want to:

Separate design from implementation details

Avoid technology lock-in as much as possible

Apply software best practices

dependency injection

encapsulation

unit and regression testing

CENTER FOR **ASTROPHYSICS**
HARVARD & SMITHSONIAN

Pulumi

"Universal" IaC tool that supports multiple cloud providers and programming languages

Provides high-level abstractions for creating, deploying, and managing cloud resources

Reusable Component Resources promote modularization

CENTER FOR **ASTROPHYSICS**

HARVARD & SMITHSONIAN

```typescript
const cluster: Cluster = ClusterFactory.createCluster("gcp", options)

const appPlatform = new ApplicationPlatform("app-platform",
    cluster,
    {dependsOn: cluster, provider: cluster.provider});

const postgres = applicationFactory.createComponent(
    "postgres", {},
    {dependsOn: [appPlatform.argoCd], provider: cluster.provider})

const jupyter = new Jupyter("jupyter", {
    cluster, ingress: appPlatform.ingressNginx
}, {dependsOn: gafaelfawr, provider: cluster.provider})

const home = new StaticWebsite("home", {
    provider: cluster.provider,
    dependsOn: [cluster, appPlatform.ingressNginx, gafaelfawr, jupyter]
})
```

CENTER FOR **ASTROPHYSICS**
HARVARD & SMITHSONIAN

Extension Points

IVOA Archive In a Box?

```
const cluster = ClusterFactory.createCluster("gcp", options);

const db = new Postgres();
const tapServer = new CadcTapServer({cluster, db});

const chandraSchema = new ChandraSourceCatalogSchema();
const oirSchema = new OirSchema();

tapServer.add({schema: chandraSchema, endpoint: "/tap-csc-2.1"});
tapServer.add({schema: oirSchema, endpoint: "/tap-oir-1.1"});
```

CENTER FOR ASTROPHYSICS

HARVARD & SMITHSONIAN

Extension Points

Authentication and Authorization ✅

Home Page ⛔

Portal ❓

Jupyter Lab (registry of trusted Docker images, conda environments)

Integrate Science Platforms: Seamless access to RSP, Nexus, etc. in the CANFAR virtual desktops?

User Data Storage

CENTER FOR **ASTROPHYSICS**
HARVARD & SMITHSONIAN