Arcade Prototype and Platform Standardization

Brian Major
October 2019
IVOA Groningen, NL

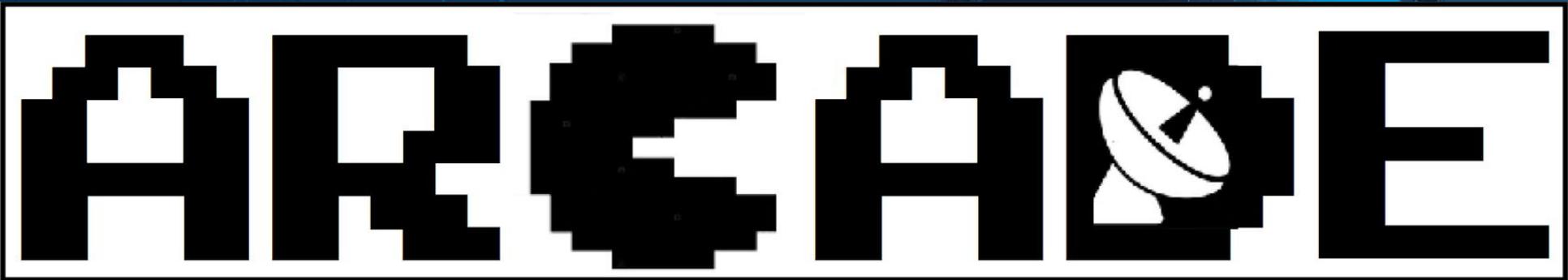National Research Council Canada   Conseil national de recherches Canada

Canada

Arcade Overview

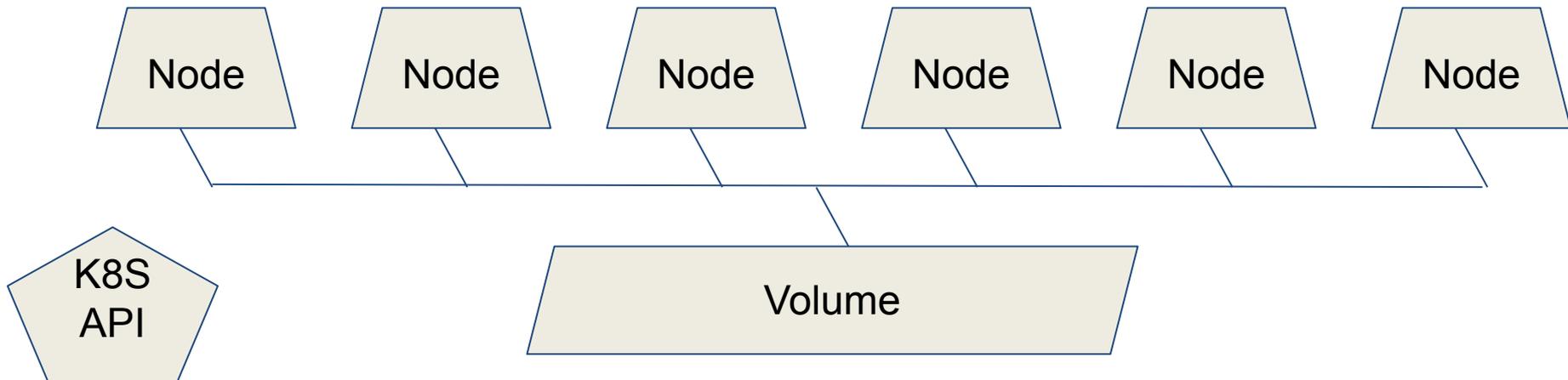Architecture and move to Kubernetes

Future for Arcade:
- Science and Operational Requirements
- Interoperability with other platforms
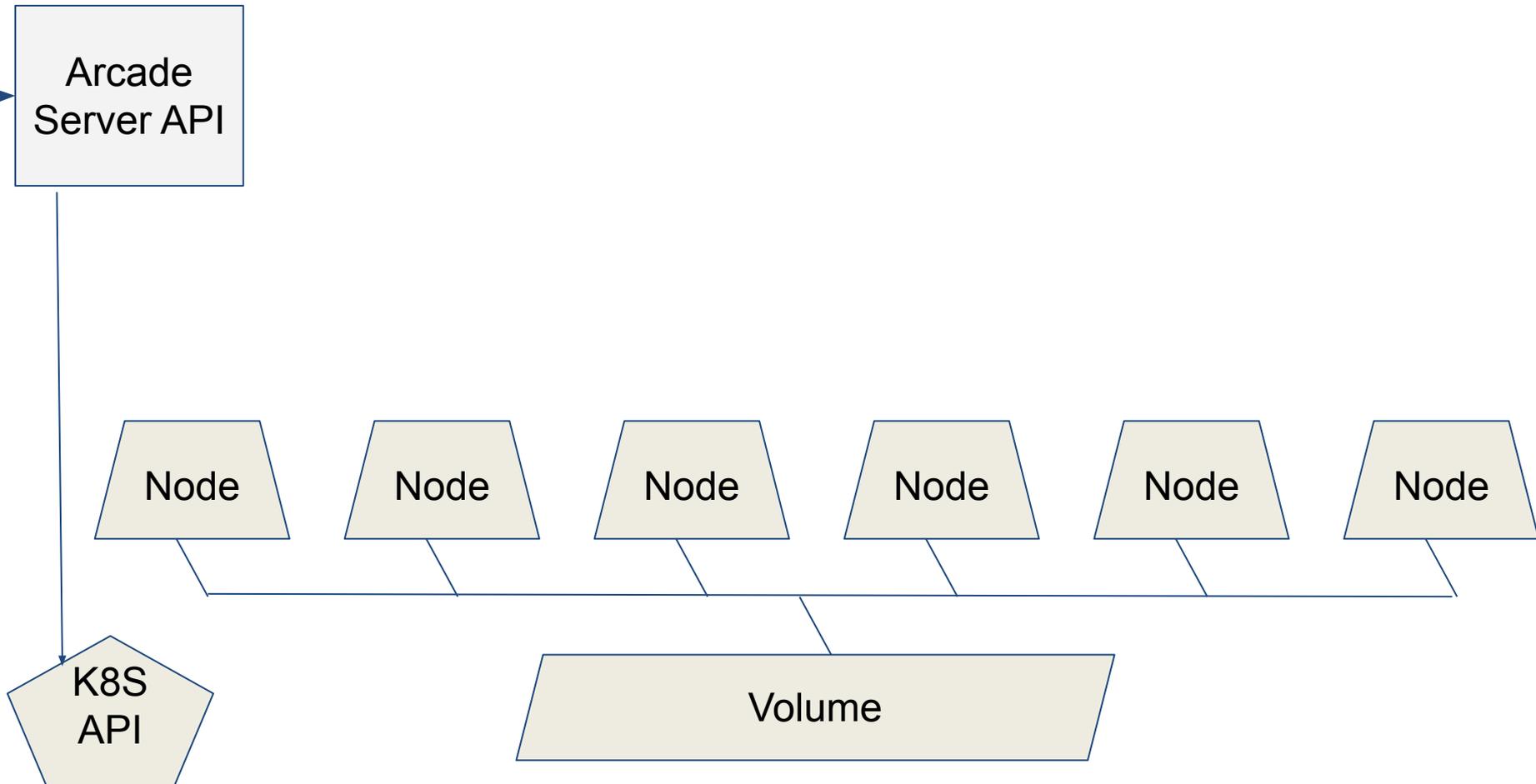- Standardization

NRC·CNRC

(logo design by Nienke van der Marel)

- Prototype data processing environment

- Astronomy software packages
  - command line and graphical

- Users interact with a remote desktop

- REST API to start sessions & launch software

NRC·CNRC

Kubernetes (K8S) Container architecture

Node

Node

Node

Node

Node

Node

K8S
API

Volume

NRC·CNRC

Kubernetes (K8S) Container architecture

Arcade Server API

Node Node Node Node Node Node

K8S API

Volume

NRC·CNRC

Canadian Advanced Network for Astronomical Research

Take the tour

Storage Management

Group Management

Data Publication
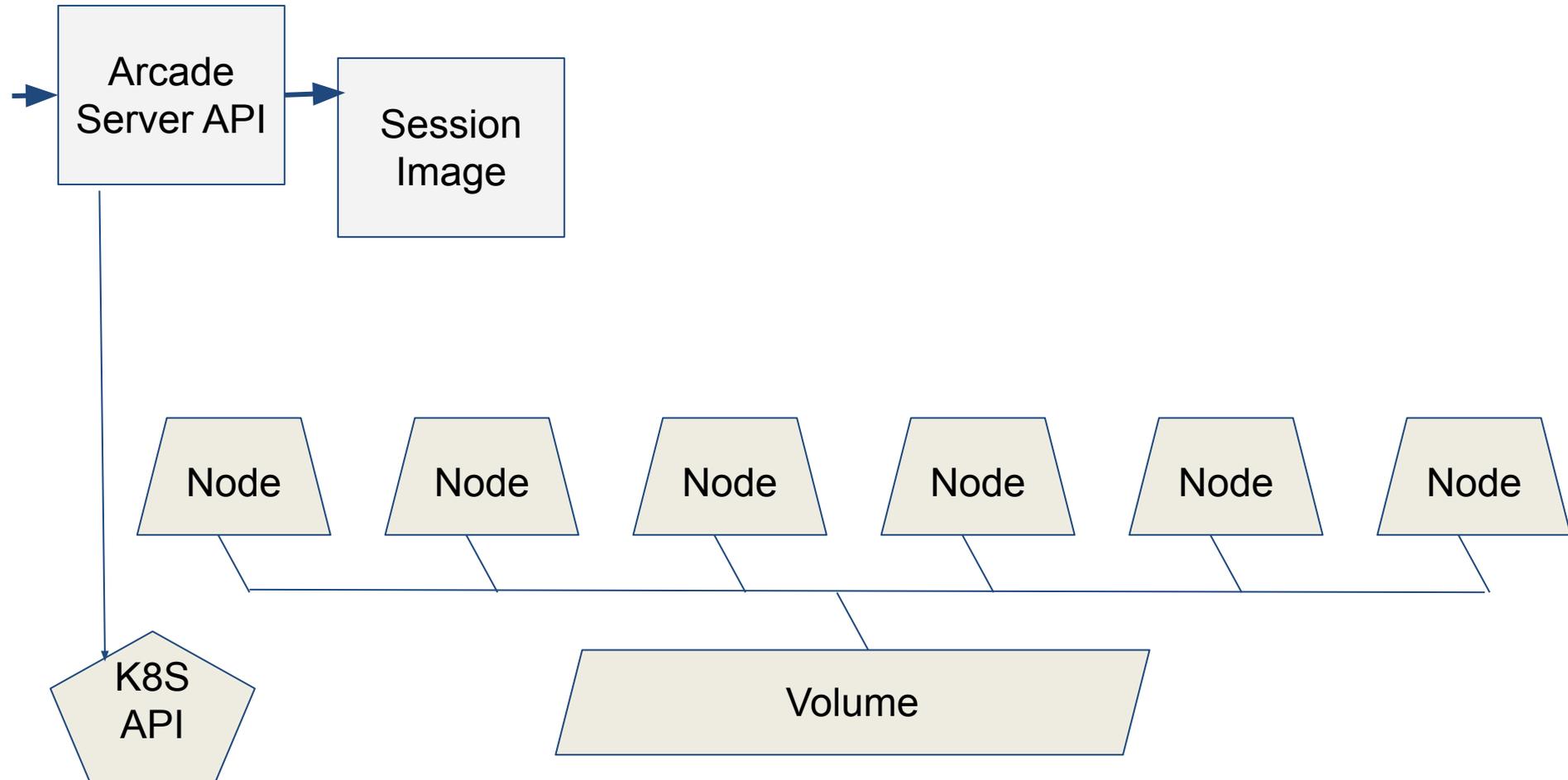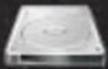
Batch Processing

Arcade

Open Stack

CADC Search

Kubernetes (K8S) Container architecture

File System

firefox

Home

arcade-
feedback

terminate

xterm

Run Program...

Terminal Emulator
File Manager
Mail Reader
Web Browser

Settings ▶
Accessories ▶
CANFAR ▶
Internet ▶
Multimedia ▶
Office ▶
Resolution ▶
System ▶

Log Out

arcade-feedback
astrodebian
casa-4.3.0
casa-4.3.1
casa-4.3.1-pipe
casa-4.4.0
casa-4.5.0
casa-4-5.1
casa-4.5.2
casa-4.5.3
casa-4.6.0
casa-4.7.0
casa-4.7.1
casa-4.7.2
casa-5.0.0-218
casa-5.1.0-74
casa-5.1.1-5
casa-5.1.2-4
casa-5.3.0-143
casa-5.4.0-70
firefox
omm2caom2
OSSOS
terminate
xterm

terminate

xterm

xterm

9fa9cec8:/$

Kubernetes (K8S) Container architecture

File System

firefox

Home

arcade-feedback

terminate
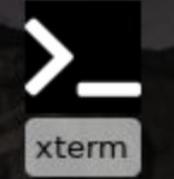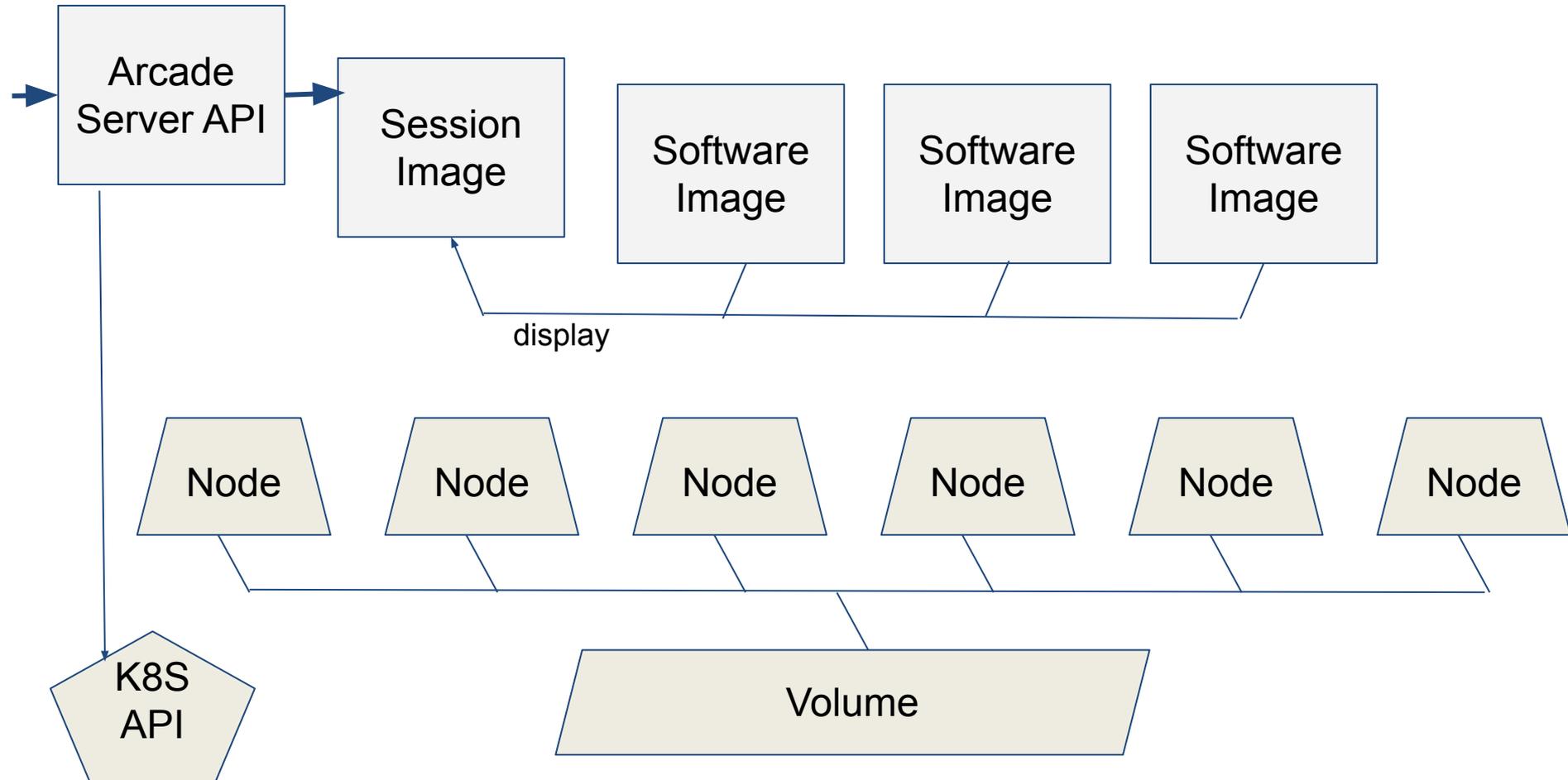
xterm

Log Messages (524818dcc50f:/home/guest/casapy-20191009-215

File   Edit   View

Search Message   Filter Time

Time	Priority	Origin	Message
	INFO	casa::::	---
	INFO	casa::::	CASA Version 4.5.2-REL (r36115)
	INFO	casa::::	Tagged on: 2016-02-09 17:39:09 -0500 (Tue, 09 Feb 2016)

xterm

```
are cached or not.
========================================


CASA Version 4.5.2-REL (r36115)
  Compiled on: Wed 2016/02/10 13:31:05 UTC

----------------------------------------------------------
    For help use the following commands:
    tasklist                - Task list organized by category
    taskhelp                - One line summary of available tasks
    help taskname           - Full help for task
    toolhelp                - One line summary of available tools
    help par.parametername  - Full help for parameter name
----------------------------------------------------------
Activating auto-logging. Current session state plus future input saved.
Filename       : ipython-20191009-215800.log
Mode           : backup
Output logging : False
Raw input log  : False
Timestamping   : False
State          : active
*** Loading ATNF ASAP Package...
*** ... ASAP (4.3.0a rev#34723) import complete ***

CASA <2>: █
```
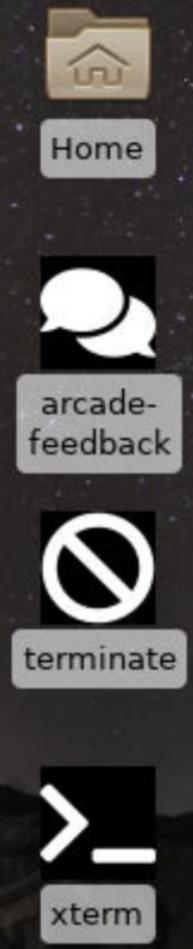
Kubernetes Container architecture

Arcade Server API

Session Image

Software Image

Software Image

Software Image

Node — Node — Node — Node — Node — Node

Resource Profile A

Resource Profile B

Resource Profile B

Resource Profile C

Resource Profile C

Resource Profile D

Basic

ALMA dataset

ALMA dataset

High RAM

High RAM

ML

NRC·CNRC
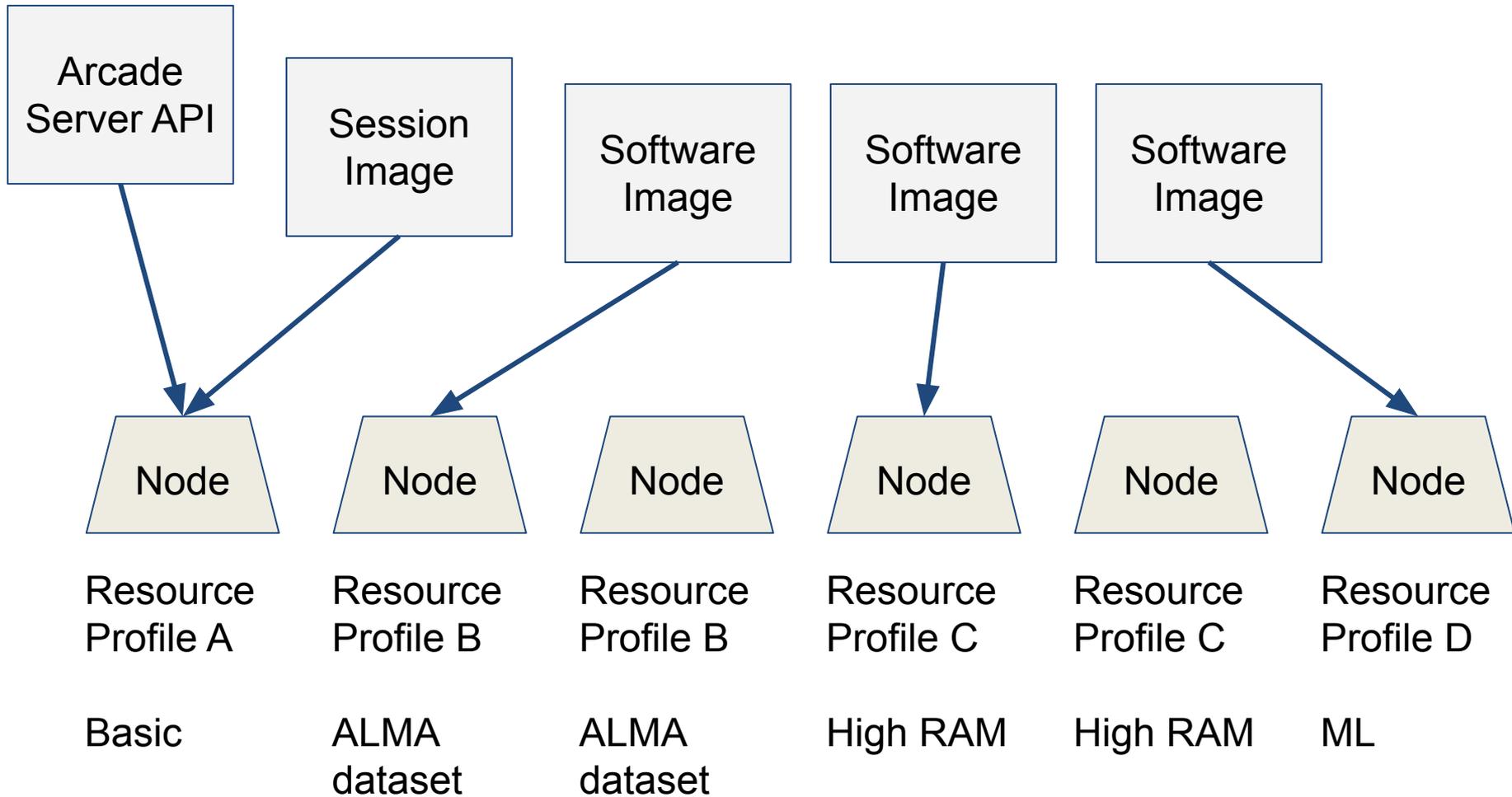
The Future for Arcade

Prototype Successful!

- Risky and unknown technical issues proved possible
- Usability and operation workings need may improvements
- Time to plan and redesign, possibly rebuild

Goals going forward:

- Scalability (leverage orchestration)
- Performance (I/O access)
- Usability (much work to do here!)

- Standardization and interoperability
- Open Source collaboration - let's share the work

NRC·CNRC

Science case: JWST Data Processing

Canadian JWST Project Scientist: Chris Willott

"Arcade will be the JWST data processing environment containing specialized software for users to process and analyze JWST datasets covering a range of science topics."

NRC·CNRC

ALMA Development Study - Starts immediately

PI: Helen Kirk

ALMA Data reduction in the CANFAR data environment

(i.e. – Improve the scalability, performance, and usability of Arcade for the ALMA data reduction workflow)

Interesting Work items:

- Development environment in a kubernetes **commercial cloud**
- Storage: Explore **storage solutions** to achieve a secure and performant user personal storage space with data sharing.
- Container/**Software maintenance**
- **Open Source** code (github)

NRC·CNRC

Challenges: Storage

- Requirements:
    - /home/guest must be the same on every container
    - Future: certain data sets available to certain containers

- System requirements
    - Fast I/O
    - Persistent
    - Shareable

- Cavern (mountable VOSpace)
- Kubernetes volumes
- Ceph FS on worker nodes

NRC·CNRC

Challenges: Storage

```
┌──────┐  ┌──────┐  ┌──────┐  ┌──────┐  ┌──────┐  ┌──────┐
│ Node │  │ Node │  │ Node │  │ Node │  │ Node │  │ Node │
└──────┘  └──────┘  └──────┘  └──────┘  └──────┘  └──────┘
```

Distributed File System
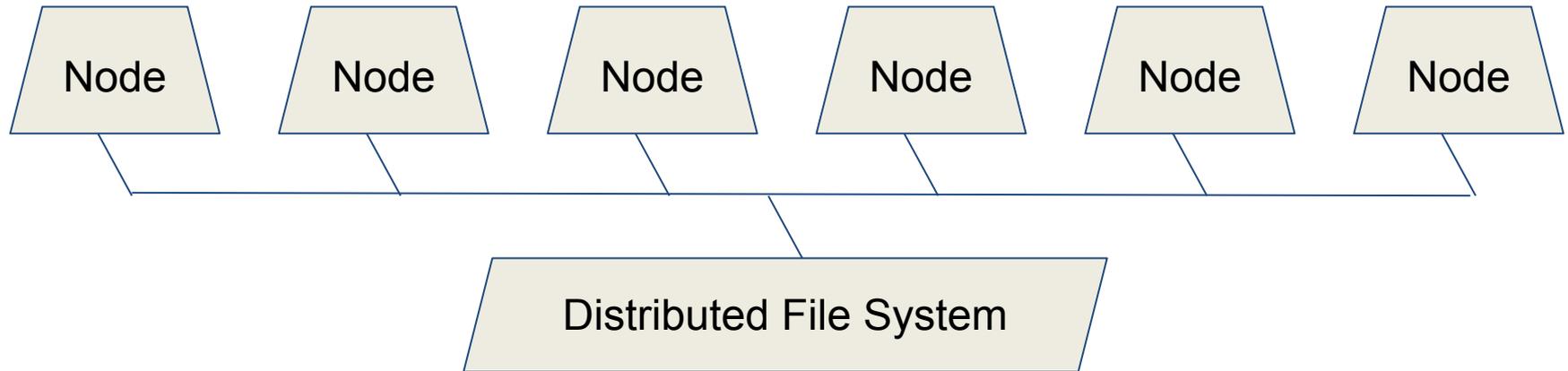
- Working with external infrastructure provider
- Kubernetes Persistent Volumes
- Partition volume mounts?
    - By user
    - By data set
    - By storage profile (I/O performance, reliability)

NRC·CNRC

Launching Interactive Images (concept only)

Enter image URL:

⦿ https://dockerhub.com/bmajor/extractor

Or pick an image:

◯ CASA version 5.4.1 ▼

Interactive or Batch?

⦿ Interactive   ◯ Batch

Image parameters:

| |

Min cores   4 ▼   Min RAM   100 GB ▼

Launch

Data set:

data link query...

Job List

ID	Image	Status	
KJ7L	ds-9	EXECUTING	details
H113	pyvo	EXECUTING	details

NRC·CNRC

Launching Batch Images (concept only)

Enter image URL:

⦿ https://dockerhub.com/bmajor/extractor

Or pick an image:

◯ CASA version 5.4.1 ▼

Interactive or Batch?

◯ Interactive   ⦿ Batch

How many instances?

1000 ▼

Data set:

data link query...

Image parameters:

|

Or parameter input file:

https://ws-canfar/vault/files/bmajor/project1/input.txt ▼

Min cores 4 ▼   Min RAM 100 GB ▼

Launch

Job List

ID	Image	Status	
KJ7L	extractor	EXECUTING	details
H113	extractor	EXECUTING	details
J26F	extractor	EXECUTING	details
98T2	extractor	EXECUTING	details
P8C3	extractor	EXECUTING	details
3SJ3	extractor	EXECUTING	details
7Y3N	extractor	ERROR	details
B382	extractor	COMPLETED	details
L1N2	extractor	COMPLETED	details
JZA3	extractor	COMPLETED	details
IUY7	extractor	COMPLETED	details

X: 587 px   Y: 512 px

NRC·CNRC

Interoperability & Standardization

Use case: Run software on data sets hosted at different institutions

Users want:

- To write and debug a software software container at a single site in interactive mode, then

- Run that container on other sites with on other data sets

NRC·CNRC

Interoperability

- We want to be a part of a greater, connected astronomy science platform, a ***Science Platform Network***.

  (Vandana Desai et al. *A Science Platform Network to Facilitate Astrophysics in the 2020s*)

- Members of the Science Platform Network have
  - certain hosted telescope archives
  - data-specific software and processing capabilities
  - certain infrastructure capabilities (eg GPUs for ML)

NRC·CNRC

What could be standardized?

1. Software/User containers (notebooks, legacy tools, custom software
    a. what are the minimum requirements?
        i. Non-root user 'guest'
        ii. ENV variables to find storage resources

2. REST API Interactive and batch container execution
    a. UI options look very similar to early CANFAR with VMs, hints at stability of requirements for general purpose processing
        i. UWS for job mgmt
        ii. job details contain connect URLs for inter
    b. Solution must integration with VO A&A recommendations

3. Storage and Input data -->

Standardization of Storage and Input data

API access -- download the data during image execution
- LSST butler
- VOSpace with HTTP protocols
- DataLink

File system / POSIX access
- Legacy programs expect a file system
- Pre-mount ahead of image startup

Hybrid:
- VOSpace with Mount protocols (Cavern)
- Mounts made at start of image execution

NRC·CNRC

Open Source Platform Development

- The process encourage standardization -- it has to work for everyone!

- Both platform infrastructure and software containers could be open-sourced

- Software containers would need to pass a set of tests to ensure they meet the minimum required to execute on different platforms at different institutes with different data sets.

- Let's get started!

NRC·CNRC

Summary

- The CADC (CANFAR and Arcade) have exciting opportunities for Science Platforms ahead.

- We would like to collaborate with other groups and be part of an astronomy science platform network

- I have outlined three areas (up for discussion) that could be standardized for interoperability:

    1. Software containers and notebooks
    2. Platform infrastructure and integration with A&A
    3. How data and storage is made available to the software

NRC·CNRC

Astronomy and Computing

Call for Papers – topical issue on astronomical science platforms

Submission deadline: **November 1st, 2019**

NRC·CNRC