Spark data analysis platform for Gaia DR3

Dave Morris
<dmr@roe.ac.uk>

D.Morris
Institute for Astronomy,
Edinburgh University

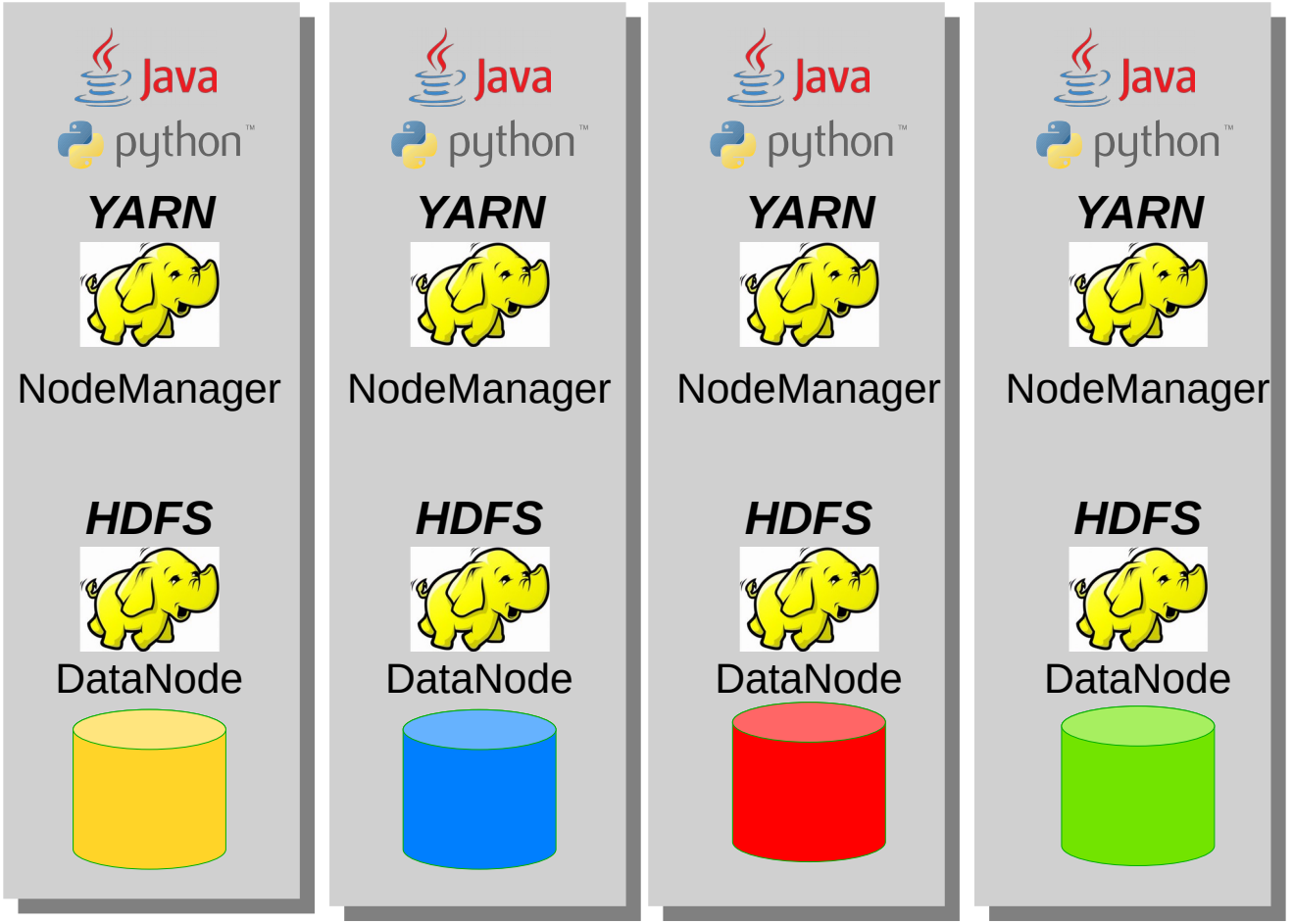
Gaia CU9:plenary
January 2020

Gaia Data Analytics Framework (GDAF)

+++ Live service running now +++

GDAF

physical hardware

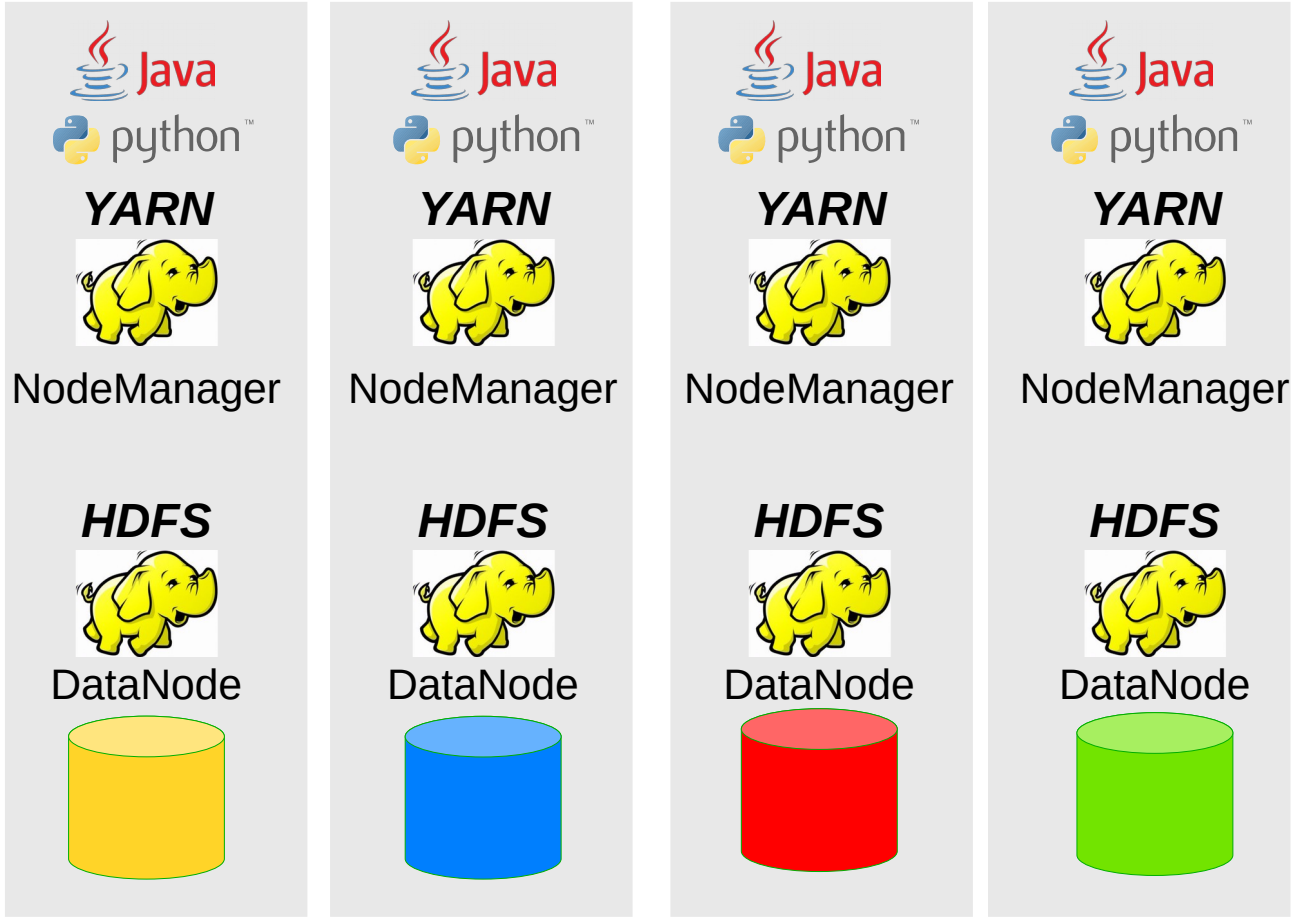


D.Morris
Institute for Astronomy,
Edinburgh University

Spark and Hadoop on OpenStack

+++ Live service running now +++

Cloud compute

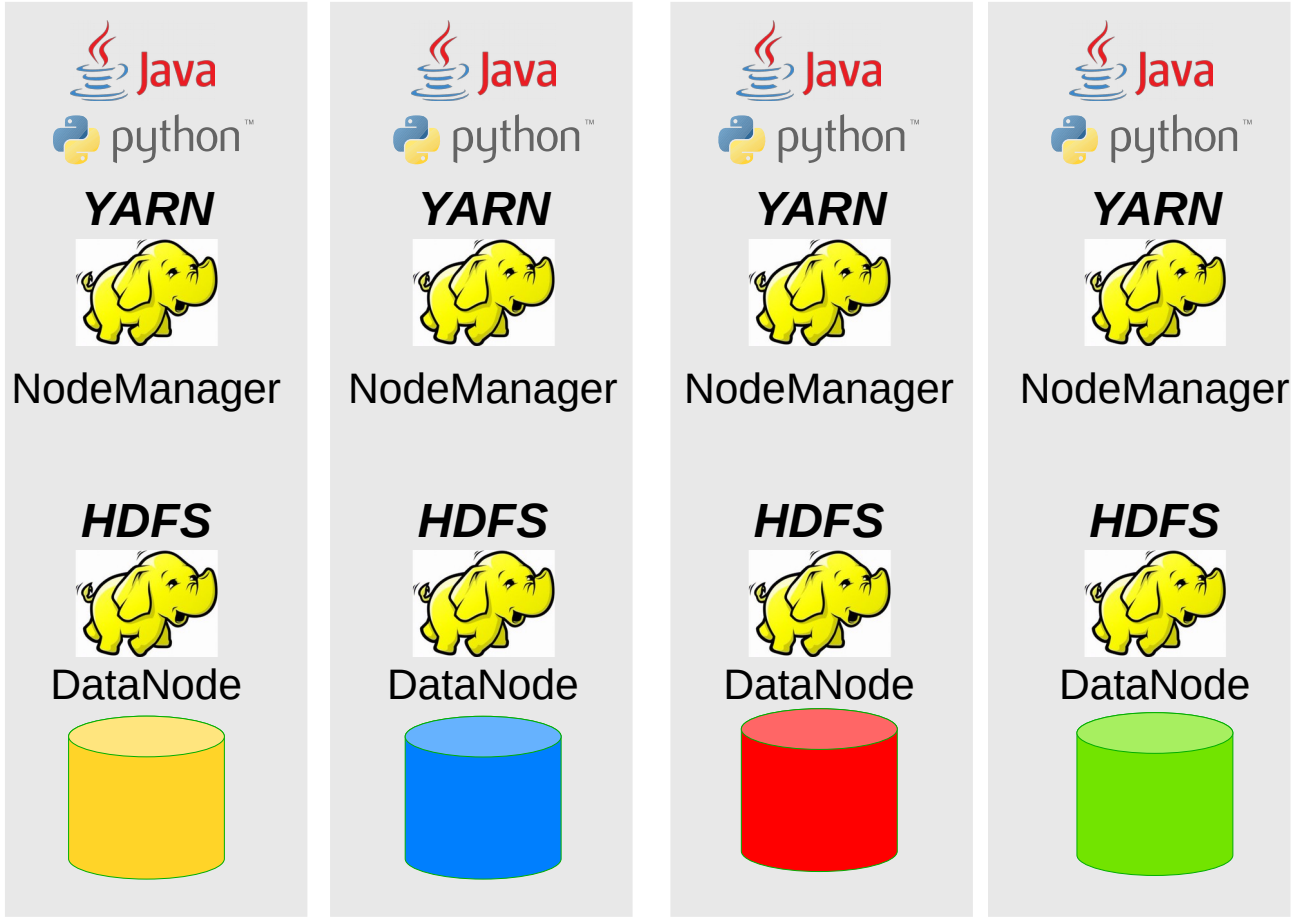


D.Morris
Institute for Astronomy,
Edinburgh University

Ansible automated deployment

+++ available next week +++

Cloud compute



D.Morris
Institute for Astronomy,
Edinburgh University

ROYAL OBSERVATORY EDINBURGH

Gaia DPAC

Automated deployment

3 separate cloud systems

APACHE Spark™

YARN hadoop

openstack™

Cloud compute

gaia-dev Development system

20 virtual machines
200 cpu cores 1Tbyte RAM
4Tbyte volume storage
4Tbyte object storage

gaia-test Continuous integration platform (*) April 2020

20 virtual machines
200 cpu cores 1Tbyte RAM
4Tbyte volume storage
4Tbyte object storage

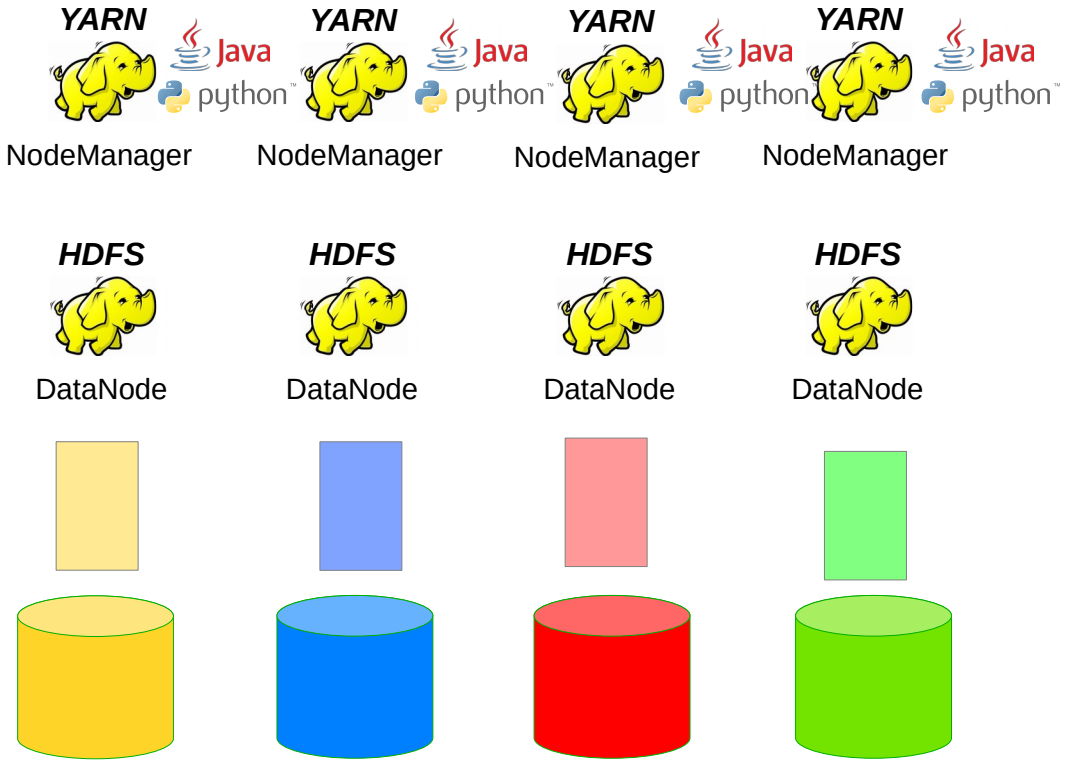
gaia-prod Production system (*) April 2020

20 virtual machines
200 cpu cores 1Tbyte RAM
4Tbyte volume storage
4Tbyte object storage

D.Morris
Institute for Astronomy,
Edinburgh University

THE UNIVERSITY OF EDINBURGH

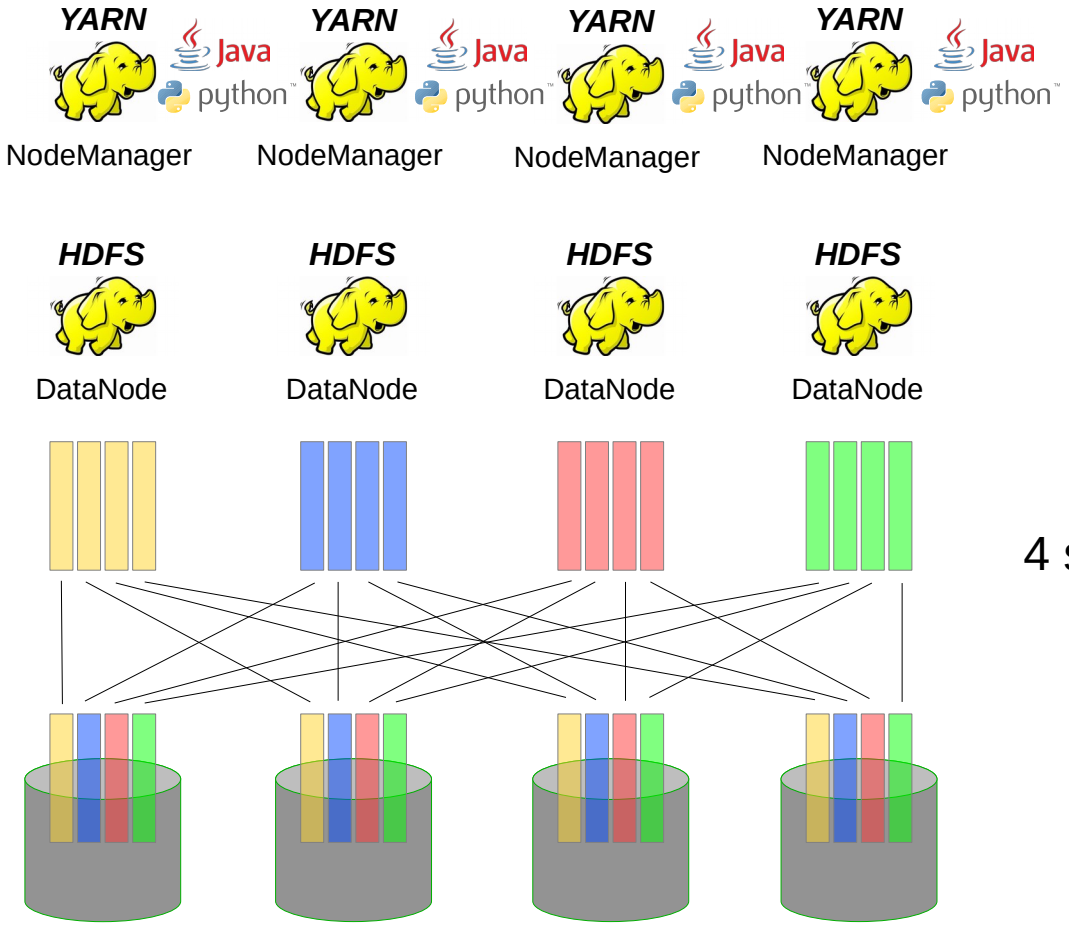
Spark and Hadoop



Logical allocation of one data disc per node

D.Morris
Institute for Astronomy,
Edinburgh University

Hadoop Distributed File System (HDFS)



4 data files

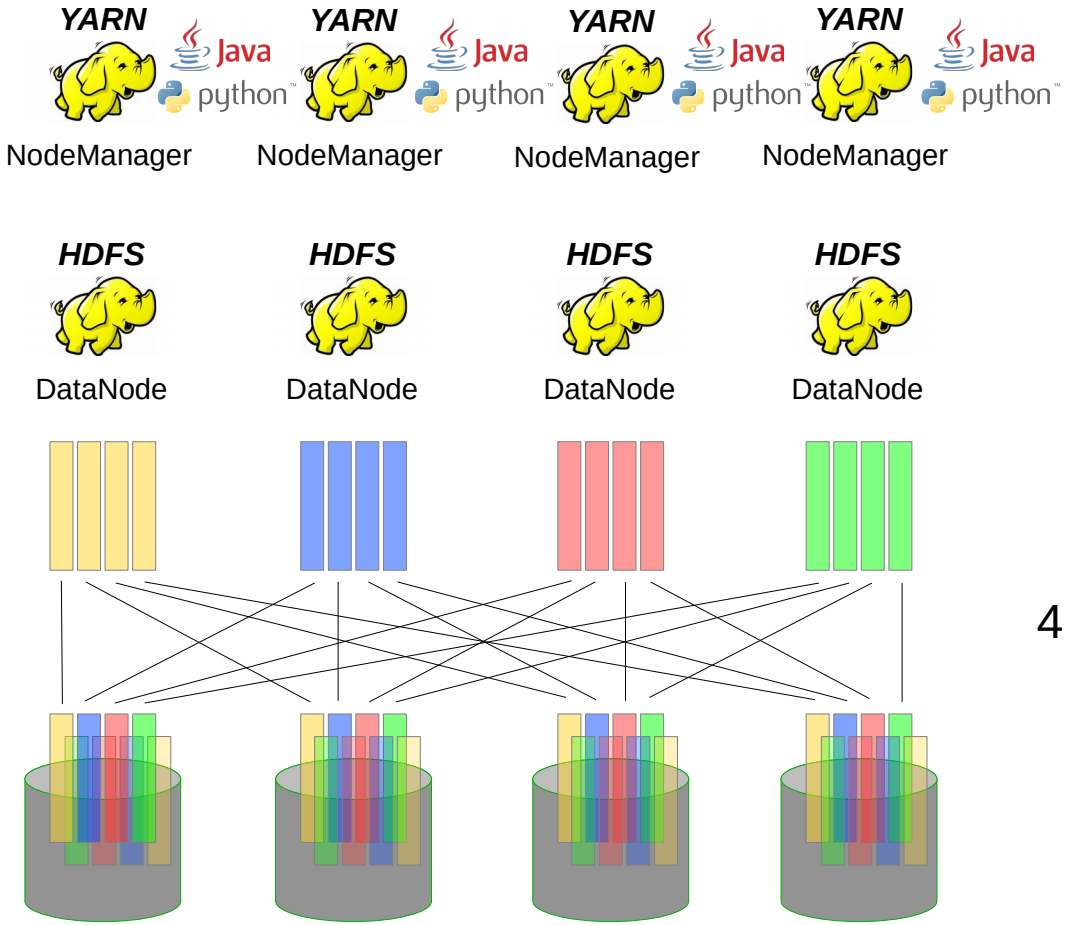
4 stripes per file

16 stripes

Data distributed in stripes

D.Morris
Institute for Astronomy,
Edinburgh University

Hadoop Distributed File System (HDFS)



4 data files

4 stripes per file

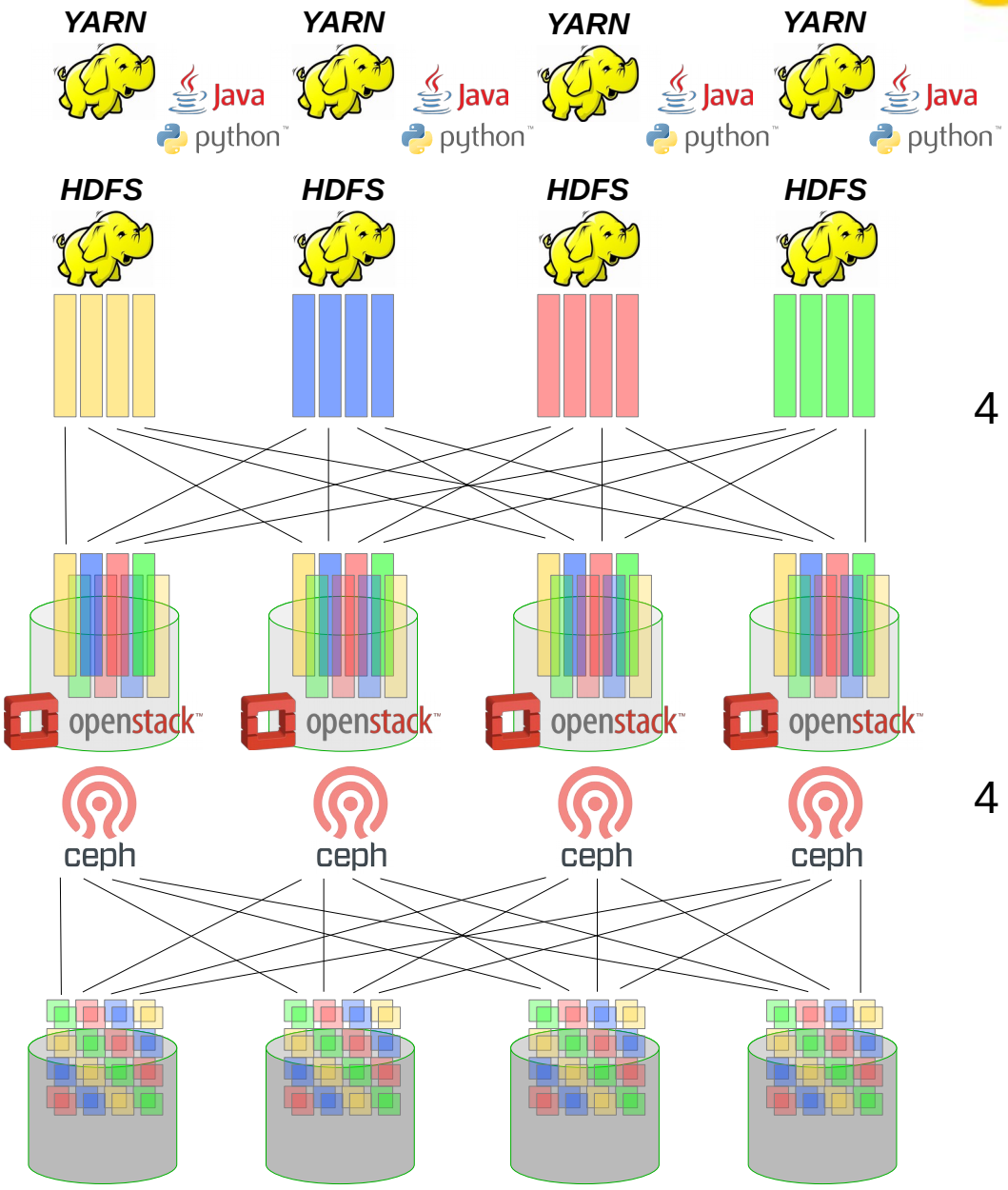
2 replicas

32 stripes

Stripes replicated for redundancy

D.Morris
Institute for Astronomy,
Edinburgh University

HDFS on top of Ceph



4 data files

4 stripes per file

2 replicas

32 stripes

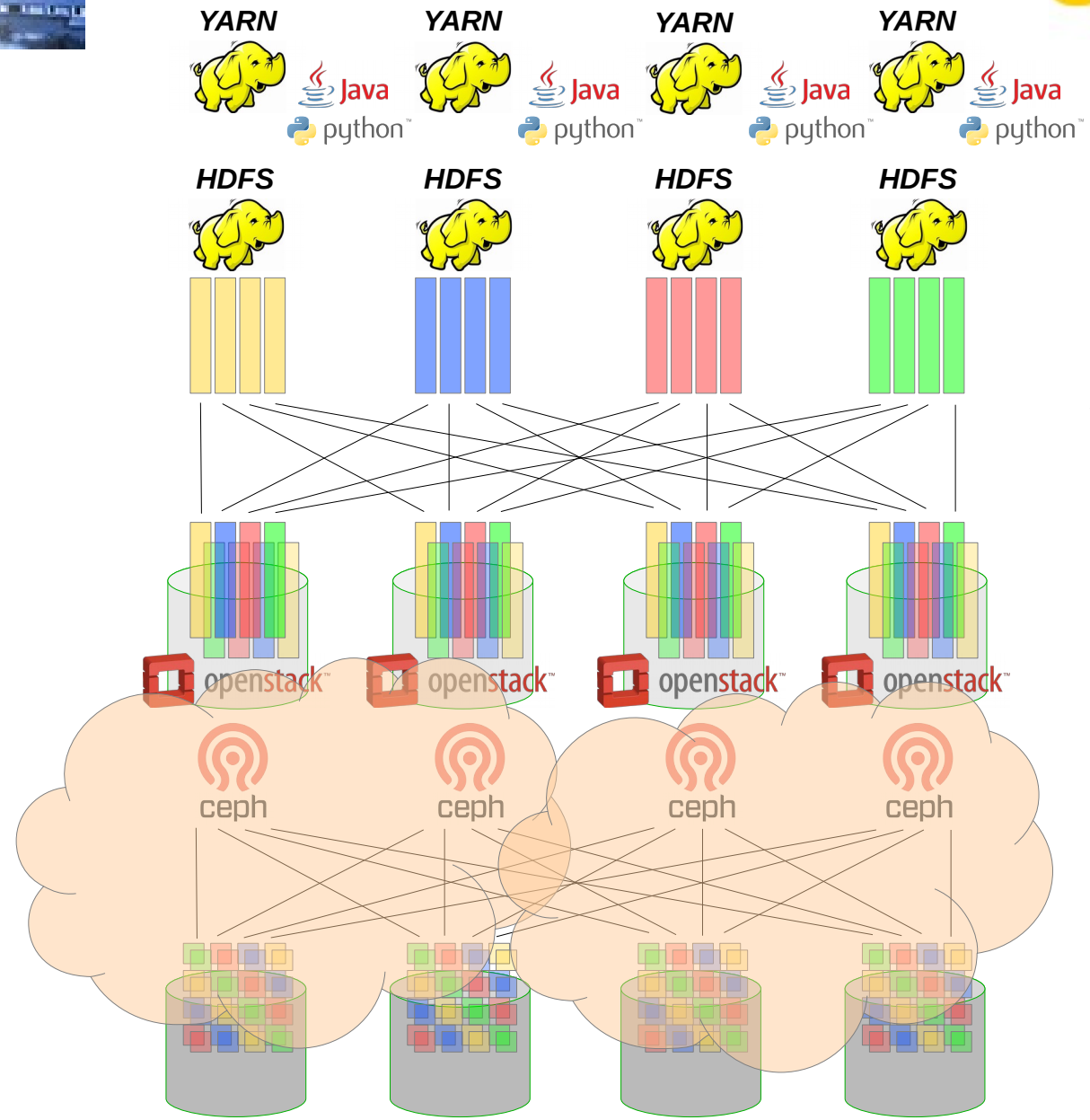
4 stripes per file

2 replicas

256 stripes

D.Morris
Institute for Astronomy,
Edinburgh University

HDFS on top of Ceph



virtual disc

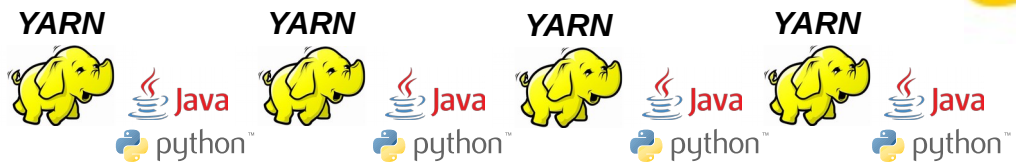
virtual server

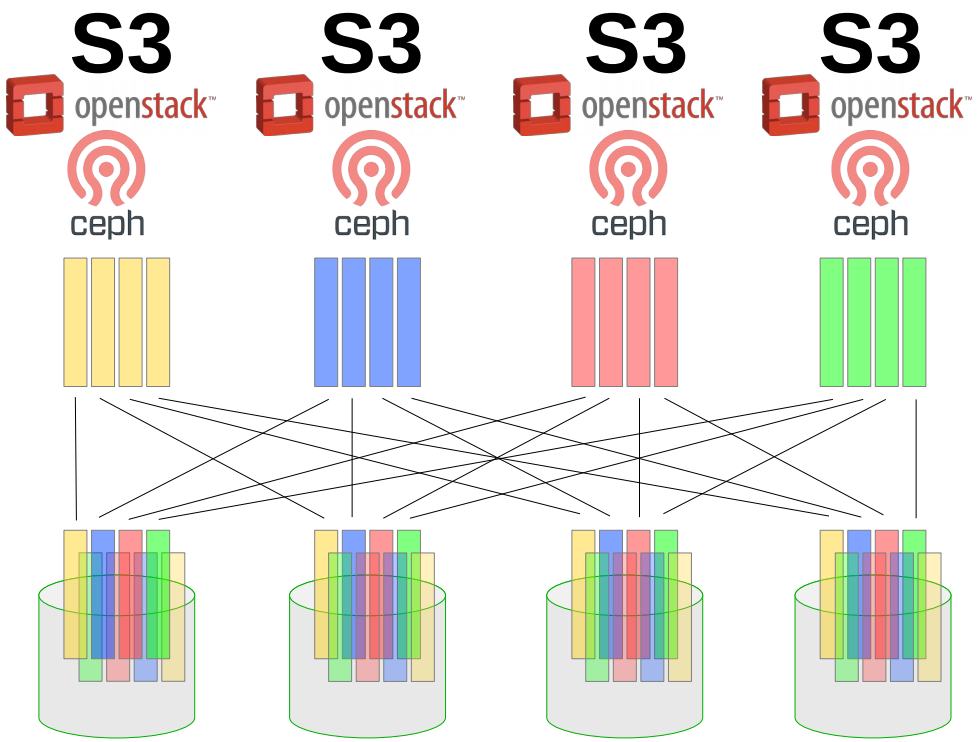
virtual rack

no location knowledge

D.Morris
Institute for Astronomy,
Edinburgh University

Simple Storage Service (S3)





OpenStack

Swift

ObjectStore

S3 interface

Ceph

RADOS
Object-based
storage system

Reliable Autonomic Distributed Object store

D.Morris
Institute for Astronomy,
Edinburgh University

Simple Storage Service (S3)

+ve Simple Storage Service – easy to implement

+ve Several providers of S3 compatible object storage

Simple Storage Service (S3)

Azure

-ve HTTP based transfers are slow

D.Morris
Institute for Astronomy,
Edinburgh University

Network File System (NFS)

YARN YARN YARN YARN

NFS NFS NFS NFS

Network File System

+ve Standard file system interface

+ve Several providers for NFS accessible storage

Network File System (NFS)

Azure

+ve NFS transfers are faster

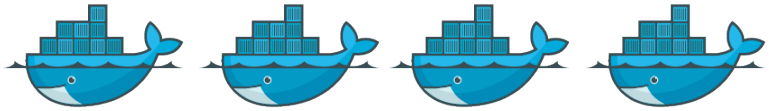
D.Morris
Institute for Astronomy,
Edinburgh University

Spark Kubernetes scheduler

Replace YARN with Kubernetes

-ve Kubernetes scheduler is still experimental

+ve ‘unit’ of computing changes from virtual machine to container



+ve dynamic scaling in response to load

D.Morris
Institute for Astronomy,
Edinburgh University

Managed Kubernetes cloud

Kubernetes manages pods of containers

+ve Several providers for managed Kubernetes cloud

Azure

D.Morris
Institute for Astronomy,
Edinburgh University

Managed Spark cloud

+ve Several providers for managed Spark cloud

Azure

+ve Managed system is very easy to set up

-ve Each provider deploys Spark slightly differently

-ve Platform specific configuration

-ve Platform specific interfaces

D.Morris
Institute for Astronomy,
Edinburgh University

Managed Kubernetes cloud

Highest inter-operable (*) common (**) abstraction interface

Azure

(*) in theory (***) probably

(**) kind of

D.Morris
Institute for Astronomy,
Edinburgh University

Container Storage Interface (CSI)

Container Storage Interface (CSI)

Inter-operable (*) standard (**) storage interface

+ve Several providers implementing CSI

Azure

D.Morris
Institute for Astronomy,
Edinburgh University

ROYAL OBSERVATORY EDINBURGH

"Kubernetes is the new OS"

Gaia DPAC

APACHE Spark™

kubernetes

CONTAINER STORAGE INTERFACE

Kubernetes is becoming
a microkernel OS

(Tanenbaum Torvalds)

Container Storage Interface (CSI)

Inter-operable (*) storage components

Prediction for 2020 …. 2022

Container Network Interface (CNI)

Inter-operable (*^2) network components

Container Compute Interface (CCI)

Inter-operable (*^3) compute components

D.Morris
Institute for Astronomy,
Edinburgh University

“In the cloud” means *“other people’s computers”*

D.Morris
Institute for Astronomy,
Edinburgh University

Something unexpected has happened

…. thank you for your patience

Gaia CU9:plenary
January 2020

Managed Kubernetes cloud

Inter-operable (*) common (**) abstraction layer

Portable deployment

If one cloud is unavailable, we can deploy on another
If one cloud is too expensive, we can deploy on another

Azure

(*) in theory (***) probably

(**) kind of

D.Morris
Institute for Astronomy,
Edinburgh University

Project goals

Immediate goal ...

OpenStack deployment capable of handling DR3

Long term goal ...

portable deployment to multiple clouds

Thank you

Dave Morris
<dmr@roe.ac.uk>

D.Morris
Institute for Astronomy,
Edinburgh University