MAX PLANCK
COMPUTING & DATA FACILITY

MPCDF

WHAT CAN HPC CLOUD DO FOR YOU?

What can you do for the HPC cloud?

Frank Berghaus
for the HPC Cloud Team

OVERVIEW

MAX PLANCK COMPUTING & DATA FACILITY

- Features of the HPC Cloud
  - Compute
    - *Images*
    - *Flavors*
    - *Server Groups*
  - Volumes, i.e. Block storage
  - Network
    - *Networks*
    - *Security Groups*
    - *Load Balancers (not discussed)*
    - *Floating IPs (not discussed)*
  - Orchestration
  - Object storage
  - Shares, i.e. shared file systems

- Recipes for the HPC Cloud
  - Kubernetes (MoM)
  - SLURM (JADE)
  - Workstations
  - VPN (?)

- Services based on HPC Cloud
  - Notebooks:
    - *BinderHub & JupyterHub*
  - GitLab runners
  - Virtual cluster
    - *Example: remote visualization service (RVS)*
  - Open Build Service (OBS)

OVERVIEW

MAX PLANCK COMPUTING & DATA FACILITY

- Features of the HPC Cloud
  - Compute
    - *Images*
    - *Flavors*
    - *Server Groups*
  - Volumes, i.e. Block storage
  - Network
    - *Networks*
    - *Security Groups*
    - *Load Balancers (not discussed)*
    - *Floating IPs (not discussed)*
  - Orchestration
  - Object storage
  - Shares, i.e. shared filesystems

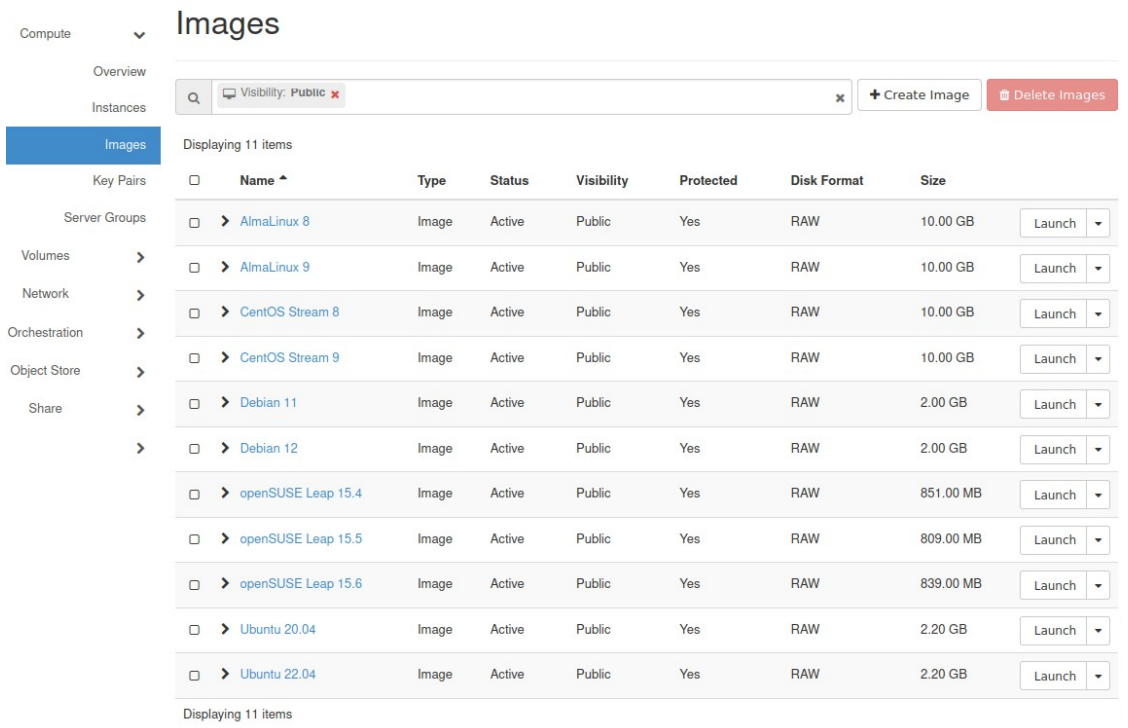
- Recipes for the HPC Cloud
  - Kubernetes (MoM)
  - SLURM (JADE)
  - Workstations
  - VPN (?)

- Services based on HPC Cloud
  - Notebooks:
    - *BinderHub & JupyterHub*
  - GitLab runners
  - Virtual cluster
    - *Example: remote visualization service (RVS)*
  - Open Build Service (OBS)

FEATURES → COMPUTE → IMAGES

MAX PLANCK COMPUTING & DATA FACILITY

- EL-Like:
  - Alma
  - Centos
  - RHEL (BYOL)
- Debian-like
  - Debian
  - Ubuntu
- SUSE-like
  - OpenSUSE
  - SLES (BYOL)
- Wishlist?
  - OS's: Arch, Container OS's, Windows
  - Local repo mirrors / regular updates
  - Pre-installed "appliances" (e.g. Prometheus)



- OS versions removed at EOL
- Updates are manual / irregular

FEATURES → COMPUTE → FLAVORS

MAX PLANCK COMPUTING & DATA FACILITY

- Defined by mapping to hardware
- Shared (default):
  - Two VCPUs share a physical CPU
- Dedicated (on request):
  - VCPUs are pinned to physical CPUs
- Devices
  - GPUs: A30/A40, A100, H100
  - NVMe: 1.8 TB & 3.6 TB
- Wishlist:
  - GPU flavor sizing??
  - Hyper-threaded flavors

RAM per VCPU [GiB]

	shared	dedicated
standard	4	6.7
high	8	---
extreme	16	30

Maximum size

	shared	dedicated	whole-node
standard	64c256g	36c240g	72c480g
high	64c512g	---	---
extreme	32c512g	32c960g	64c1920g

FEATURES → COMPUTE → SERVER GROUPS

MAX PLANCK COMPUTING & DATA FACILITY

- Hint to the OpenStack scheduler

```
o server group create demo –-policy soft-affinity
O server create demo –-server-group demo ...
```

policy	effect
soft-affinity	Attempt to schedule instances near each other, for example on the same hypervisor
soft-anti-affinity	Attempt to schedule instances on different hardware, for example separate racks

MAX PLANCK COMPUTING & DATA FACILITY

- Block storage provided by Ceph
- Three variants:
  1) Ceph
  2) CephSSD
  3) CephIntensive
- Little difference in performance
- Faster block devices
  - ➔ Flavors with NVMe
- Wishlist
  - ➔ NVMeOF (?)

- Volume as root disk
  - ➔ Create a volume from an image
  - ➔ Mark the volume “bootable”
  - ➔ Create server from the volume
- Multi-attach volumes
  - ➔ Available on request
  - ➔ Must set up parallel file system

FEATURES → NETWORK → NETWORKS

MAX PLANCK COMPUTING & DATA FACILITY

- Networks shared by all projects
  - cloud-local-1
    - *very full*
  - cloud-local-2
    - *full*
  - cloud-local-3
- Public IPs
  - cloud-public
- You can define private networks
  - Network
  - Subnet
  - Router

- Special networks for shares
  - Access given on request

```
o network create NETWORK
o subnet create SUBNET \
      --network NETWORK \
      --subnet-range 192.168.0.0/24 \
      --dns-nameserver 130.183.9.32 \
      --dns-nameserver 130.183.1.21
o router create ROUTER
o router set ROUTER --external-gateway cloud-public
o router add subnet ROUTER SUBNET
```

FEATURES → NETWORK → SECURITY GROUPS

MAX PLANCK COMPUTING & DATA FACILITY

- Firewall rules
- We setup up three for you
  - default
  - web
  - web-campusonly
- Notes on default group
  - Allows ingress from your hosts
  - Blocks ingress from other projects
  - Allows SSH from campus

- AutoDNS:
  - DNS entry for FIPs and Servers
  - Runs every 5 minutes
  - Servers:

VM_NAME.**PROJECT_NAME**.*hpccloud.mpg.de*

  - FIPs:

FIP_DESCRIPTION.**PROJECT_NAME**.*hpccloud.mpg.de*

- Certificates
  - CertBot & Let's Encrypt
  - CA of your choice

MAX PLANCK COMPUTING & DATA FACILITY

- Addressing the storage
  - S3: Amazon's protocol
    - *de facto standard*
  - SWIFT: OpenStack protocol
- Credentials issues by OpenStack
  - Provide full access to project space
- Globally accessible
- Wishlist:
  - How to make this useful to you?

Paul Müller, Wed @ 9:00

Storage in the Cloud, Wed @ 15:00

- Share content
  - Handle access control in separate application
  - Generate links with temporary access tokens

MAX PLANCK COMPUTING & DATA FACILITY

- Shared file system as a service
- Three flavors:
  - NexusPOSIX
    - *Native on HPCs*
    - *Configured for large data files*
    - *Uses MPCDF user & group IDs*
    - *Projects can be exported to cloud via NFS*
  - Manila & NFS
    - *Easy to set up*
    - *Widely compatible*
    - *Low performance (bottleneck)*
  - Manila & CephFS
    - *Requires compatible client*
    - *More configuration*
    - *Good performance*

- Wishlist:
  - Recipes / examples (?)

Storage in the Cloud, Wed @ 15:00

FEATURES → ORCHESTRATION

MAX PLANCK COMPUTING & DATA FACILITY

HEAT

- Defined in yaml
- State tracked in OpenStack
- *Example:* Kubernetes (MoM)

Terraform

- Defined in hcl
- State local to client
  - ➔ Shared state via GitLab
- *Example:* SLURM (JADE)

Automation & Infrastructure as code, Wed @ 13:30

- Define compute, storage, and network infrastructure as code
- Collaborate, reproduce, or share systems of compute resources

OVERVIEW

MAX PLANCK COMPUTING & DATA FACILITY

- Features of the HPC Cloud
  - Compute
    - *Images*
    - *Flavors*
    - *Server Groups*
  - Volumes, i.e. Block storage
  - Network
    - *Networks*
    - *Security Groups*
    - *Load Balancers (not discussed)*
    - *Floating IPs (not discussed)*
  - Orchestration
  - Object storage
  - Shares, i.e. shared filesystems

- Recipes for the HPC Cloud
  - Kubernetes (MoM)
  - SLURM (JADE)
  - Workstations
  - VPN (?)

- Services based on HPC Cloud
  - Notebooks:
    - *BinderHub & JupyterHub*
  - GitLab runners
  - Virtual cluster
    - *Example: remote visualization service (RVS)*
  - Open Build Service (OBS)

MAX PLANCK COMPUTING & DATA FACILITY

- MoM Recipe:
  gitlab.mpcdf.mpg.de/mpcdf/cloud/kubernetes
- Two branches:
  - production:
    - *Meant to host productive services*
    - *Three control plane nodes*
    - *Problem: warnings from etcd*
  - test:
    - *Meant to try out Kubernetes*
    - *Single control plane*

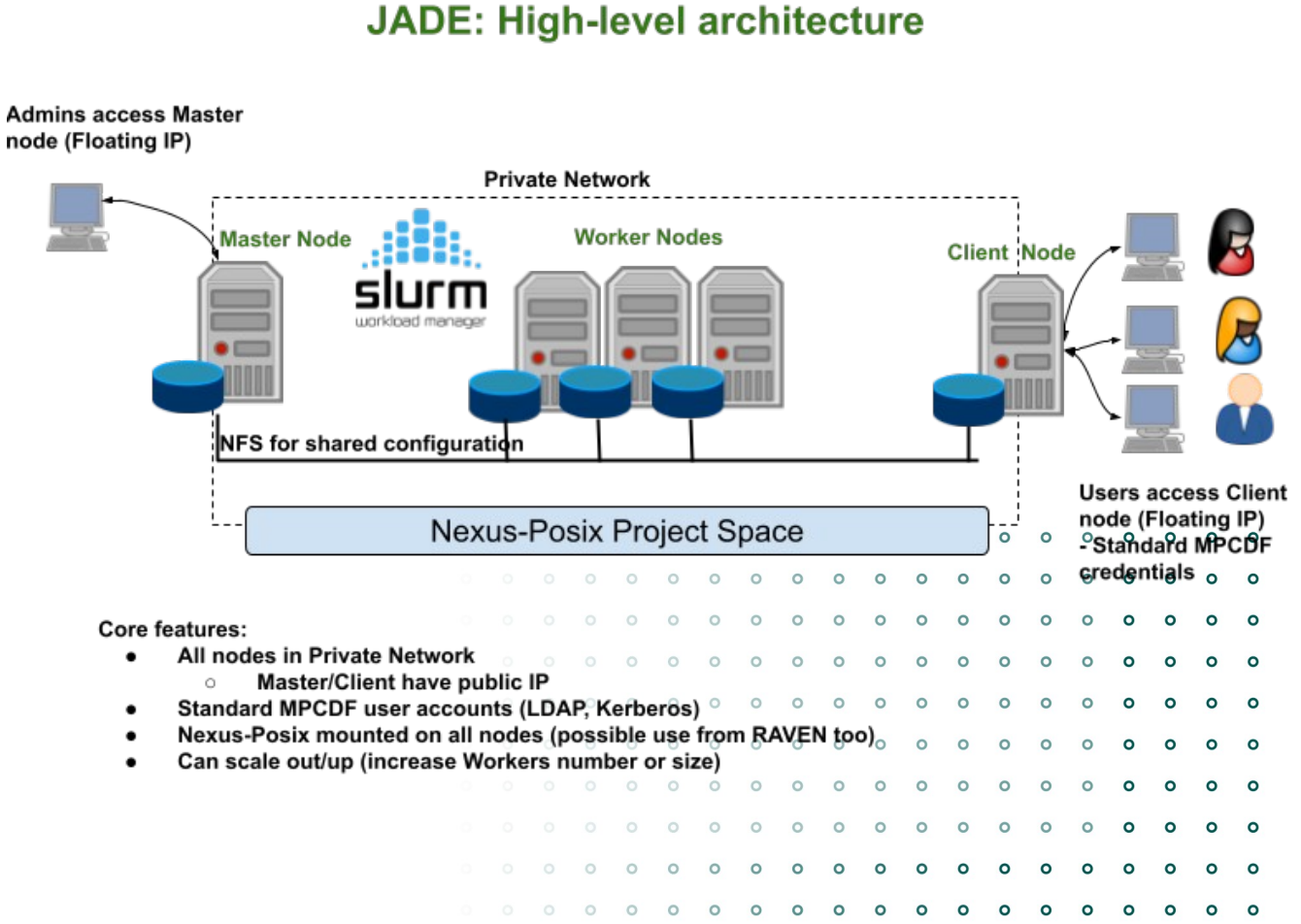
Kubernetes, Containers etc., Wed @ 13:00

- Use:
  - Notebooks
  - Many projects, at the workshop:
    - *Flaski*
    - *Nomad*
    - *SciServer*
- Wishlist
  - Improve existing recipe
  - Other deployment options (?)
    - *Magnum*
    - *Rancher*
    - *ClusterAPI*

MAX PLANCK COMPUTING & DATA FACILITY

- JADE Recipe:
  - gitlab.mpcdf.mpg.de/mpcdf/cloud/jade
- Slurm Cluster in the Cloud
  - Full control of your cluster config and software environment
  - Shared Storage on NexusPOSIX
    - *Also mounted on Raven or Cloud VMs*
  - Good for single-node workloads
  - Terraform based
    - *Scale out/in easily*
    - *Cleanly spin up or tear down clusters*

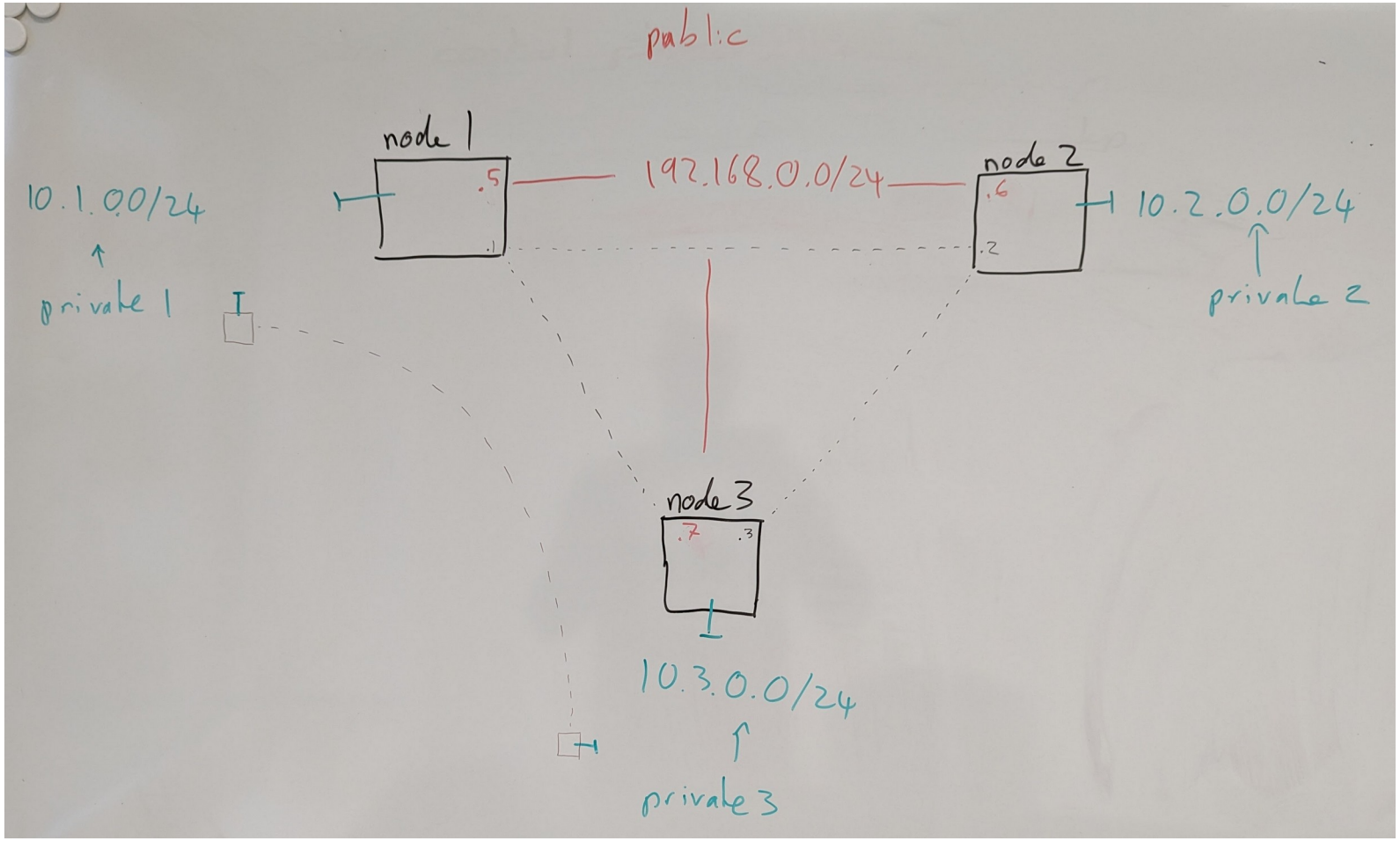
Clusters in the Cloud, Wed @ 15:00



MAX PLANCK COMPUTING & DATA FACILITY

- Workstations in the cloud
  - ➔ High degree of control over infrastructure
    - *Size of VM*
    - *Storage*
  - ➔ Local or MPCDF Accounts (LDAP)
  - ➔ NexusPOSIX mount possible
  - ➔ Ideal for long running post processing jobs
    - *Production on Raven, data store in NexusPOSIX, post processing on workstation*
  - ➔ Possible mount of block storage volumes
  - ➔ Full control over software environment
  - ➔ Remote Desktop (VNC)
  - ➔ Access via SSH through Gateways

RECIPES → VPN (?)



- WIP Example:
  gitlab.../fberg/vpn-demo
- Wishlist:
  - Recipe (WG mesh)
  - OpenStack VPNaaS
- Interest (?)

Maurits & Simeon, Tue @ 16:30

OVERVIEW

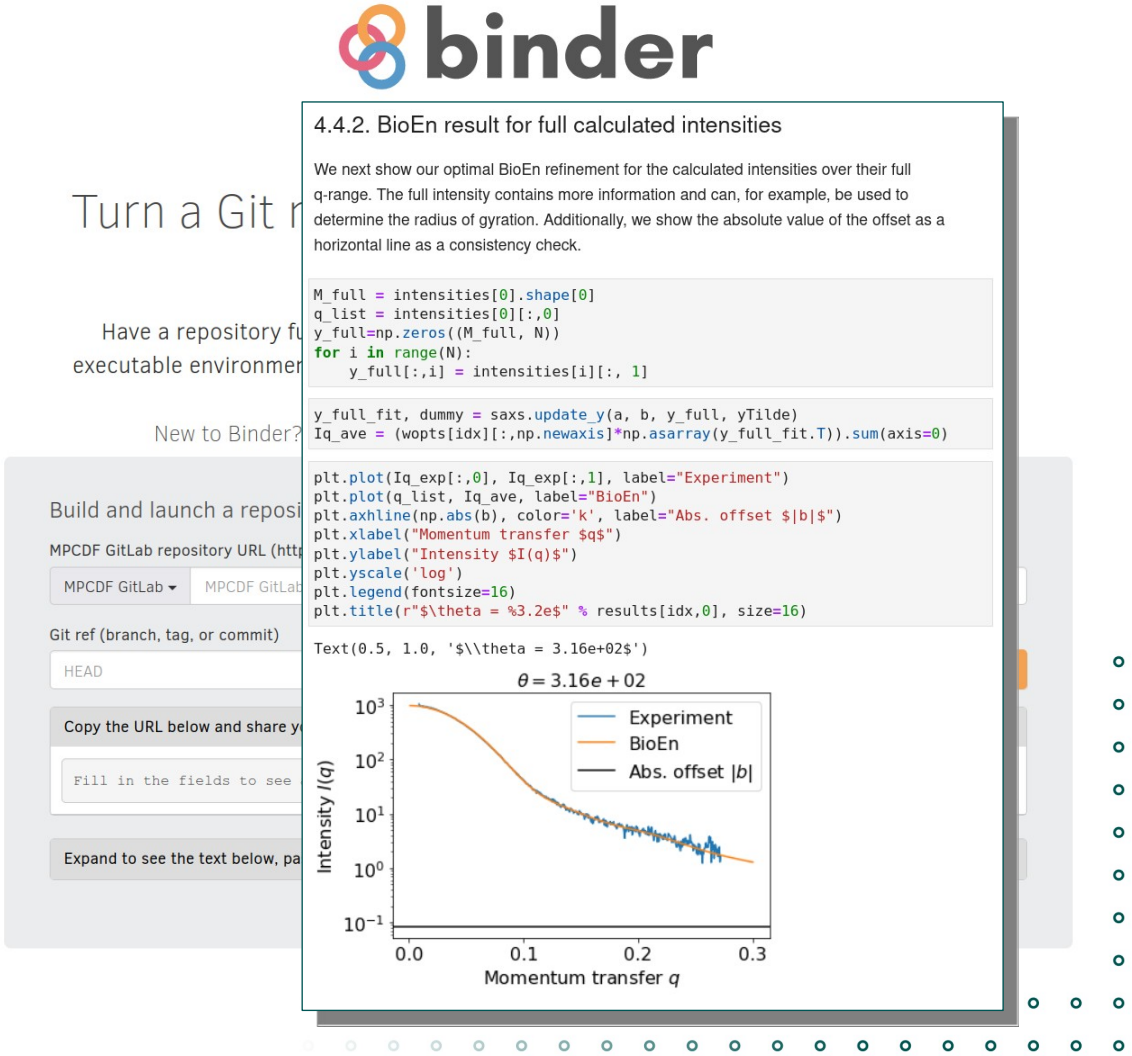
- Features of the HPC Cloud
  - Compute
    - *Images*
    - *Flavors*
    - *Server Groups*
  - Volumes, i.e. Block storage
  - Network
    - *Networks*
    - *Security Groups*
    - *Load Balancers (not discussed)*
    - *Floating IPs (not discussed)*
  - Orchestration
  - Object storage
  - Shares, i.e. shared file systems

- Recipes for the HPC Cloud
  - Kubernetes (MoM)
  - SLURM (JADE)
  - Workstations
  - VPN (?)

- Services based on HPC Cloud
  - Notebooks:
    - *BinderHub & JupyterHub*
  - GitLab runners
  - Virtual cluster
    - *Example: remote visualization service (RVS)*
  - Open Build Service (OBS)

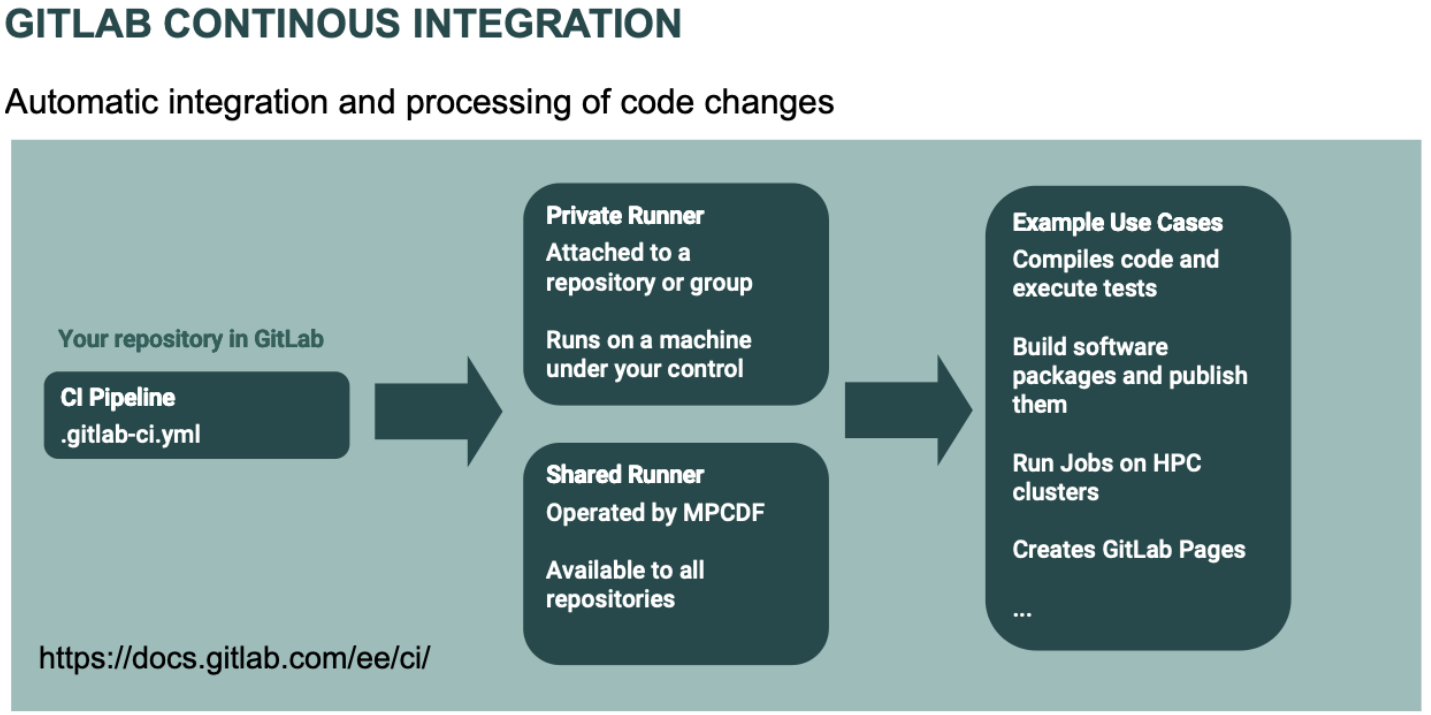
MAX PLANCK COMPUTING & DATA FACILITY

- Service:
  - notebooks.mpcdf.mpg.de/binder
- Binder:
  - Run notebooks from git repo
  - Restricted to MPCDF GitLab
- JupyterHub
  - Create JupyterHubs for projects
  - e.g. Workshops
- Wishlist
  - Authentication
    - *EduGAIN for Binder*
  - Plugin to Storage (?)



MAX PLANCK COMPUTING & DATA FACILITY

- Service:
  - gitlab.mpcdf.mpg.de
- Run CI/CD jobs
  - Shared runners
  - Group/projects runners



MAX PLANCK COMPUTING & DATA FACILITY

- Managed SLURM cluster
  - For large projects
    - *Solution provided by MPCDF Cluster team on demand when fitting*
  - Caveats:
    - *Bound to MPCDF accounts*
    - *Limited customization*
- Examples:
  - FHI General
  - Pirol
  - Robin / RVS
- Alternative:
  - SLURM (JADE) recipe
  - Build your own from scratch

Clusters in the Cloud, Wed @ 15:00

MAX PLANCK COMPUTING & DATA FACILITY

- Internal project
- Builds software for HPC systems

WRAP UP

MAX PLANCK COMPUTING & DATA FACILITY

- Features of the HPC Cloud
  - Compute
    - *Images*
    - *Flavors*
    - *Server Groups*
  - Volumes, i.e. Block storage
  - Network
    - *Networks*
    - *Security Groups*
    - *Load Balancers (not discussed)*
    - *Floating IPs (not discussed)*
  - Orchestration
  - Object storage
  - Shares, i.e. shared file systems

- Recipes for the HPC Cloud
  - Kubernetes (MoM)
  - SLURM (JADE)
  - Workstations
  - VPN (?)

- Services based on HPC Cloud
  - Notebooks:
    - *BinderHub & JupyterHub*
  - GitLab runners
  - Virtual cluster
    - *Example: remote visualization service (RVS)*
  - Open Build Service (OBS)

THANK YOU