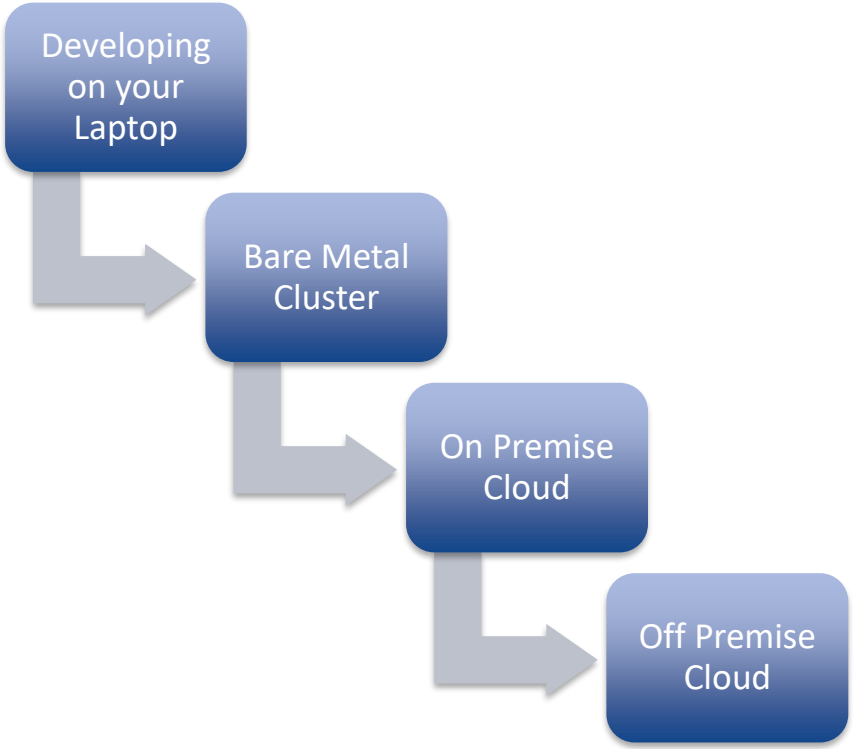
NEXUS Deployments

Frank Greguska
Jet Propulsion Laboratory, California Institute of Technology

Team:
Nga Quach
Thomas Huang
Joseph Jacob
Brian Wilson

Overview

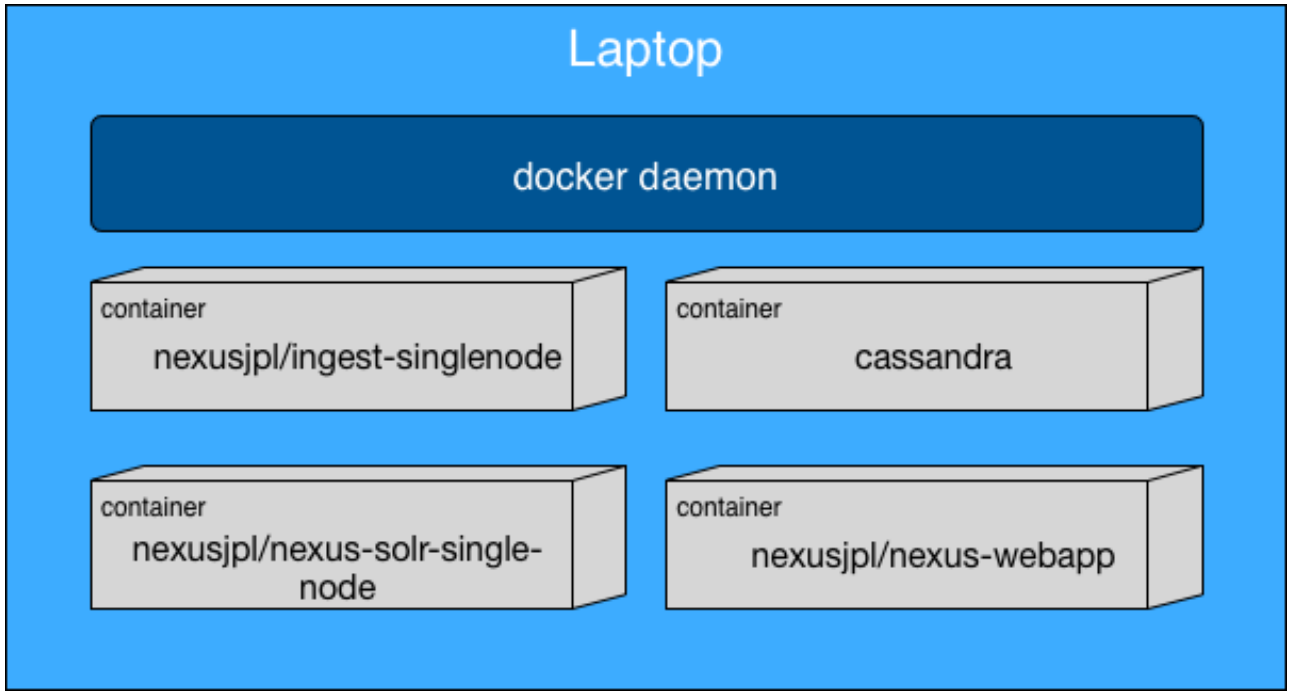
- Installing NEXUS from Local to Cluster to Cloud



- Other Installations

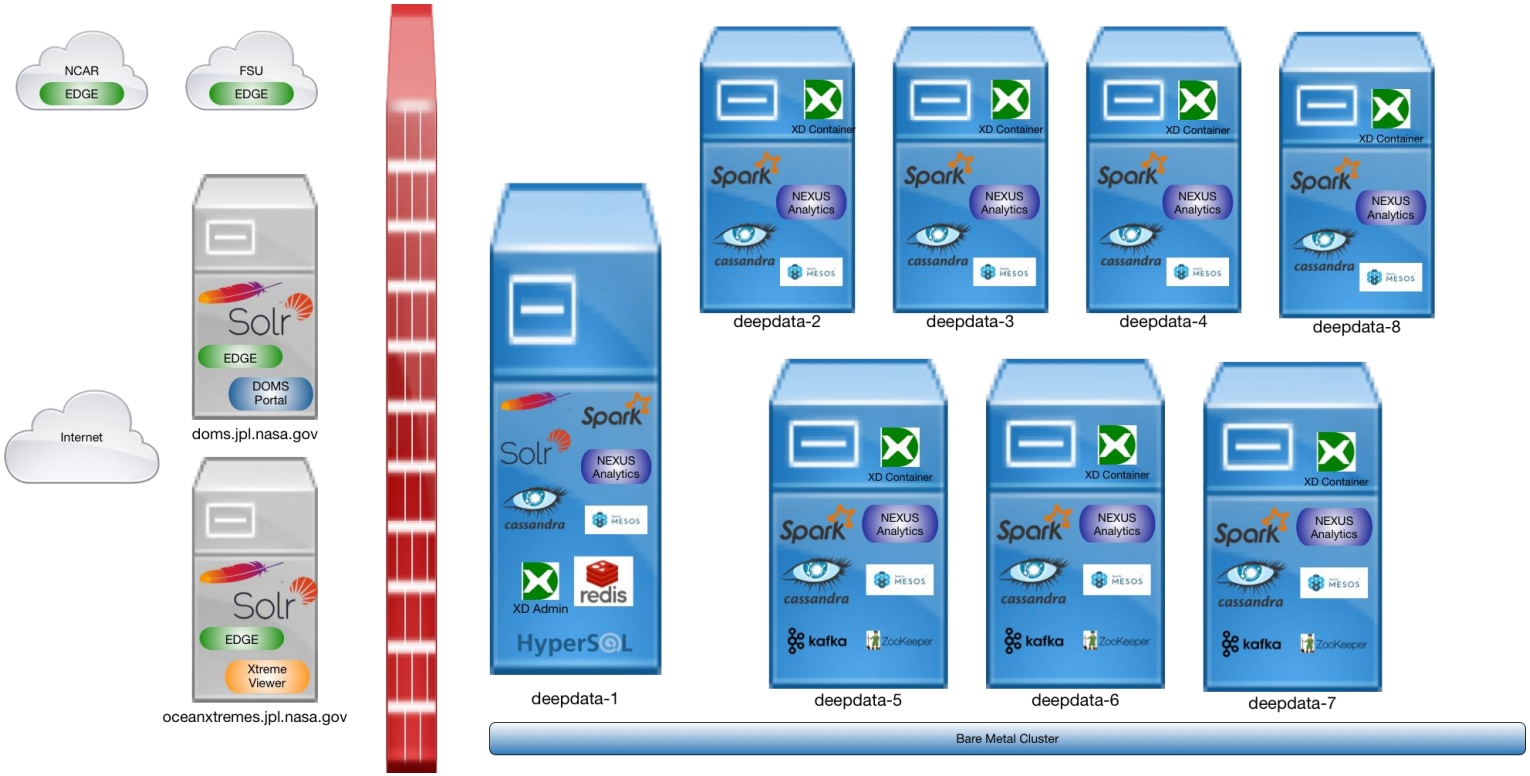
Local

- Running NEXUS on your laptop
  - Native with Vagrant
    - https://github.com/dataplumber/nexus#developer-installation
  - Docker
    - https://github.com/dataplumber/nexus/tree/master/docker
    - https://hub.docker.com/u/nexusjpl/dashboard/



Bare Metal

- Bare Metal NASA AIST-funded Deep Data Computing Environment (DDCE) at JPL



Bare Metal

Pros

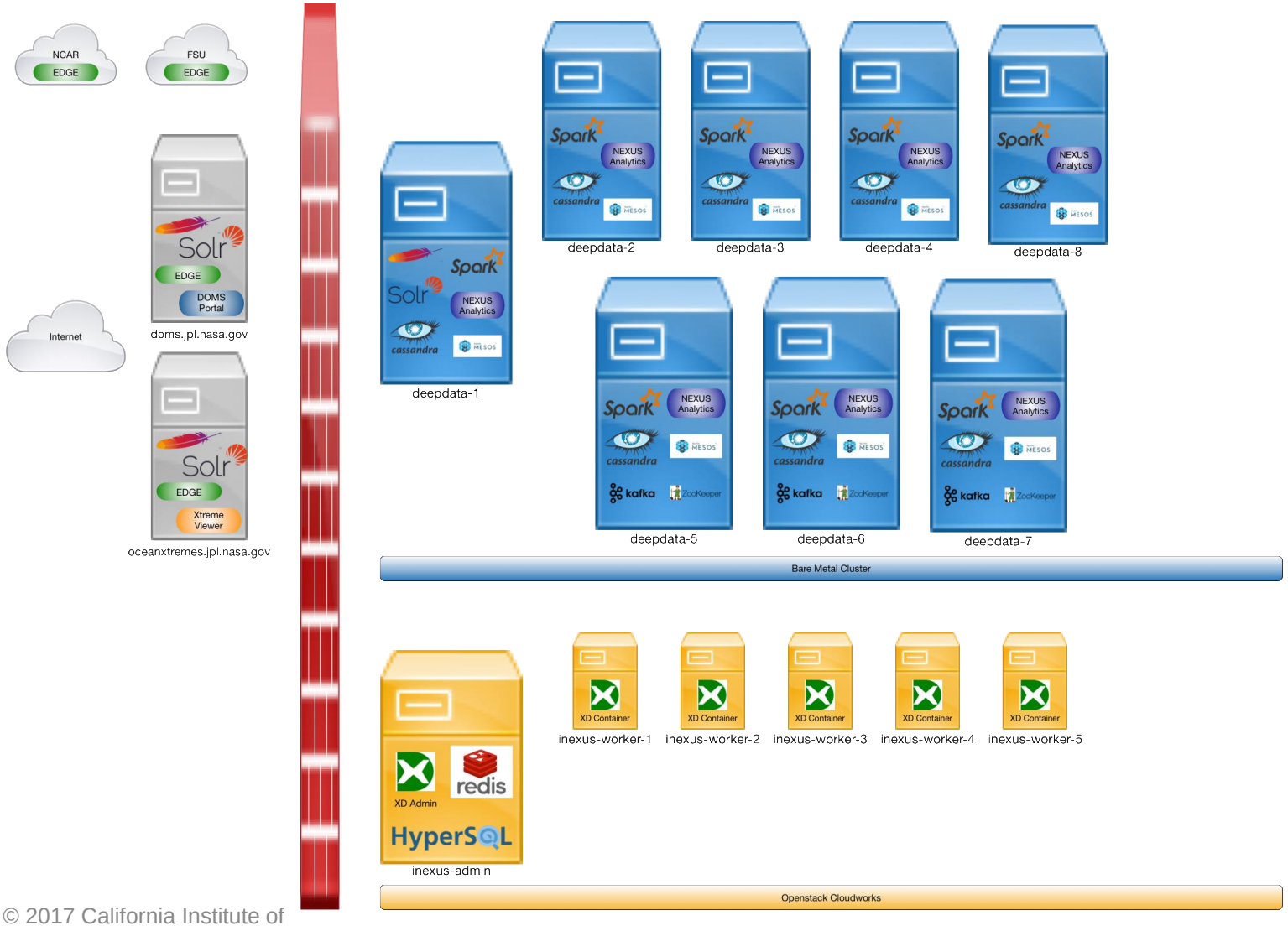
- Full control over operating system, software, and configuration
- No additional software overhead
  - Local disk access
  - Direct network adapter access
- Cost

Cons

- Management is difficult
  - Operating system patches
  - Custom startup scripts
  - Lots of SSH sessions
  - Adding new machines
- Clusters competing for resources
- Cost

On Premise Cloud

- Ingestion Cluster moved to OpenStack



On Premise Cloud

Pros

- Virtual - easier to add new machines
- Similar to bare metal installation

Cons

- Virtualization adds layer of abstraction – i.e. overhead
  - Kafka performance issues
- Similar to bare metal installation

Amazon Web Services (AWS) Elastic Compute Cloud (EC2)

- AIST Managed Cloud Environment (AMCE)
  - Fully Docker-ized Deployment

Amazon Web Services (AWS) Elastic Compute Cloud (EC2)

Pros

- Very easy installation
  - Write Dockerfile once, deploy anywhere Docker can run
  - Host machines only need to be able to run Docker
- Easy to add new containers
- Flexible deployment architecture
  - Choose the size that is right for you
- Cost

Cons

- Container Orchestration is hard
  - Docker Swarm and Docker Stack not ready for primetime
  - Don’t kill the swarm manager
  - Docker logs
- Debugging is harder
  - Especially when using Docker defined networking
- Additional overhead between code and infrastructure
  - In practice, not significant with our workload
- Cost

Other Deployments

- Amazon Web Services (AWS) Elastic Compute Cloud (EC2)
  - Used beefy machines
    - 6 x i2.4xlarge
      - Memory: 122 GB, vCPUs: 16, 4 * 800 GB SSD per instance
  - Compared Cassandra vs. ScyllaDB
  - Similar to bare metal installation

- Sea Level Change Portal
  - Bare metal installation at JPL
  - Small cluster due to nature of data
    - 1 Solr instance
    - 1 Cassandra node
    - No Spark/Mesos

- This Workshop!
  - Single EC2 instance per student (group)
  - ”Mimic” a full cluster deployment