

# BIG-Data @INFN-Bari

*Giacinto DONVITO*

*INFN-Bari*



# Outline

- ◆ IaaS @INFN-Bari
  - ◆ Architecture and characteristics
- ◆ Hadoop
- ◆ Sahara
- ◆ Conclusions



# IaaS @INFN-Bari

- ◆ The INFN-Bari is actively involved in two national Cloud project
- ◆ with the aim to support both Public Administration and scientific data analysis workflows on an Open Source IaaS solution



# IaaS @INFN-Bari:

## Software solutions

- ◆ Cloud IaaS Solution:
  - ◆ OpenStack (Icehouse at the moment)
    - ◆ KVM based virtualization
- ◆ Storage:
  - ◆ GlusterFS 3.4 (replica 2 and 3) both posix and iSCSI export
  - ◆ CEPH Firefly release (replica 3)
  - ◆ Swift: Supported both S3 and CDMI interface
- ◆ Operating System:
  - ◆ Ubuntu 12.04 LTS



# IaaS @INFN-Bari:

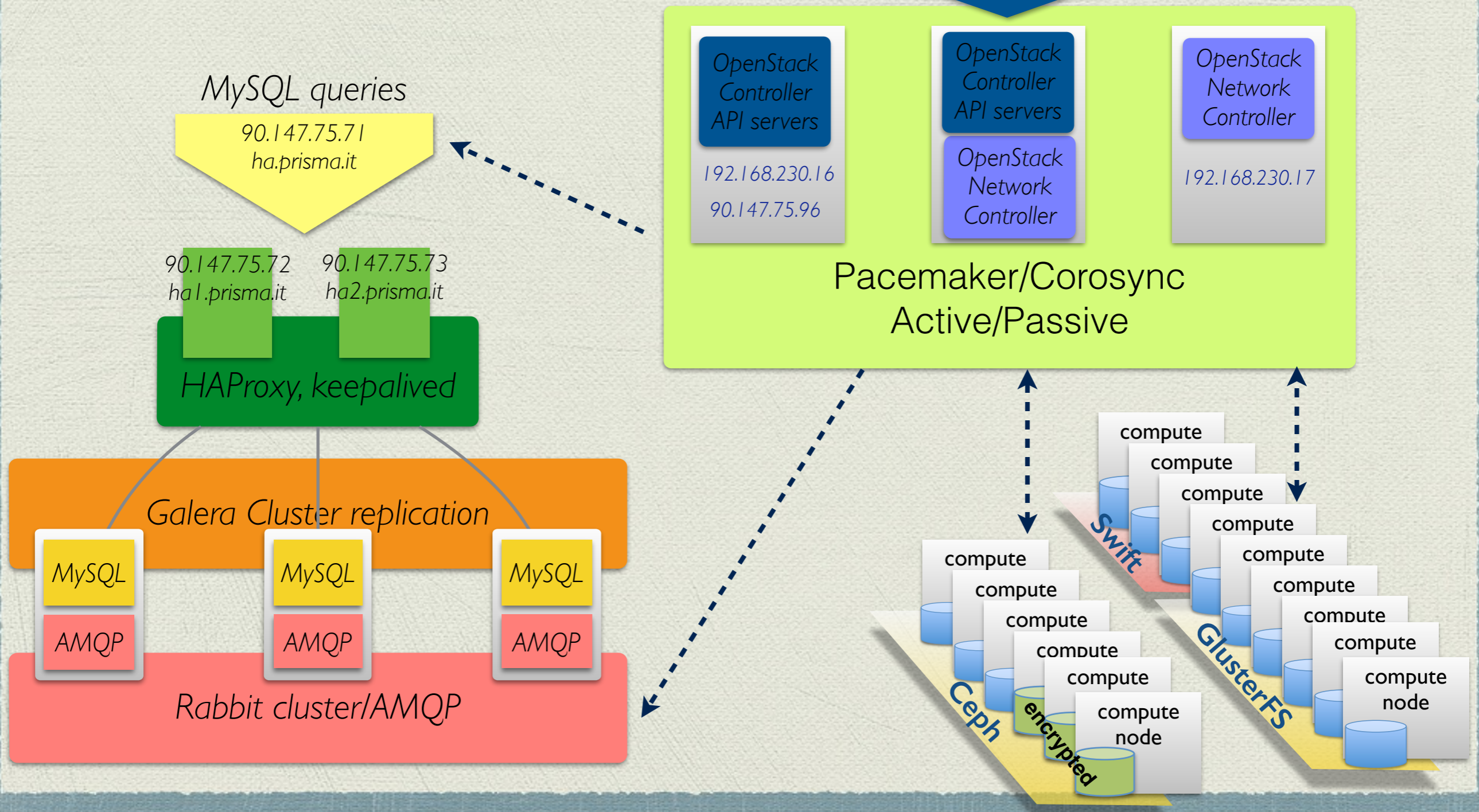
## Hardware solutions

- ◆ Base services:
  - ◆ MySQL + RabbitMQ
  - ◆ 3 hosts with 8 Cores and 18GB of RAM each
- ◆ Core services:
  - ◆ 3 hosts with 24 Cores and 80GB of RAM each
- ◆ Compute Node:
  - ◆ 12 nodes with 32 Cores and 256 GB of RAM each + 15 nodes with 24 Cores and 80GB of RAM each
  - ◆ **About 700 cores and 4TB of RAM**
- ◆ Network:
  - ◆ Each physical hosts has 1x10Gbit/s wire-speed guaranteed bandwidth
- ◆ Storage:
  - ◆ **~150 disks, for a total of ~470TB of overall storage**



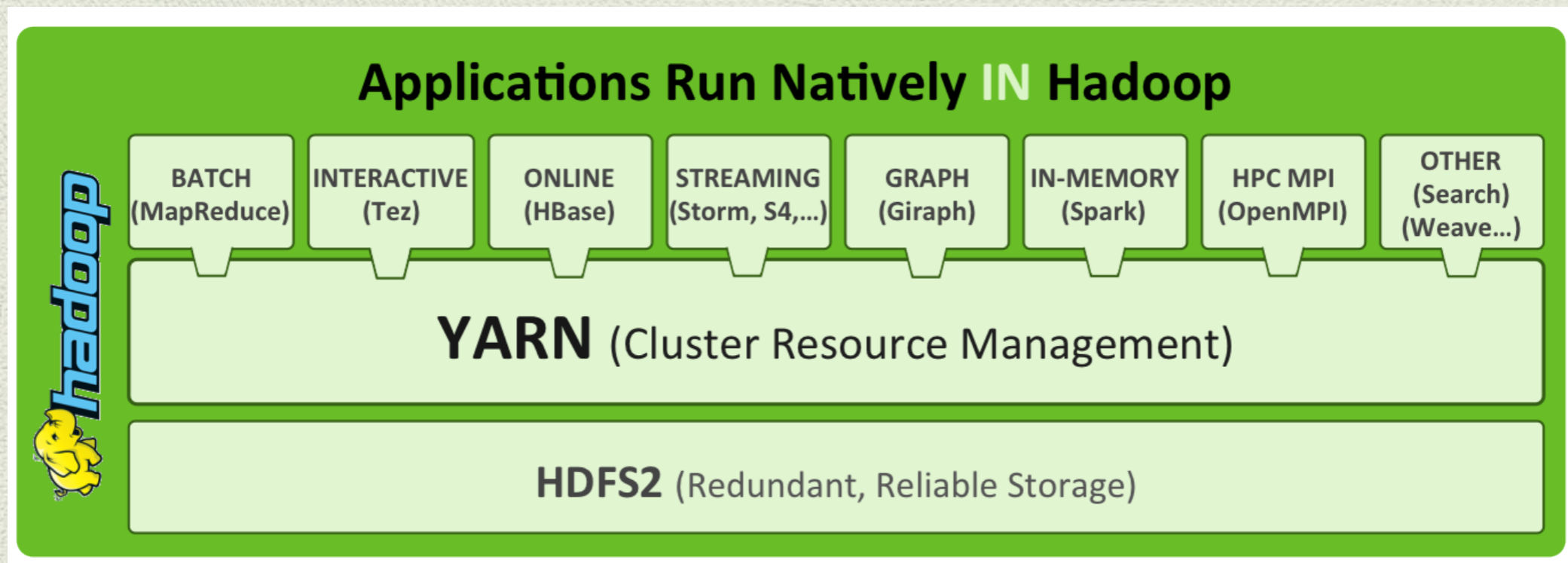
# IaaS @INFN-Bari: layout

## OpenStack Public APIs



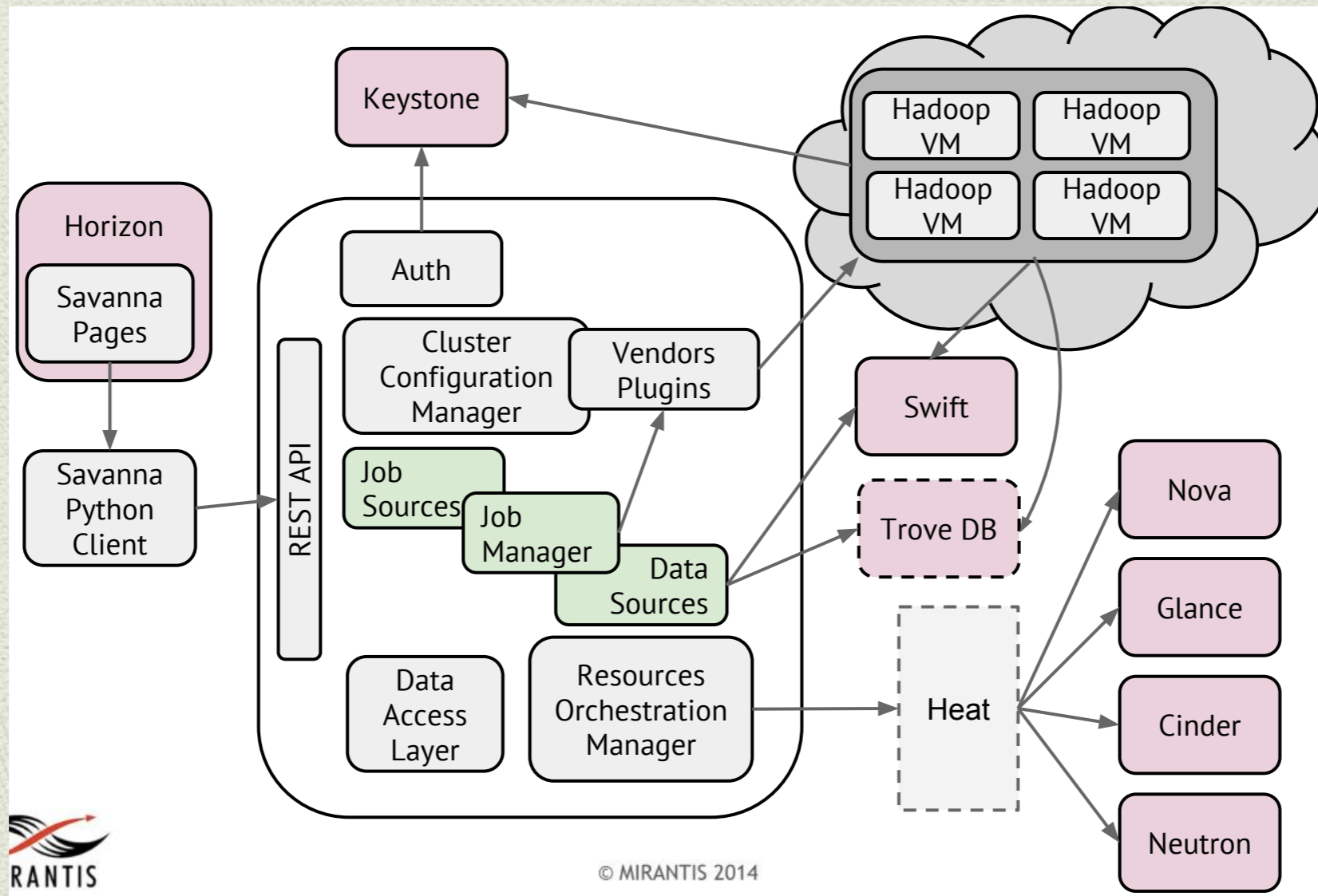


# Overview on Hadoop



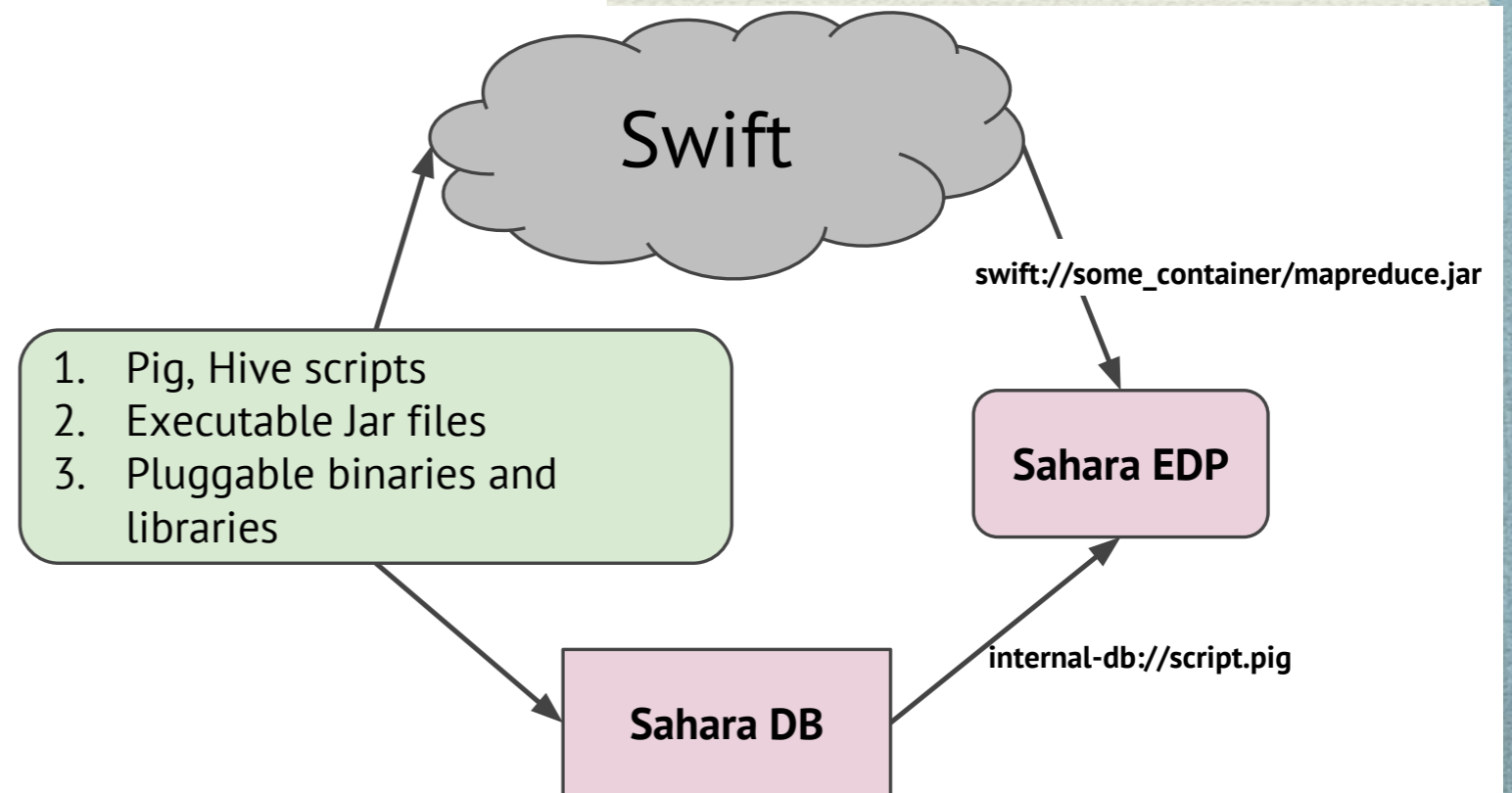
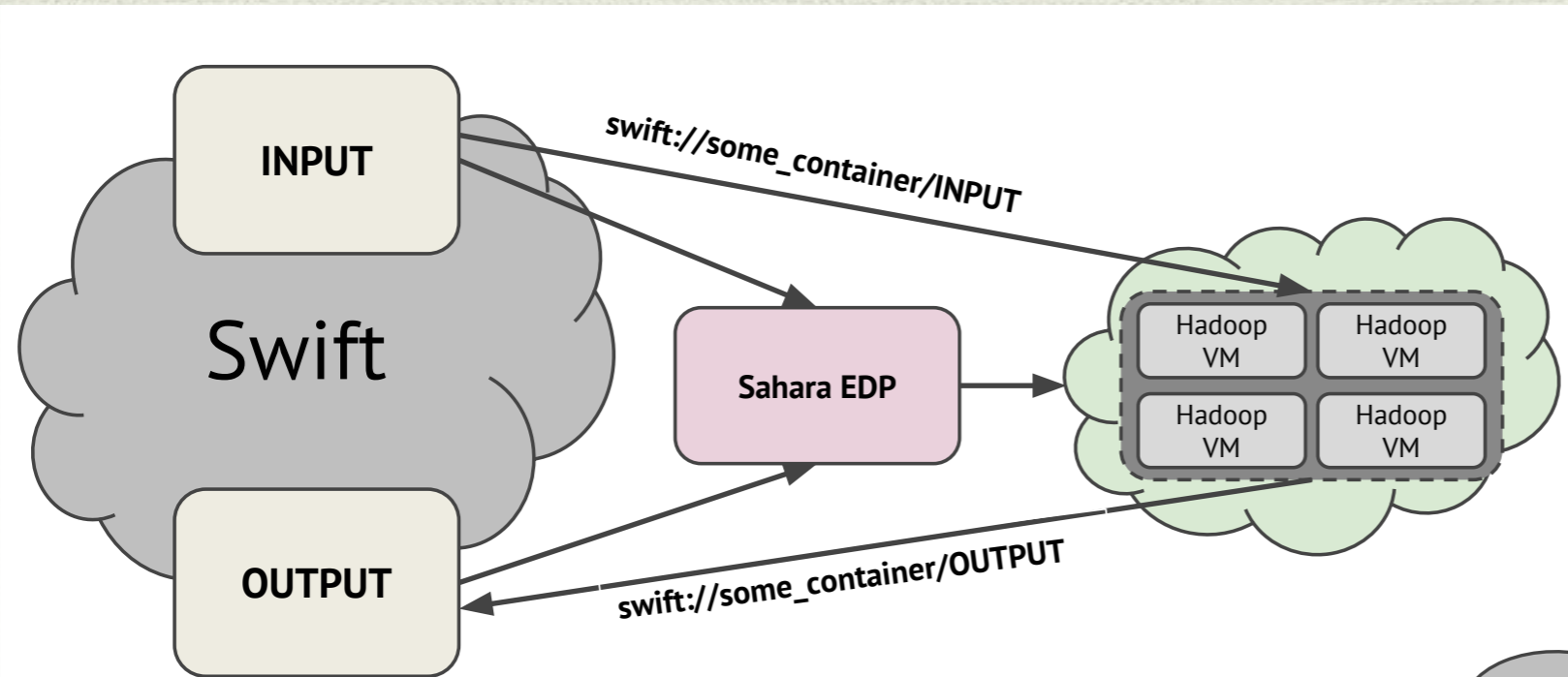


# Hadoop on OpenStack: Sahara





# Hadoop on OpenStack: Sahara





# Future works

- ◆ Test Sahara on OpenStack
  - ◆ Both in terms of data storage and executing code
  - ◆ Measures scalability in terms of performances on our hw IaaS
- ◆ Test the installation of NoSQL DB on dynamic cluster instantiated with Heat
  - ◆ at least MongoDB and Cassandra
    - ◆ as we already has experiences on these two



# Conclusions

- ◆ Our hw Cloud infrastructure aims to provide horizontal scalability
- ◆ we expect it to be well-suited for the Big Data use-cases
- ◆ CEPH will provide Block storage via RBD
  - ◆ that should be scalable and performant