DIRAC and Clouds

DIRAC User Workshop 29-31 Oct 2012. CPPM. Marseille

Víctor Méndez Muñoz
Port d'Informació Cientifica (PIC)

1. Introduction

2. VMDIRAC status

2.1 Federated cloud tested Architecture

2.2 Work in progress branches

3 Integration of the Federated Cloud model with DIRAC

3.1 UseCase LHCb Monte Carlo in EGEE FedCloud

3.2 VMDIRAC integration of Federated Cloud Overview

4 VMDIRAC job submission

1. INTRODUCTION

DIRAC is a proven Grid solution, which currently is providing transparent access and interoperability between resources as:

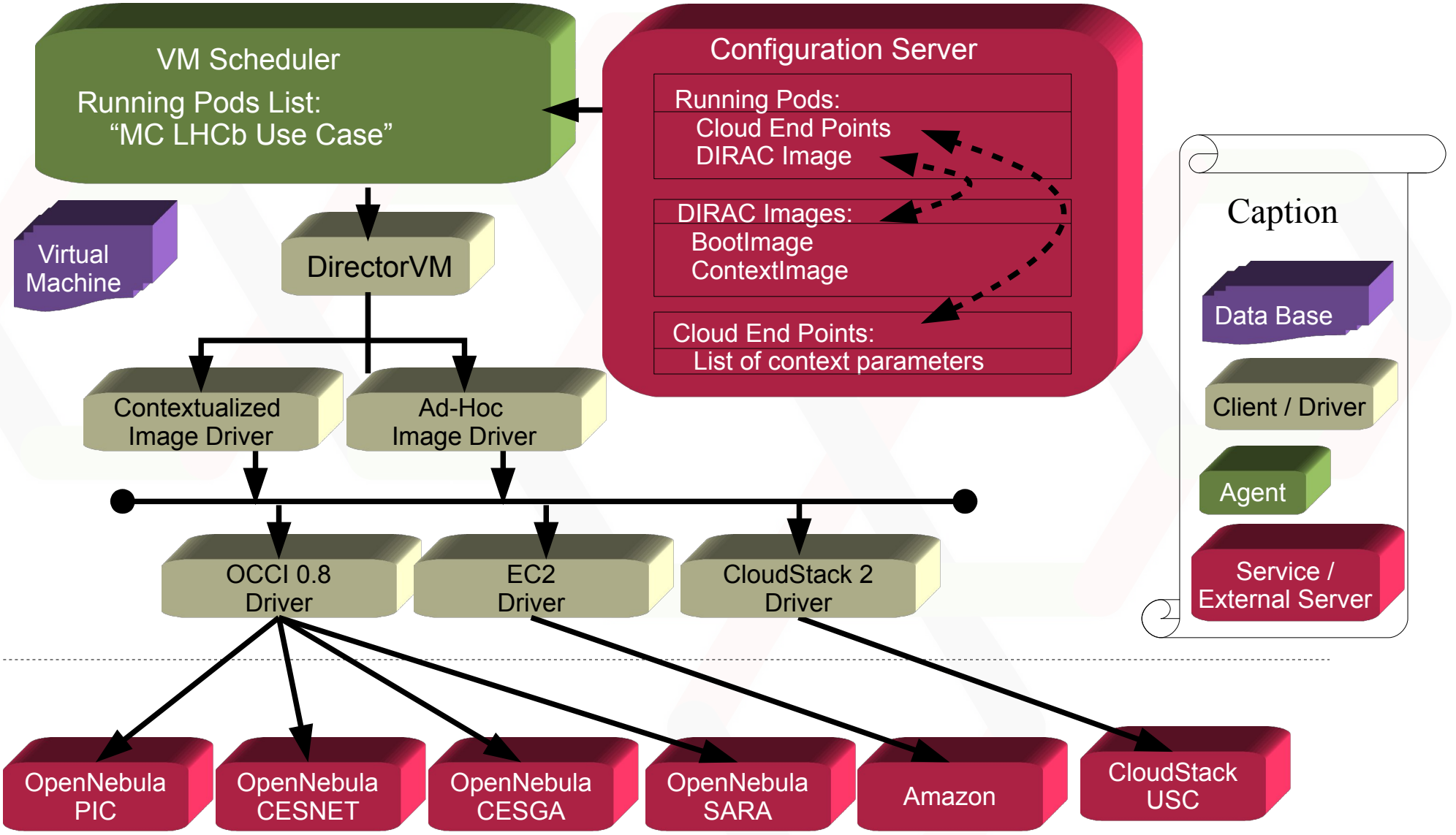
- Grid resources (EGI, OSG, batch system)
- Cloud: Amazon EC2, OCCI 0.8, CloudStack 2

VMDIRAC new release: A federated cloud model integration

- Multi cloud endpoint management
- OCCI 1.1
- CloudStack 3
- Federated authentication
- SSL
- Contextualization as the image maintenance approach (CernVM)
- Software repositories as part of the contextualization system (cvmfs)

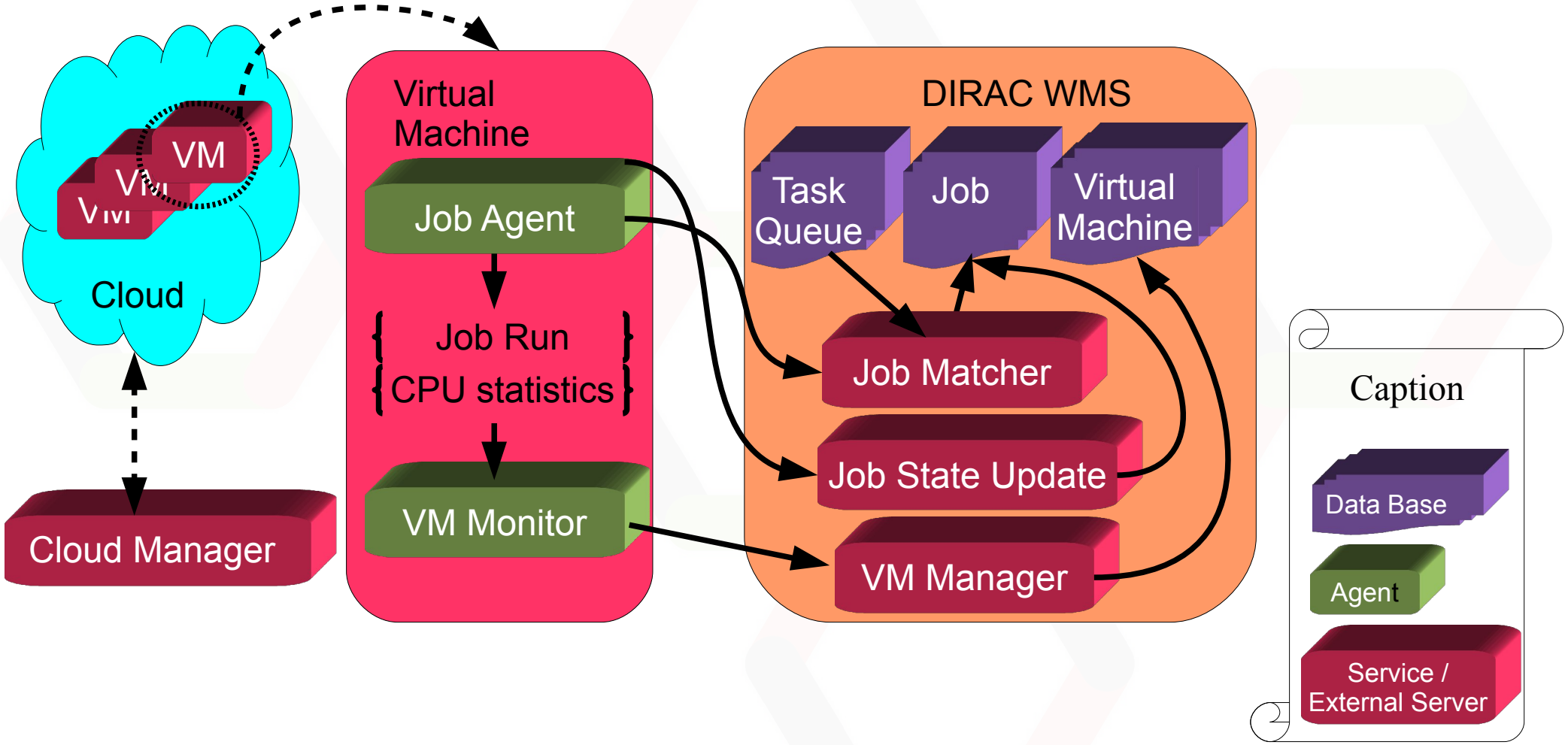
2. VMDIRAC status:
2.1 Federated cloud tested Architecture

- Multi endpoint management design -



2. VMDIRAC status:
2.1 Federated cloud tested Architecture

- Job Running in a VM -



2. VMDIRAC status:
2.2 Work in progress branches

Stable release VMDIRAC v0r4
https://github.com/DIRACGrid/VMDIRAC.git

Federated cloud tested architecture (vmendez/VMDIRAC.git v0r5)
- CloudStack Multi-endpoint
- OCCI 0.8 Multi-endpoint, Contextualization

Federated cloud developing to be tested
- OCCI 1.1 (rOCCI OpenNebula and OpenStack)
- EC2 Contextualization
- Authentication X509
- SSL connections

Next of VMDIRAC on the federated cloud
- OCCI 1.1 (rOCCI CloudStack)
- VOMS proxy authentication
- Marketplace polling integration (StratusLab as the federated marketplace image metadata)

2. VMDIRAC status:
2.2 Work in progress branches

Current Cloud APIs standardization main issues

- OpenStack contextualization: amiconfig + HEPiX scprits + EC2
  - There is no implementation in OCCI
  - Alternatively one can use OCCI mixin:
    - Golden Image + Context image (Without context section for dynamic)
    - Dynamic context variables via Online Portal

- CernVM 2.6.1 (development release) is including EC2 contextualization
  - usual amiconfig plugins for CernVM contextualization
  - hepix plugin for amiconfig
  - Dynamic context via Online Portal

- CloudStack OCCI 1.1 is going to be developed using rOCCI
- VOMS proxy authentication prototype deployed at FedCloud
- StratusLab marketplace (only image catalog) deployed at FedCloud

3 Integration of the Federated Cloud model with DIRAC

EGEE FedCloud Task Force

- Currently in testbed status
- To be in production in April 2013
- Aggregation of many private cloud providers with different cloud managers (currently: OpenNebula, OpenStack, StratusLab, WNoDeS, Okeanos...)
- Aggregation of multiple Web services for eScience purposes (Marketplace, Information System, Authorization....)

EGEE FedCloud Use Case:

- LHCb Monte Carlo in EGEE FedCloud
  - OCCI 1.1 and EC2 Multi Cloud Manager
  - Multi endpoint submission
  - Federated authentication

https://wiki.egi.eu/wiki/FedCloudDIRAC

3.1 UseCase LHCb
Monte Carlo in EGEE FedCloud

- Golden Image: Cernvm batch node
- Context image for OCCI 0.8, OCCI 1.1 and EC2
- Cvmfs with the lhcb.cern.ch repository

- Functional Test: multi endpoint cloud management (OCCI 0.8)
  - PIC, CESNET, CESGA, SARA

Lessons learned:
- DHCP VM network interface setup
- Outbound communication automatic setup is necessary
- Debugging contextualization policies
  - Mandatory context parameters: occiURI, maxEndpointInstances, instanceType, driver (OCCI 0.8), occiUser, occiPasswd, netId
  - Optional parameter: networking parameters, CVMFS_HTTP_PROXY, networkFilter (libvirt), contextFiles

3.1 UseCase LHCb
Monte Carlo in EGEE FedCloud

- Scaling test: Monte Carlo LHCb workflow
  - LbLogin -> SetupProject -> Gaudi Gauss
  - 1000 DIRAC jobs with 100 events each
  - 954 jobs "done" (without rescheduling)

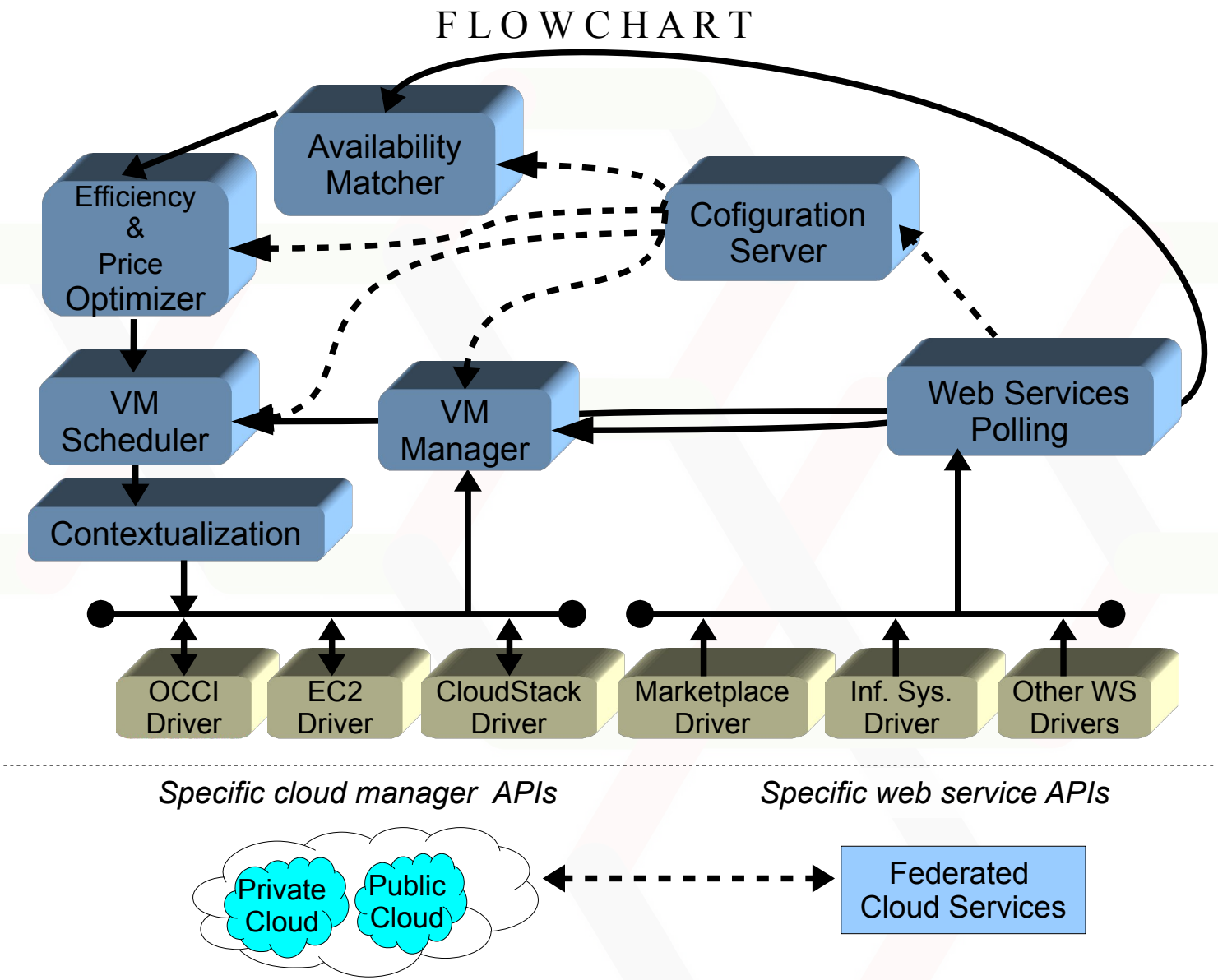
Lessons learned:

- Different provider behaviour
  - Network, VM types, mounting devices
- Issues observed in the scaling test
  - Large startup latency, up to many hours
  - IO issues

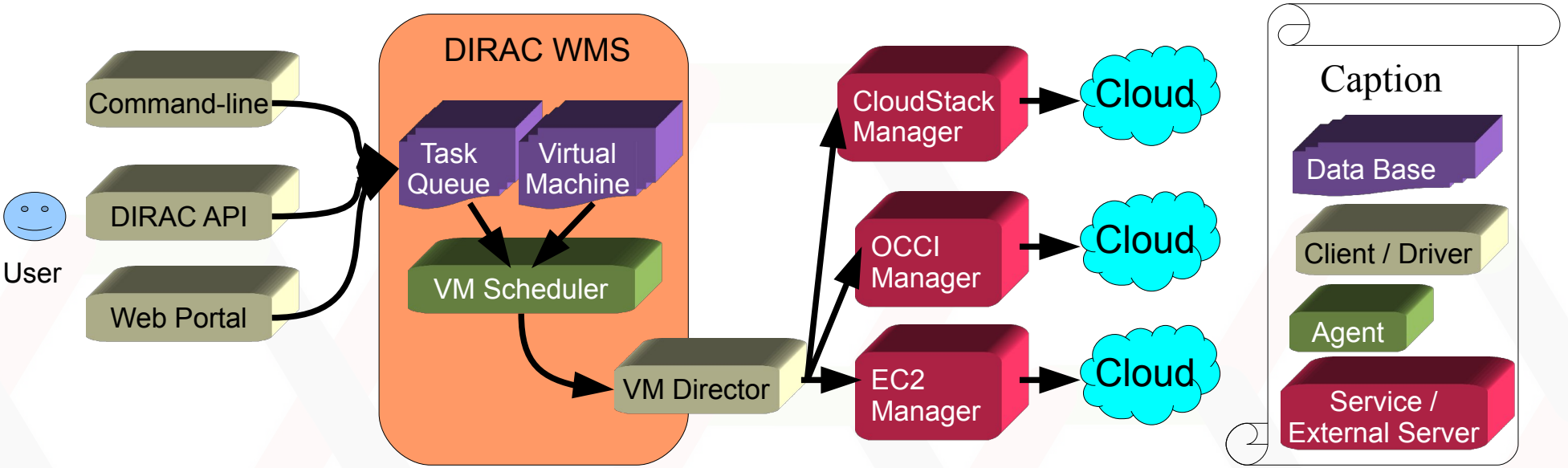
- Further work in progress on the Cloud endpoints to reach a production level. The current LHCb Monte Carlo Use Case is providing feedback on the scaling conditions and performances

10 DIRAC User Workshop, Marseille 2012 DIRAC and Clouds

3.2 VMDIRAC integration of Federated Cloud Overview



4 VMDIRAC submission



JDL job: SubmitPools = “Grid”, "ContextualizedCloud", “MyImageCloud”;
Where the last two pools are cloud pools, would be defined in CS:
Systems / WorkloadManagement / Production / Agents
VirtualMachineScheduler
SubmitPools = ContextualizedCloud, MyImageCloud
ContextualizedCloud
RunningPods = ONECernVMContext, EC2CernContext
MyImageCloud
RunningPods = CentosDIRAC_IN2P3, UbuntuDIRAC_USC