



21st International Conference on Computing in High Energy and Nuclear Physics **CHEP2015** Okinawa Japan: April 13 - 17, 2015



Pilots 2.0: DIRAC pilots for all the skies

*Federico Stagni,
A. McNab, C. Luzzi, A. Tsaregorodtsev*

*On behalf of the DIRAC consortium
and the LHCb collaboration*

History class

(with possibly an LHCb bias)

We didn't have a big variety of resources:

- WLCG sites - AKA the "Grid"
 - EDG, EGEE, EGEE-II, EGEE-III, EGI, EMI, EMI2, EMI3...

Submitting Jobs (from a central queue) to:

- The LCG resource broker (aka WMS): a queue for dispatching to
- CEs (LCG, then CREAM): a queue for dispatching to
- Batch queues in front of the WNs
- Finally, running on a WN

→ Number of queues: 4

- LCG inefficiencies exposed to end users
- High load on LCG brokers

→ So, pilot jobs came

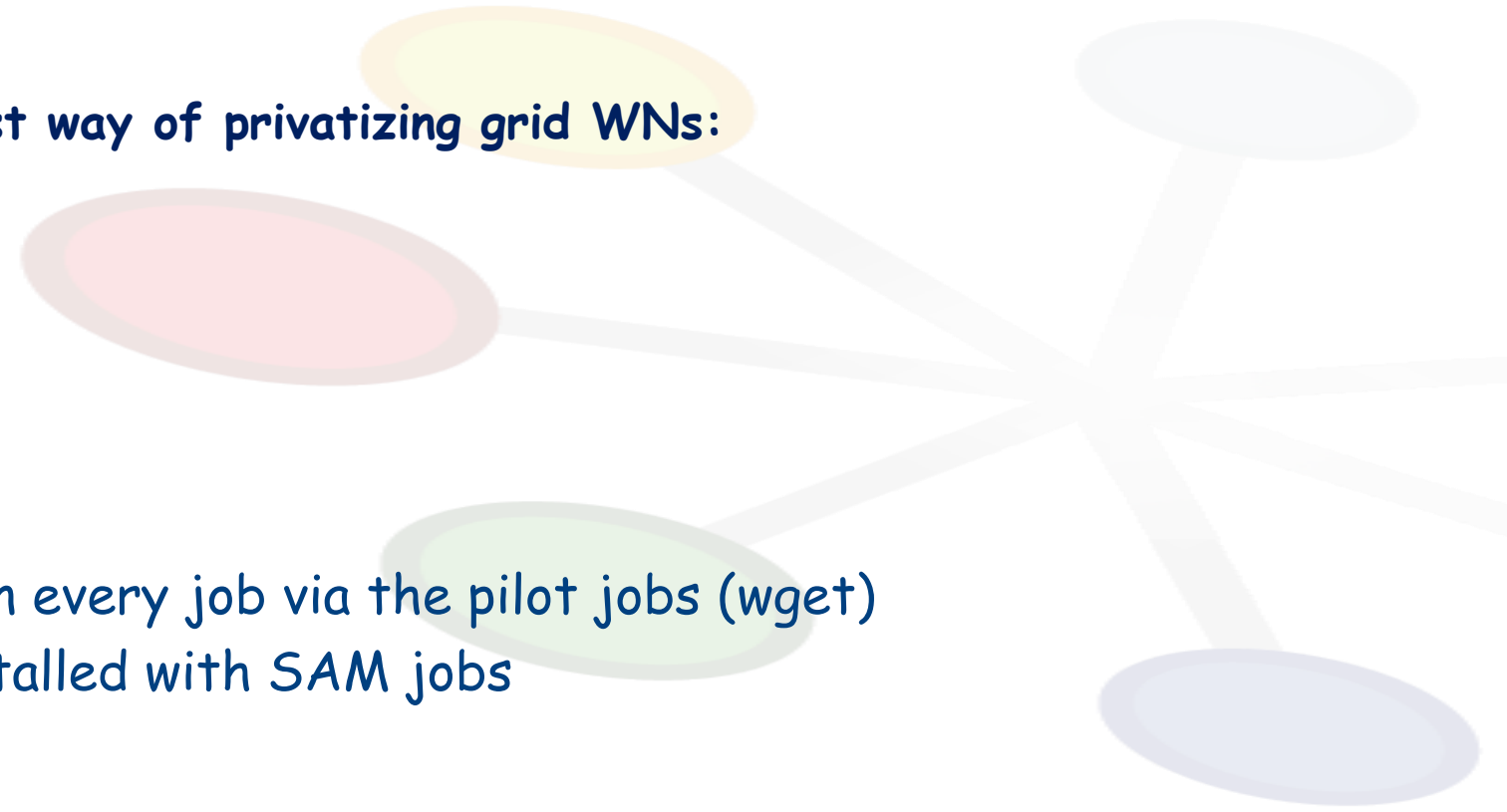
Pilot jobs came also because **VOs** wanted their "own" machines for their Grid jobs

- Or, at least, to privatize them for some hours

Pilot jobs became the first way of privatizing grid WNs:

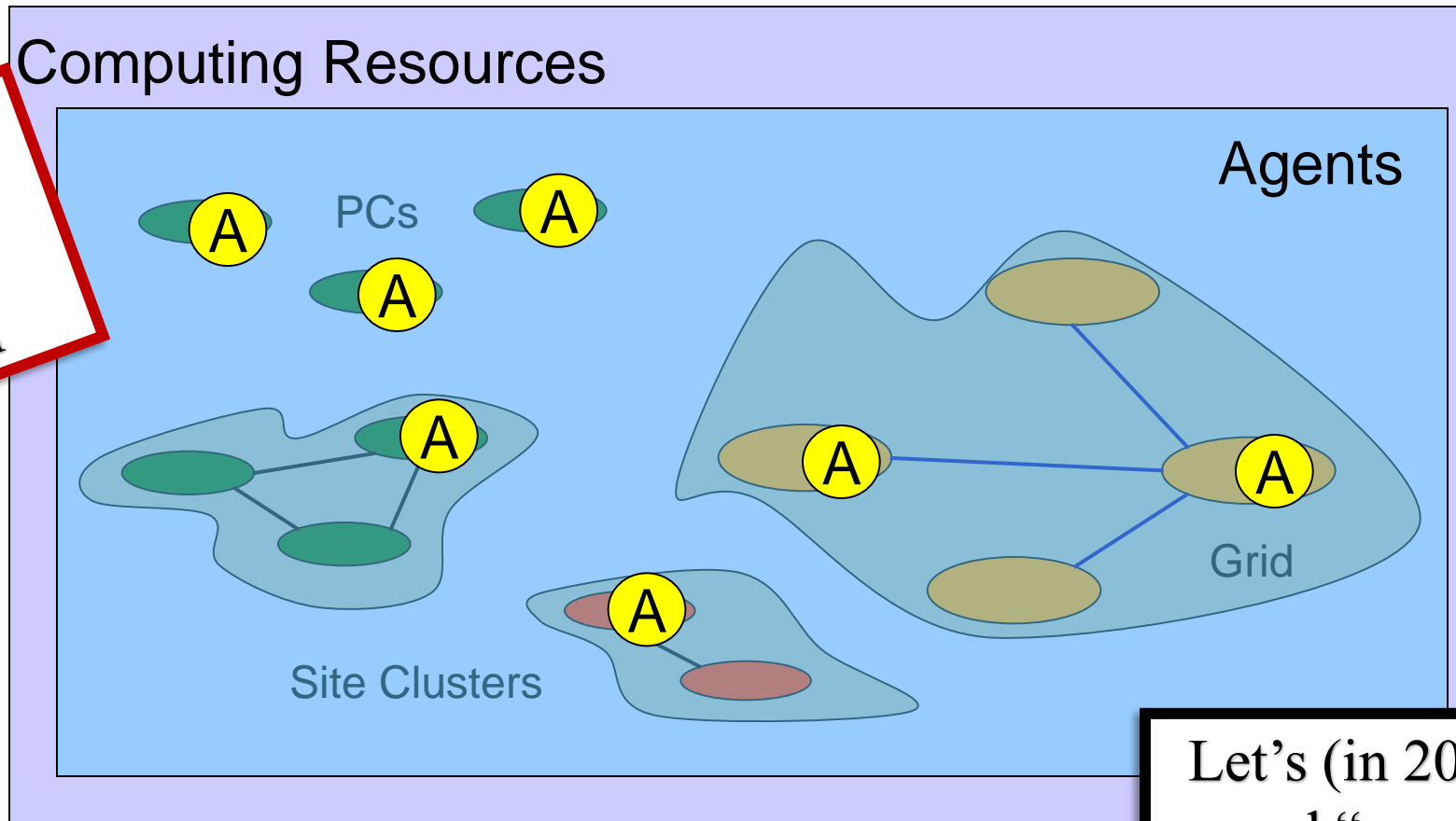
- set the environment
- install the middleware

- (LHCb)DIRAC: installed in every job via the pilot jobs (wget)
- Application software: installed with SAM jobs



Overlay network paradigm

Agents form an overlay layer hiding the underlying diversity



A. Tsaregorodtsev,
CHEP 2006,
DIRAC
presentation

Let's (in 2015) substitute the word "agent" with the word "pilot"

A DIRAC pilot has, at a minimum, to:

install DIRAC

configure DIRAC

run and agent: the "JobAgent"

- That fetches (matches, in fact) a job from the central jobs queue
 - ↳ Or, more than one job only

In DIRAC, a pilot has to run on each and every computing resource type.

Agents Pilots form an overlay layer hiding the underlying diversity

- ...or not?

→ Well, not completely... until Pilots 2.0

Many LHC and non-LHC communities started having quite some variety of resources

- WLCG sites
 - CREAM CE: direct pilot jobs submission
 - ↳ and no more central brokers are needed → one less queue ☺
 - But also other CEs, e.g. ARC CEs are a popular choice among sites
- WNs are VMs that the experiment provides on VAC, Cloud
 - VAC: an IAAC
 - Cloud: an IAAS
 - Condor-based systems
- Various forms of opportunistic computing
 - HLT farms
 - ↳ Some experiments made a cloud, other (i.e. LHCb not)
 - (HPC) opportunistic sites
 - ↳ Usually, not used as real HPC, anyway...
 - BOINC
 - ↳ both IAAS and IAAC

Managing VMs with VAC and Vcycle
A. McNab, Track 7, Mon, 17:00

Track 7, Tue, 14:00 – 16:00

It seems like the grid is not anymore “The Grid”

Heterogeneity is the norm



Exercise: monitoring of pilots as example of heterogeneity

- Exercise: get the Logging Info of the pilot
 - LCG: `edg-job-logging-info -v 2 --noint <ref>`
 - gLite: `glite-wms-job-logging-info -v 3 --noint <ref>`
 - CREAM: `glite-ce-job-status -L 2 <ref>`
 - ARC: a feature request
 - DIRAC CEs: ...nope
 - CLOUDs: depends from cloud to cloud and from what you expect
 - BOINC: NA
 - HLT: ...not?
 - Opportunistic : ...
 - ... resources of tomorrow: ??

So, we can keep adding support for each and every type of resources, or

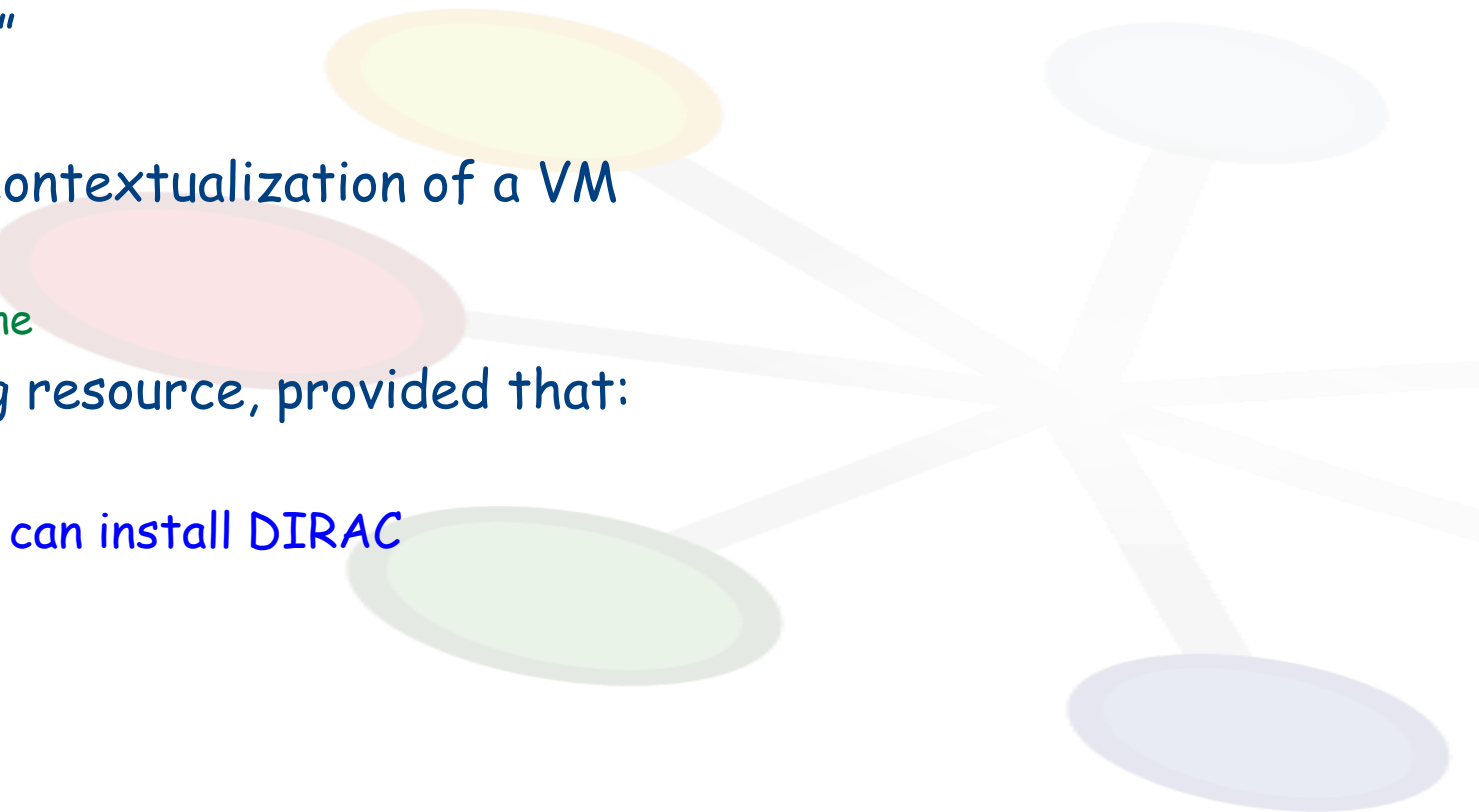
→ we can just embed this functionality in the pilot

- 2nd exercise: get the log output file
 - Same story as above!

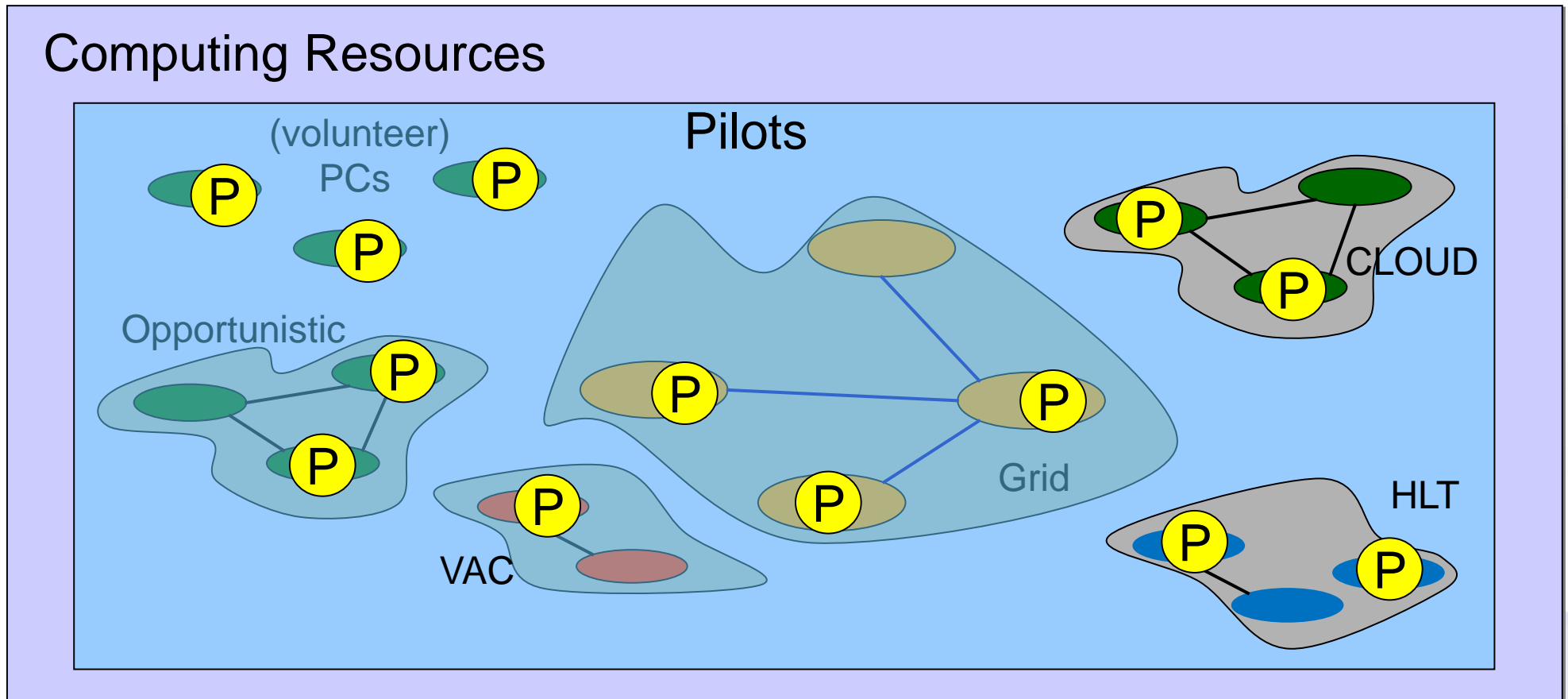
Pilots 2.0 as overlay layer

A pilot is what creates the possibility to run jobs on a worker node.

- A pilot 2.0 is a standalone script
- Can be sent, as a "pilot job"
 - To all "Grid" CEs
- Can be run as part of the contextualization of a VM
 - Or on an HPC machine
 - ↳ Or on a ...whatever machine
- Can run on every computing resource, provided that:
 - Python 2.6+ on the WN
 - It is an OS onto which we can install DIRAC



- The same pilot used everywhere



A toolbox of pilots capabilities (that we will call "commands") is available to the pilot script

Each command implements a single, atomic, functions, e.g.:

- Run an environment test
 - Install DIRAC (or its extension)
 - Configure DIRAC
 - Run the JobAgent
 - Run monitoring thread
 - Report usage
 - ... and whatever it is needed
-
- Communities can easily extend the content of the toolbox, adding more commands
 - If necessary, different computing resource types can run different commands
 - All configurable

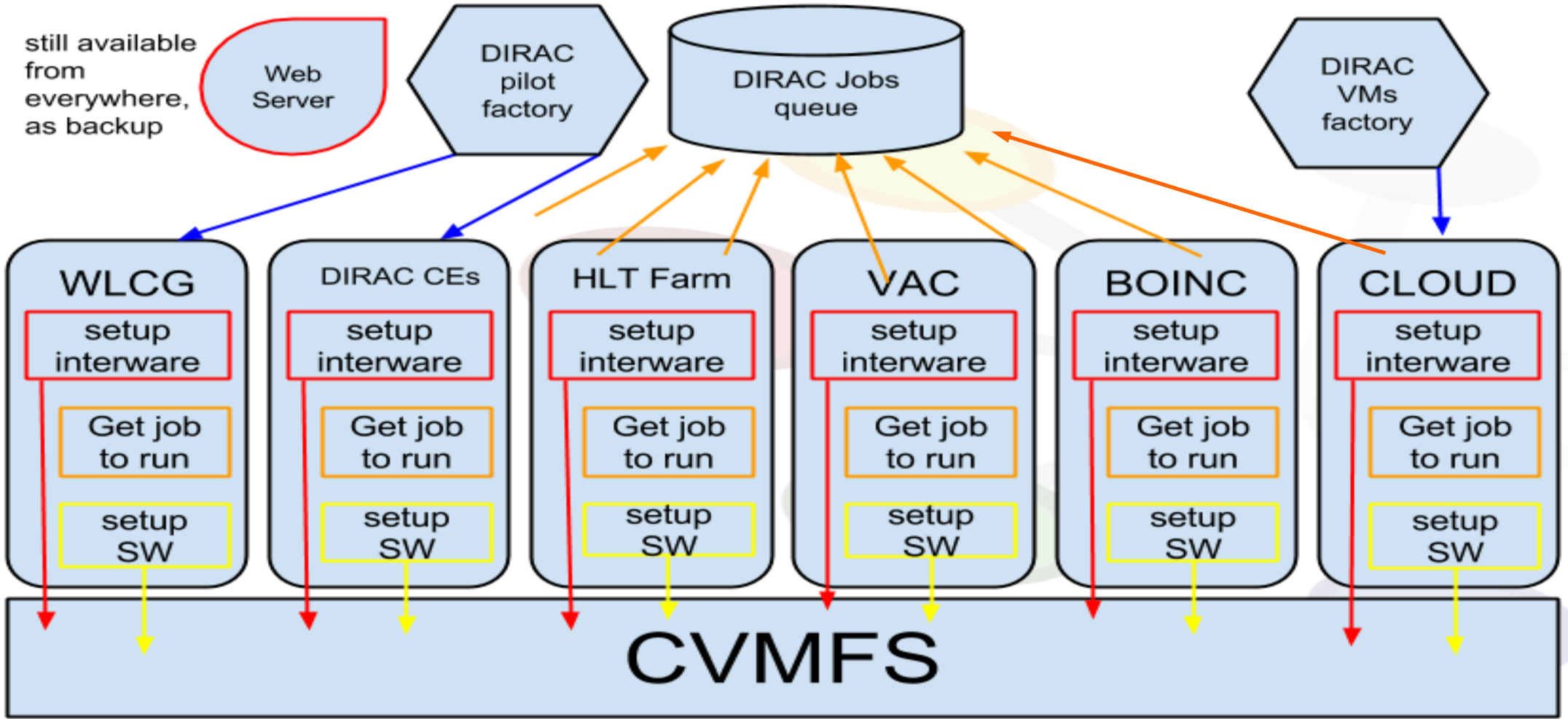
LHCb pilots 2.0



- LHCb requirement: CVMFS
 - For the distribution of all LHCb software, including LHCbDIRAC
 - LHCb Pilots will try to “install” (set up) LHCbDIRAC from CVMFS
 - ↳ If it fails (e.g., it's a just deployed version, and the cache is cold – or a test version), will fall back installing the old way
 - LHCb won't download CAs: instead it uses what is on CVMFS
 - ...

LHCb experience with running jobs in VMs
A. McNab, Track 7, Tue, 17:30

“Pilots to fly in all the skies”



- Pilots 2.0 are the “pilots to fly in all the skies”
- Available to all DIRAC communities as of DIRAC v6r12
 - ↳ Some may have not noticed the change...
- Easy to extend
 - Actively developed
- Pilots 2.0 are the real federators
 - WNs look the same everywhere
 - ↳ Make a pilot to run, and you will monitor all of them the same way
- Extended by LHCb
 - Mainly for fully using CVMFS

Prospects:

- Add more generic commands
- One will be for running SAM jobs

Question, comments

