Pilots 2.0

Pilots to fly in all the skies

A new generation of pilots

Why?

- Pilots 1.0 have been developed 10 years ago
  - Presented at CHEP 2006 in Mumbai...
- With the idea of pilot "jobs" only
- One long, out-dated python script
  - Lots of "old garbage" in
  - So anyway, worth a revision

- In the meantime, the grid is not anymore "The Grid"
  - HLT, IaaS, IaaC, vac...

Requirements

- Same same but different - and better
- Configurable
- Easy to test and extend
- Run everywhere
  - Pilot jobs
    - CREAM CEs
    - DIRAC CEs
    - ARC CEs
  - CLOUD
  - VAC
  - BOINC
  - HLT
  - …

Since when

- One of the main "products" of DIRAC v6r12
  - Released 6 months ago

- Multiple efforts:
  - Cinzia, Andrew, Andrei, Federico
- Long, not easy development

How

- A standalone set of scripts
  - NOT dependent from DIRAC...
- Can be sent, as a "pilot job"
- Can be run as part of (the contextualization) of a (V)M
- Requirements:
  - Python 2.6+ on the WN
  - It is an OS onto which we can install DIRAC
- Make this script to run (with the proper switches) and it will probably get you a job

Coding rules

Discussed within this [RFC](#)

> But implementation is a bit different

Basically, **an implementation of the command pattern**

A **toolbox of pilots capabilities** (the "commands") is available to the pilot script. Each command implements one function:

- Check the environment
- Get the pilot version to install
- Install (VO)DIRAC
- Configure (VO)DIRAC
  - In fact, there are several configuration commands
- Configure CPU capabilities
  - the famous "dirac-wms-cpu-normalization"
- Run the JobAgent

Just a script

```
python dirac-pilot.py
    -S $DIRACSETUP
    -r $projectVersion
    -C $CSURL
    -N jenkins.cern.ch
    -Q jenkins-queue_not_important
    -n DIRAC.Jenkins.ch
    -M 1
    --cert --certLocation=/home/dirac/certs/
    -X GetPilotVersion,CheckWorkerNode,InstallDIRAC,ConfigureBasics,
ConfigureSite,ConfigureArchitecture,ConfigureCPURequirements,LaunchAgent
```

DIRAC
THE INTERWARE

Capabilities

- Communities can easily extend the content of the toolbox, adding more commands
- If necessary, different computing resources types can run different commands
- Highly configurable, parameters for "everything"

Examples of extensions

- ## LHCb
  - requirement for WNs: 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 (dirac-install)
  - LHCb won't download CAs: instead it uses what is on CVMFS
- ## GridPP
  - GitInstallDIRAC command
- ## Others…?

DIRAC

THE INTERWARE

The good

- Easy to add/extend/modify commands
- Easy to test
  - unit-test for commands
  - integration tests are also possible
- Easy to deploy
  - Pilots 2.0 are the real federators
- Better logs
- Removed some "old garbage"

Still unsolved

- Monitoring:
  - No common logging
    - We have "WLCG" logging
      - In fact only CREAM CESs
        - And that's it
  - Pilot logs are not uploaded automatically
    - We can retrieve "WLCG" logs
      - ...and that's it
        - Yes, we can get logs from vcycle, but...

- Missing stats, e.g.:
  - SetupProject (its successor) OK?
  - how long it takes?

Prospects

Pilots logging

Main User story:
- AS Administrator I WANT TO have access to the pilot logging info for all the pilots, independently of where they run

Which translates into having a better, more generic system. So, a better federator

What's Pilots logging

A list of messages like

> "I've booted up" …
>
> "I found the DIRAC pilot ok" ...
>
> "I'm about to shutdown"...

and also things like

> "I installed DIRAC via SetupProject/dirac-install"...
>
> "This machine has power of 11 HS06"....
>
> "This machine is SLC6/CC7"...
>
> "I matched a job" or
>
> "I failed to match a job"... and so on.

Implementation glimpse

- Use RabbitMQ
  - Security issues being addressed using OpenSSL
    - Server deployment in the pipeline
  - pika and stomp libraries being experimented client side
- Standard DISET service+MySQL behind for storing the messages
- First PRs for v6r14 ready

Work done by

- Wojciech Krzemien
- Milosz Zdybal
- Federico Stagni
- Zoltan Mathe

Other developments

- Run monitoring thread? Maybe useful for VMs
- Report usage

  ... and whatever it is needed

- Other suggestions?

Summary

Pilots 2.0 are the "pilots to fly in all the skies", available to all DIRAC communities as of DIRAC v6r12

- Easy to extend
- Actively developed
- Pilots 2.0 are the real federators
  - WNs look the same everywhere
  - Some communities may have not noticed the change...
- Extended by LHCb
  - Mainly for fully using CVMFS
- Work in the pipeline for v6r14

?