



DIRAC Pilot 3.0 + VMs: Getting Started

Andrew McNab

LHCb, GridPP,
University of Manchester



Steps to get VMs running with Pilot 3.0

- Get a host certificate for VMs
- Configure your DIRAC CS
- Create your pilot web directory
- Copy files into it
- Edit `user_data_vm` if starting VM by hand
- Create `pilot.json`
- Start VM (`Vac`, `Vcycle`, by hand?)

Host certificate

- Currently a host certificate is used by the pilots in VMs to authenticate to DIRAC
 - In the future, this will be a proxy of a hostcert/key
- The “host” does not need to exist
- The hostname does not need to be in DNS
- However, it should be a domain name you are entitled to get a certificate for
- In almost all cases use a host certificate from an IGTF CA (the usual grid ones)
 - For example: /C=UK/O=eScience/OU=Manchester/L=HEP/CN=lhcb-vm.tier2.hep.manchester.ac.uk

Configure your DIRAC CS: host

- Allow your host cert/key to get pilots, in /Registry/Hosts/ :

```
lhcb-vm.tier2.hep.manchester.ac.uk
```

```
{
```

```
    DN = /C=UK/O=eScience/  
OU=Manchester/L=HEP/CN=lhcb-  
vm.tier2.hep.manchester.ac.uk
```

```
    Properties = GenericPilot
```

```
    Properties += LimitedDelegation
```

```
}
```

Configure your DIRAC CS: site

```
VAC.Manchester.uk
{
  CE = vac01.tier2.hep.manchester.ac.uk
  CEs
  {
    vac01.tier2.hep.manchester.ac.uk
    {
      CETYPE = Vac
      Queues
      {
        default
        {
          maxCPUTime = 1000
        }
      } } } }
```

“wget” bootstrapping

- Pilot 3.0 does this by

```
wget --recursive https://some.wh.ere/some/directory/  
./dirac-pilot.py ...
```

- You need an HTTPS web server where you can put the pilot files
 - The cert should be from an IGTF CA (a grid cert)
- The files are in <https://github.com/DIRACGrid/Pilot/tree/master/Pilot>
 - This is used by LHCb in production and still contains some LHCb specific code which should be ignored or fail silently for other VOs
- The key to the whole thing is the file `user_data_vm`

user_data file

- For Vac and Vcycle, this file is preprocessed before being passed to the VMs
- Patterns like `##user_data_***##` are replaced with defined values or removed
- If you're using Vac or Vcycle to start the VMs, you don't need to do anything with the file
- If you're starting VMs by hand, then you should replace
 - `##user_data_space##` with the virtual CE name of your site
 - `##user_data_uuid##` with a unique string
 - `##user_data_file_hostkey##` and `##user_data_file_hostcert#` with the PEM-encoded cert and key
 - And remove all other `##user_data_***##` patterns



Bootstrapping configuration

- For the Vacuum VM case, have a configuration bootstrapping issue
 - Generic pilot code needs to know the CS URL etc of this XyzDIRAC instance, DIRAC version, ...
 - Needs to know the site it's running at (eg may need to run different commands at different types of site)
- Pilot 3.0 includes a file pilot.json in the directory wget fetches
- This is minimal dump of the CS with enough info to configure the pilot
 - DIRAC version, commands to run, Setup, and all the CE to Site mappings

Example pilot.json file to adapt

```
{
  "Setups": {
    "Prod": {
      "CheckVersion": "True",
      "Commands": {
        "Vac": [
          "CheckWorkerNode",
          "InstallDIRAC",
          "ConfigureBasics",
          "CheckWNCapabilities",
          "ConfigureSite",
          "ConfigureArchitecture",
          "ConfigureCPURequirements",
          "MultiLaunchAgent"
        ]
      },
      "Version": "v6r17p18"
    },
    "Defaults": {
      "ConfigurationServer": "dips://xyz.example.com:9135/Configuration/Server"
    }
  },
  "CEs": {
    "vac04.tier2.hep.manchester.ac.uk": {
      "Site": "VAC.Manchester.uk",
      "GridCEType": "Vac"
    }
  },
  "DefaultSetup": "Prod"
}
```

Start VM

- For Vac and Vcycle, just do setup a machinetype as normal, giving the URL of the user_data file on your pilot directory webserver
- By hand using libvirt or VirtualBox or
 - Follow the usual CernVM instructions for the platform
 - <https://cernvm.cern.ch> then “Virtual Machine” menu then “How to run on ...”
 - Supply user_data in an additional ISO image / optical device as a Config Drive
 - Cloud Init specifies where to put this:
 - <https://cloudinit.readthedocs.io/en/latest/topics/datasources/configdrive.html>