EDF HPC

HPC System: Puppet Configuration

Date:

Pages: 34

Authors: CCN-HPC

Contents

Chapter 1

About this document

1.1 Purpose

This document contains a generic description of an HPC system in terms of its architectural views.

1.2 Typographic conventions

- Files or directories names are written in italics: */admin/restricted/config-puppet*.
- Hostnames are written in bold: **genbatch1**.
- Groups of hostnames are written using the nodeset syntax from clustershell. For example **genbatch[1-2]** refers to the servers **genbatch1** and **genbatch2**.
- Commands, configuration files contents or source code files are written in the format below:

```
$ cp /etc/default/rcS /tmp
```

1.3 Build dependencies

On a Debian Jessie system, these packages must be installed to build this documentation : - graphviz - inkscape - rubber - rubygems - texlive-latex-extra - texlive-latex-recommended - ruby-tilt - asciidoctor - ruby-puppet-strings

1.4 License

1.5 Authors

In alphabetical order:

- Benoit BOCCARD
- Thomas HAMEL
- Cécile YOSHIKAWA

Chapter 2

Overview

2.1 What is provided?

The goal of the configuration is to provide a common configuration for HPC clusters. The Puppet HPC configuration provides:

- Generic Puppet modules for tools commonly used in HPC clusters, these modules can be provided directly or coming from third party sources (mainly the puppetlabs forge)
- Specialized Puppet modules that can only work with the rest of the configuration, these modules rely on data provided by **hiera**
- A generic **hiera** configuration to provide some defaults
- A set of scripts and tools to apply, fetch and debug the configuration

2.2 What is missing?

The Puppet HPC configuration can not be applied directly, it must be configured for a specific cluster or site. This configuration is done by using **hiera**. This configuration should at least define the cluster (See: Cluster Definition).

It's also possible to expand the puppet modules to include custom modification or support for software and product not implemented in the common configuration.

2.3 Workflow

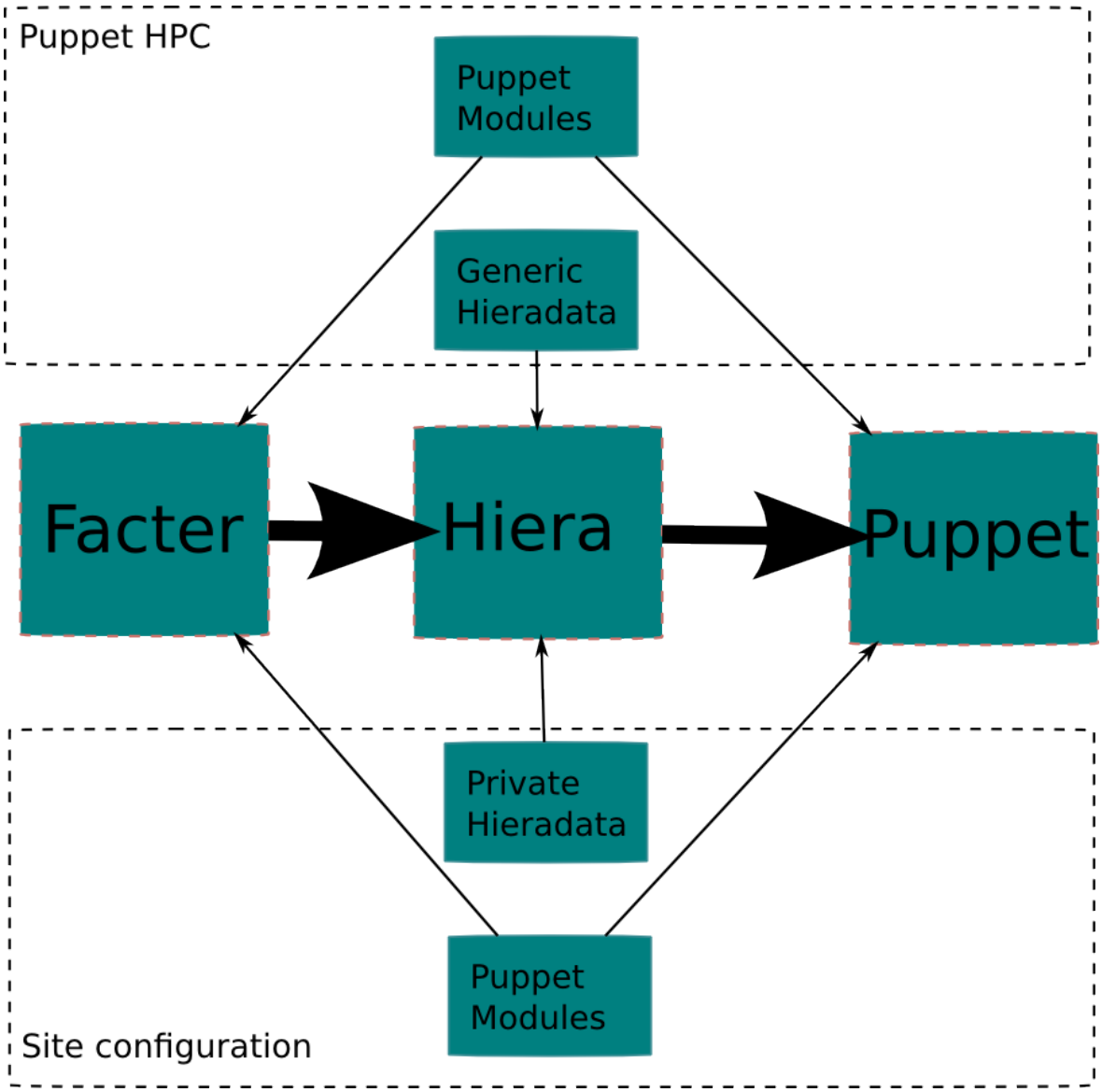

Figure 2.1:

The final configuration in Puppet is called the **catalog**, the catalog is computed from:

- The site main **manifest** (`site.pp`)
- The puppet modules included
- **hiera**
- The facts (native or from the modules)

In the hiera configuration, it is possible to reference facts and top level variables. Top level variable are defined by the site manifest.

The Puppet HPC configuration is particular because some facts (from the `hpclib` module) are defined from the cluster definition that is stored in hiera.

2.4 Customization

The aim of the configuration is to provide a common configuration base between multiple clusters and to cleanly separate configuration data from logic. This means that the primary tool for a system administrator should be to modify the hiera private configuration.

If a configuration can not be handled only by modifying hiera, it should be added as a custom module and not by modifying the Puppet HPC manifest directly.

Systems administrators are encouraged to open issues with the Puppet HPC configuration project on github in those situation. The modification might make sense on other sites or be a new default.

Chapter 3

Cluster Definition

This cluster configuration is meant to be use with a standard cluster architecture, deviation from this architecture should be minimum. Some constraints are planned to be relaxed in future.

Here, we are going to describe this architecture and how it should be defined to be used by the configuration.

3.1 Architecture

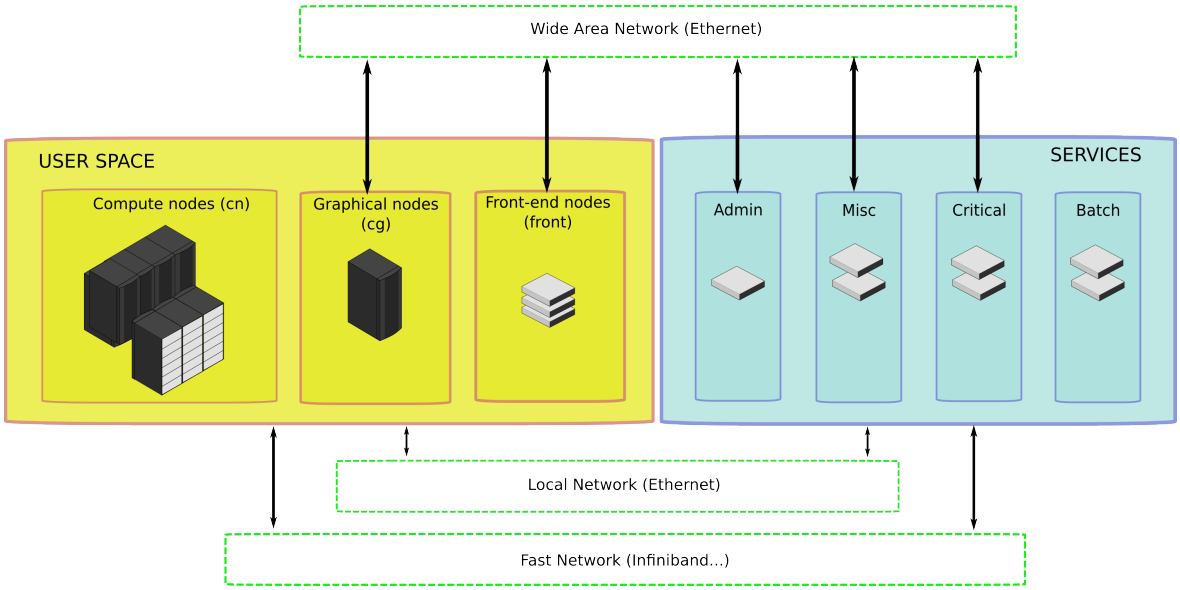

Figure 3.1:

3.2 Network definitions

3.2.1 Topology

Network topology is defined in the *cluster* level of the Hiera hierarchy. This means it is common to all nodes.

```
## Network topology of the cluster
net::allloc::ipnetwork: '172.16.0.0'
net::allloc::netmask: '255.255.0.0'
net::administration::ipnetwork: '172.16.0.0'
net::administration::netmask: '255.255.248.0'
net::administration::prefix_length: '/21'
net::administration::broadcast: '172.16.7.255'
net::lowlatency::ipnetwork: '172.16.40.0'
net::lowlatency::prefix_length: '/21'
```

```
net::management::ipnetwork: '172.16.80.0'
net::management::netmask: '255.255.240.0'
net::management::broadcast: '172.16.95.255'
net_topology:
    'wan':
        'name':           'WAN'
        'prefixes':       'wan'
        'ipnetwork':      '172.17.0.0.0'
        'netmask':        '255.255.255.0'
        'prefix_length':  '/24'
        'gateway':        '172.17.0.1'
        'broadcast':      '172.17.0.255'
        'ip_range_start': '172.17.0.1'
        'ip_range_end':   '172.17.0.254'
        'firewall_zone':  'wan'
    'allloc':
        'ipnetwork':      '172.16.0.0'
        'netmask':        '255.255.0.0'
    'administration':
        'name':           'ADM'
        'prefixes':       ''
        'ipnetwork':      '172.16.0.0'
        'netmask':        '255.255.248.0'
        'prefix_length':  '/21'
        'gateway':        '172.16.0.1'
        'broadcast':      '172.16.7.255'
        'ip_range_start': '172.16.0.1'
        'ip_range_end':   '172.16.7.254'
        'firewall_zone':  'clstr'
        'pool0':
            'ip_range_start':
                          '172.16.0.1'
            'ip_range_end':
                          '172.16.5.254'
        'pool1':          # IP reserved for the discovery process
            'ip_range_start':
                          '172.16.6.1'
            'ip_range_end':
                          '172.16.7.254'
    'lowlatency':
        'name':           'IB'
        'prefixes':       'ib'
        'ipnetwork':      '172.16.40.0'
        'netmask':        '255.255.248.0'
        'prefix_length':  '/21'
        'gateway':        ''
        'broadcast':      '172.16.47.255'
        'ip_range_start': '172.16.40.1'
        'ip_range_end':   '172.16.47.254'
        'firewall_zone':  'clstr'
    'management':
        'name':           'MGT'
        'prefixes':       'mgt'
        'ipnetwork':      '172.16.80.0'
        'netmask':        '255.255.240.0'
        'prefix_length':  '/20'
        'gateway':        ''
        'broadcast':      '172.16.95.255'
        'ip_range_start': '172.16.80.1'
        'ip_range_end':   '172.16.95.254'
        'firewall_zone':  'clstr'
    'bmc':
        'name':           'BMC'
        'prefixes':       'bmc'
        'ipnetwork':      '172.16.80.0'
```

```
        'netmask':          '255.255.248.0'
        'prefix_length':    '/21'
        'gateway':          ''
        'broadcast':        '172.16.87.255'
        'ip_range_start':   '172.16.80.1'
        'ip_range_end':     '172.16.87.254'
        'firewall_zone':    'clstr'
```

3.2.2 Bonding

Some network interfaces are bonded together for load balancing and high availability. The bonding definition is done in hiera. If the bonding is uniform (same bond interface on same slaves interfaces) between nodes, this can be done at the *cluster* level. In case of differences between nodes, it must be done higher in the hierarchy (*role* or *node*).

```
network::bonding_options:
    'bond0':
      'slaves':
         - 'eth0'
         - 'eth1'
      'options':  'mode=active-backup primary=eth0'
    'bond1':
      'slaves':
         - 'eth2'
         - 'eth3'
      'options':  'mode=active-backup primary=eth2'
```

This value from hiera uses *Auto Lookup* to be passed to the `network` class.

3.3 Node definitions

3.3.1 Master Network

Nodes are listed in a hiera array called `master_network`. This structure is derived from an internal CSV file format. Each array "line" defines one node and its network configuration.

Each line consist of five comma separated lists of values, and three lists of associations between those values.

The value lists are:

- MAC addresses
- Interface devices
- Hostnames
- IPv4 Addresses
- IPv4 Net Masks

The associations reference each list with an index starting at 0. The associations are:

- DHCP Configuration, "MAC addresses" ←→"Hostnames" ←→"IPv4 Addresses"
- Node Configuration, "Interfaces devices" ←→"IPv4 Addresses" ←→"IPv4 Netmask" ←→"External Configuration"
- DNS/Hosts Configuration, "Hostnames" ←→"IPv4 Addresses"

The interface is the device where the configuration must be applied, this means that with a bonded interface, the configuration must be applied on the `bond` interface. The interfaces enslaved to the `bond` interfaces can be ommited from this configuration.

External configuration means the interface is configured on the system but should not be setup by the Puppet HPC Configuration. It's usefull if another subsystems sets-up the network interface: VPN, libvirt... On Debian, it means the interface is not added to `/etc/network/interfaces`. This boolean can take the value `true` or `false` and is considered `false` if ommited.

Example:

```
master_network:
  #MAC_Addr;Interfaces;Hostnames;Addresses;NetMask;DHCP(Mac_Addr@Hostname@Address);Config(
    Interface@Address@Netmask);Hosts(Hostname@Address)
```

```
  - 52:54:00:ba:9d:ac,52:54:00:43:d9:45,52:54:00:8a:aa:30,52:54:00:8a:0b:d2;bond0,bond1;
    genmisc1,wangenmisc1;172.16.2.21,172.17.42.45;255.255.248.0,255.255.255.0;0@0@0;0@0@0,1
    @1@1@false;0@0,1@1
```

This example define one node (genmisc1) with the following configuration:

- DHCP
  - 52:54:00:ba:9d:ac genmisc1 172.16.2.21
- Network configuration on the node
  - bond0 172.16.2.21 255.255.248.0 External Config: false
  - bond1 172.17.42.45 255.255.255.0 External Config: false
- DNS and Hosts
  - genmisc1 172.16.2.21
  - wangenmisc1 172.17.42.45

All lists are optional, so it's possible to define element that just define a Hosts/DNS configuration (for virtual IP addresses for instance):

```
master_network:
  #MAC_Addr;Interfaces;Hostnames;Addresses;NetMask;DHCP(Mac_Addr@Hostname@Address);Config(
    Interface@Address@Netmask);Hosts(Hostname@Address)
  - ;;genmisc;172.16.2.20;;;;0@0
```

3.3.2 Roles and Profiles

Each host in the cluster must provides features to fullfill its role. Each feature is called a profile. Each role is defined by a name and a set of profiles.

Each node has a fact puppet_role that is referenced in the hiera configuration. This way it's possible to define hiera variable with values common between all the hosts with the same role. This is used to define a hiera variable profiles that list the name of all the profile applied to host with this role.

As an example, we define in hieradata/default/roles/cn.yaml:

```
profiles:
  - profiles::cluster::common
  - profiles::network::base
  - profiles::dns::client
  - profiles::access::base
  - profiles::ntp::client
  - profiles::ssmtp::client
  - profiles::jobsched::exec
  - profiles::openssh::server
  - profiles::environment::base
  - profiles::environment::codes
  - profiles::environment::modules
  - profiles::environment::vtune
  - profiles::environment::limits
  - profiles::auth::client
  - profiles::metrics::collect_base
  - profiles::metrics::collect_jobsched_exec
  - profiles::hardware::user_tuning
  - profiles::hardware::ipmi
  - profiles::log::client
```

Roles themselves are defined by the name of the hosts. The name is analysed by searching for a pattern <prefix><role name><numerical ID> (Example: gencritical1). This is done by the hpclib module which define the puppet_role fact. The prefix is the hiera variable: cluster_prefix.

Chapter 4

Glossary

- **Admin Server**, Node the system administrators use to connect interactively to the cluster
- **Batch Server**, Server hosting services related to the job scheduling system
- **gn**, see *Graphical Node*
- **cn**, see *Compute Node*
- **Compute Node**, Node with a lot of CPU power and mempory to handle the actual jobs
- **Critical Server**, Server hosting critical services used by the nodes
- **Front-End Node**, Node where the users log in interactively to manage jobs, edit files, compile codes or do command line pre/post processing. This type of node is sometime called a *Login Node*.
- **front**, see *Front-End Node*
- **Graphical Node**, Node with a GPU to handle pre/post processing tasks or GPGPU jobs
- **Misc Server**, Server hosting miscellaneous services used by the nodes

Chapter 5

Dependencies

5.1 Debian

```
# apt-get install clustershell python-yaml
```

5.2 Puppet

This is a list of current puppet Module dependencies not provided.

- puppetlabs-stdlib (debian: puppet-module-puppetlabs-stdlib)
- puppetlabs-concat (debian: puppet-module-puppetlabs-concat)
- puppetlabs-apache (debian: puppet-module-puppetlabs-apache)
- puppetlabs-apt (debian: puppet-module-puppetlabs-apt)

stdlib is special, because the version in Jessie is too old, it is in the current git archive.
Installing and using `hiera-eyaml` is recomended:

```
# apt-get install hiera-eyaml
```

If you wish to generate the documentation, installing the `ruby-puppet-strings` package is necessary.
With the Development Push & Apply installation pattern, the debian modules are only needed on the host doing the push.

5.3 Apply script

The apply script depends on Python 3 and the urllib3 module.
On debian:

```
# apt-get install python3 python3-urllib3 hiera-eyaml puppet clustershell
```

It is also provided as a package: `puppet-hpc-apply`.

Chapter 6

Installation

Installing the configuration will depend of your cluster topology. This chapter describes two installation patterns, the first uses a shared Posix storage, the second pushes the configuration on a central storage and pull it on the node with Posix or HTTP(S).

6.1 Shared `/admin`

In this setup, a storage space is mounted on every nodes of the cluster and the configuration is applied directly from this storage space. By default this space is mounted on `/admin`, using another mount point should not be difficult.

On simple systems, it is possible to use an NFS server to make `/admin` available on all nodes. It is also possible to bootstrap the cluster with a `/admin` on the *Admin Server* exported by NFS and later move it to a more resilient location (HA NFS, CephFS or GPFS).

6.1.1 Directory layout

The layout setup should be done on the first node with `/admin` available. This is generally the *Admin Server*.

- `/admin`
- `restricted`
  - `puppet-hpc` (A git clone of the puppet-hpc repository)
    - `puppet-config`
    - `hieradata`
  - `hpc-privatedata` (Frequently another git repository)
    - `hieradata`
    - `files`
  - `hieradata`
    - `generic` (Symbolic link to `/admin/restricted/puppet-hpc/hieradata`)
    - `private` (Symbolic link to `/admin/restricted/hpc-privatedata/hieradata`)
  - `privatefiles` (Symbolic link to `/admin/restricted/hpc-privatedata/files`)
- `public`
  - `http`

6.1.2 Puppet

Puppet must be configured to search for the modules in the shared `/admin`. The following file can be used on debian and also search modules installed with debian packages:

```
[main]
logdir=/var/log/puppet
vardir=/var/lib/puppet
```

```
ssldir=/var/lib/puppet/ssl
rundir=/var/run/puppet
basemodulepath=$confdir/modules:/usr/share/puppet/modules:/admin/restricted/puppet-hpc/puppet
    -config/cluster:/admin/restricted/puppet-hpc/puppet-config/modules:/admin/restricted/
    puppet-hpc/puppet-config/modules_3rdparty
prerun_command=/etc/puppet/etckeeper-commit-pre
postrun_command=/etc/puppet/etckeeper-commit-post
stringify_facts=false
hiera_config=/etc/puppet/hiera.yaml

[master]
# These are needed when the puppetmaster is run by passenger
# and can safely be removed if webrick is used.
ssl_client_header = SSL_CLIENT_S_DN
ssl_client_verify_header = SSL_CLIENT_VERIFY
```

6.1.3 Hiera-eyaml

It is recomended to use Hiera EYAML to store secret values. The keys must be created on the first node.

```
# mkdir /etc/puppet/secure
# cd /etc/puppet/secure/
# eyaml createkeys
[hiera-eyaml-core] Created key directory: ./keys
Keys created OK
# chown -R puppet:puppet /etc/puppet/secure/keys
# chmod -R 0500 /etc/puppet/secure/keys
# chmod 0400 /etc/puppet/secure/keys/*.pem
```

To configure `eyaml(1)` itself, the following file should be created: `/etc/eyaml/config.yaml`

```
---
pkcs7_private_key: '/etc/puppet/secure/keys/private_key.pkcs7.pem'
pkcs7_public_key: '/etc/puppet/secure/keys/public_key.pkcs7.pem'
```

Hiera is configured to search for values in the generic configuration repository, then in a few files for all nodes, then in files specific for each *role*.

```
:backends:
  - eyaml
:eyaml:
  :datadir:           /admin/restricted/hieradata
  :pkcs7_private_key: /etc/puppet/secure/keys/private_key.pkcs7.pem
  :pkcs7_public_key:  /etc/puppet/secure/keys/public_key.pkcs7.pem
  :extension:         'yaml'
:hierarchy:
  - private/default/roles/%{puppet_role}
  - generic/default/roles/%{puppet_role}
  - private/cluster
  - private/network
  - private/default
  - generic/common
  - generic/%{osfamily}/common
```

6.1.4 Common steps

Some common steps must be performed.

6.1.5 Node bootstraping

Setting up the directory layout can be done once, but you will still have to do some bootstraping on other newly installed nodes. Those steps will be handled by the bootsystem eventually.

The steps are:

- Installing puppet

- Distributing the puppet configuration
- Distributing the hiera configuration and keys
- Mounting `/admin`
- Apply puppet

6.2 Development Push & Apply

With this pattern, modifications to the Puppet HPC Configuration are pushed from local directories, to a central storage. The push is done to a remote POSIX file system with ssh/scp. It will eventually be possible to push with HTTP.

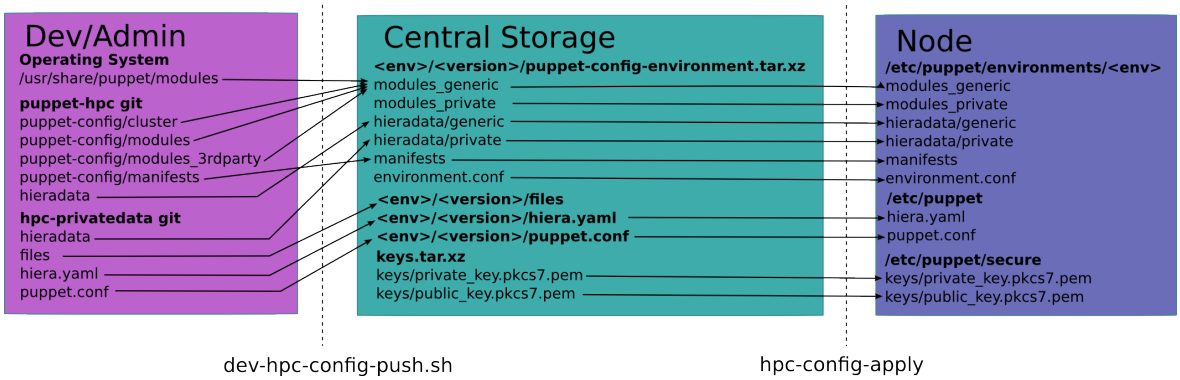

Figure 6.1:

The push script (`dev-hpc-config-push.sh`) will create a tarball with the content of the modules and the hieradata configuration. Applying the configuration uses the script that will eventually be used in production (`hpc-config-apply`).

In this setup, the apply script downloads the puppet configuration (modules and hiera data) as a tarball and installs it as an environment in `/etc/puppet`. Private data files are not downloaded with the configuration. Private data files are available from the central storage and are directly downloaded from the puppet modules.

6.2.1 Directory layout

Source directories (git puppet-hpc and git hpc-privatedata), can be placed anywhere on the Admin/Development machine. They should be in the same place though.

- `<SOME DIRECTORY>`
  - `puppet-hpc` (a git clone of the puppet-hpc repository)
  - `hpc-privatedata` a directory containing cluster specific data, frequently another git repository)
    - `hieradata`
    - `files`

The destination should be shared between the central storage servers. This can be an NFS space. This restriction is meant to be removed eventually. This is because the push to destination is a simple `scp`, the push will entually use HTTP to push to a replicated object system.

The destination is by default `/admin/public/http/mirror/hpc-config`. The path and the destination host can be modified by defining the environment variables:

- `DESTINATION_HOST`
- `DESTINATION_PATH`

6.2.2 Puppet

Puppet must be installed on the target node, if the package is used. The dependencies should take care of that.

The puppet configuration will be installed by the apply script directly from the central storage.

6.2.3 Hiera eyaml

It is recomended to use Hiera EYAML to store secret values. The keys must be created on the first node.

```
# mkdir /etc/puppet/secure
# cd /etc/puppet/secure/
# eyaml createkeys
[hiera-eyaml-core] Created key directory: ./keys
Keys created OK
# chown -R puppet:puppet /etc/puppet/secure/keys
# chmod -R 0500 /etc/puppet/secure/keys
# chmod 0400 /etc/puppet/secure/keys/*.pem
```

Hiera is configured to search for values in the generic configuration repository, then in a few files for all nodes, then in files specific for each *role*. This is done in the file `/etc/puppet/hiera.yaml`. This file is fetched by the apply script from the central storage.

The apply script will fetch the keys with the configuration. You must create a tarball with the keys and put it in a place where the apply script can fetch it. In the example below, it is placed with the puppet-configuration:

```
# cd /etc/puppet/secure
# tar cJf /admin/public/http/mirror/keys/keys.tar.xz key
```

Please not this setup is **insecure**, on a production system, the central storage for this file should only be accessible by root users on node. The apply script can prove an http request originates from a root user by using a source port in the 1-1023 range.

6.2.4 Common steps

Some common steps must be performed.

6.2.5 Usage

To push, from the admin/development host:

```
$ <SOME DIRECTORY>/puppet-hpc/scripts/dev-hpc-config-push.sh
## Creating env structure
## Creating env modules generic
cp: cannot overwrite directory '/tmp/puppet-config-production-zARV/production/modules_generic
    /stdlib' with non-directory
## Creating env modules private
## Creating env hieradata
## Creating env config
## Building Destination
## Cleaning
$
```

From a node, you must call the `hpc-config-apply` script with sources for keys and Puppet HPC configuration:

```
# <SOME DIRECTORY>/hpc-config-apply -k http://genmisc/keys -s http://genmisc/hpc-config -v
INFO:root:Getting secure keys from http://genmisc/keys/keys.tar.xz
INFO:root:Getting Puppet HPC configuration environment from http://genmisc/hpc-config/
    production/latest/puppet-config-environment.tar.xz
INFO:root:Applying puppet configuration.
Info: Loading facts
Info: Loading facts
Info: Loading facts
Info: Loading facts
Info: Loading facts
Info: Loading facts
Notice: Compiled catalog for genadmin1.hpc.example.com in environment production in 4.49
    seconds
Info: Applying configuration version '1469783844'
Notice: /Stage[main]/Apt::Update/Exec[apt_update]/returns: executed successfully
Notice: /Stage[main]/Network::Service/Service[ifup-hotplug]/ensure: ensure changed 'stopped'
    to 'running'
```

```
Info: /Stage[main]/Network::Service/Service[ifup-hotplug]: Unscheduling refresh on Service[
    ifup-hotplug]
Notice: Finished catalog run in 11.30 seconds
```

The -v switch increases verbosity.

You can use a configuration file to not always give the source, by default in /etc/hpc-config.conf:

```
[DEFAULT]
environment=production

[production]
source=http://genmisc/hpc-config/
keys_source=http://genmisc/keys/
```

The full usage for hpc-config-apply is:

```
usage: hpc-config-apply [-h] [--config [CONFIG_FILE]] [--source [SOURCE]]
                        [--environment [ENVIRONMENT]]
                        [--deploy-step [{production,usbdisk}]]
                        [--keys-source [KEYS_SOURCE]] [--verbose]

Apply the HPC Configuration.

optional arguments:
  -h, --help            show this help message and exit
  --config [CONFIG_FILE], -c [CONFIG_FILE]
                        Configuration file
  --source [SOURCE], -s [SOURCE]
                        Configuration source URL
  --environment [ENVIRONMENT], -e [ENVIRONMENT]
                        Environment name
  --deploy-step [{production,usbdisk}], -d [{production,usbdisk}]
                        Deploy step
  --keys-source [KEYS_SOURCE], -k [KEYS_SOURCE]
                        Secret keys source
  --verbose, -v         More output, can be specified multiple times.
```

6.3 Common steps

This sections presents steps that should be applied with all installation patterns.

6.3.1 Cluster Decrypt Password

A decrypt password is used by clara to decrypt files. Once you have generated this password, it should be in your hiera under this structure:

```
cluster_decrypt_password: 'PASSWORD_TO_PUT_IN_EYAML'
clara::password_options:
  ASUPASSWD:   "%{hiera('cluster_decrypt_password')}"
```

This password is used elsewhere in the hiera, generally under the name decrypt_password (for example: openssh::server::decrypt_passwd). So we define a top level variable (cluster_decrypt_password) to reuse it more easily.

6.3.2 Cluster keyring

The cluster must use a private cluster keyring. This keyring is used to sign packages generated locally and the local repositories.

You should generate it in your privatedata. You will be asked for a passphrase, this passphrase must be provided interactively when you call clara repo add|del. The following command can be pretty long to execute if you don't use a hardware Random Number Generator (RNG).

```
 # LANG=C gpg --no-default-keyring --keyring files/repo/cluster_keyring.gpg --secret-keyring
    files/repo/cluster_keyring.secret.gpg --gen-key                    gpg (GnuPG) 1.4.18;
     Copyright (C) 2014 Free Software Foundation, Inc.
```

```
This is free software: you are free to change and redistribute it.
There is NO WARRANTY, to the extent permitted by law.

gpg: keyring ‘files/repo/cluster_keyring.secret.gpg’ created
gpg: keyring ‘files/repo/cluster_keyring.gpg’ created
Please select what kind of key you want:
   (1) RSA and RSA (default)
   (2) DSA and Elgamal
   (3) DSA (sign only)
   (4) RSA (sign only)
Your selection? 1
RSA keys may be between 1024 and 4096 bits long.
What keysize do you want? (2048) 4096
Requested keysize is 4096 bits
Please specify how long the key should be valid.
         0 = key does not expire
      <n>  = key expires in n days
      <n>w = key expires in n weeks
      <n>m = key expires in n months
      <n>y = key expires in n years
Key is valid for? (0)
Key does not expire at all
Is this correct? (y/N) y

You need a user ID to identify your key; the software constructs the user ID
from the Real Name, Comment and Email Address in this form:
    "Heinrich Heine (Der Dichter) <heinrichh@duesseldorf.de>"

Real name: HPC Team Example cluster
Email address: hpc@example.com
Comment:
You selected this USER-ID:
    "HPC Team Example cluster <hpc@example.com>"

Change (N)ame, (C)omment, (E)mail or (O)kay/(Q)uit? O
You need a Passphrase to protect your secret key.

passphrase not correctly repeated; try again.
We need to generate a lot of random bytes. It is a good idea to perform
some other action (type on the keyboard, move the mouse, utilize the
disks) during the prime generation; this gives the random number
generator a better chance to gain enough entropy.
..+++++
..........+++++
We need to generate a lot of random bytes. It is a good idea to perform
some other action (type on the keyboard, move the mouse, utilize the
disks) during the prime generation; this gives the random number
generator a better chance to gain enough entropy.
+++++
.+++++
gpg: key 241FB865 marked as ultimately trusted
public and secret key created and signed.

gpg: checking the trustdb
gpg: public key of ultimately trusted key 1F2607DD not found
gpg: public key of ultimately trusted key 94DEFA86 not found
gpg: 3 marginal(s) needed, 1 complete(s) needed, PGP trust model
gpg: depth: 0  valid:   3  signed:   0  trust: 0-, 0q, 0n, 0m, 0f, 3u
pub   4096R/241FB865 2016-05-19
      Key fingerprint = D192 11C0 2EB6 BE80 A3BC  7928 1CB4 3266 241F B865
uid                  HPC Team Example cluster <hpc@example.com>
sub   4096R/C7027D3A 2016-05-19
```

Clara will use this key in its encrypted form, if there is a working `clara enc`, it is possible to use `clara enc encode` directly. Otherwise the following command will perform the encryption:

```
$ openssl aes-256-cbc -in cluster_keyring.secret.gpg -out cluster_keyring.secret.gpg.enc -k <
    cluster decrypt password>
```

6.3.3 Self signed certificate

A production system should use a certificate validated by the operating system, either through a public CA or one internal to the cluster organization. It is possible to generate a self signed certificate when that is not possible or for testing purposes.

```
root@genservice1:~/hpc-privatedata/files/ssl# openssl req -x509 -newkey rsa:2048 -keyout ssl-
    cert-generic.key -out ssl-cert-generic.pem -days 3650 -nodes
Generating a 2048 bit RSA private key
..............................+++
....+++
writing new private key to 'ssl-cert-generic.key'
-----
You are about to be asked to enter information that will be incorporated
into your certificate request.
What you are about to enter is what is called a Distinguished Name or a DN.
There are quite a few fields but you can leave some blank
For some fields there will be a default value,
If you enter '.', the field will be left blank.
-----
Country Name (2 letter code) [AU]:FR
State or Province Name (full name) [Some-State]:France
Locality Name (eg, city) []:Paris
Organization Name (eg, company) [Internet Widgits Pty Ltd]:Example
Organizational Unit Name (eg, section) []:HPC
Common Name (e.g. server FQDN or YOUR name) []:genservice1.gen.hpc.example.com
Email Address []:Example HPC <hpc@example.com>
```

The Puppet HPC configuration will use this key in its encrypted form, if there is a working `clara enc`, it is possible to use `clara enc encode` directly. Otherwise the following command will perform the encryption:

```
$ openssl aes-256-cbc -in ssl-cert-generic.key -out ssl-cert-generic.key.enc -k <cluster
    decrypt password>
```

6.3.4 SSH Keys generation

To have stable host keys between nodes reinstallations, host keys are stored encrypted in the private data. It is possible keys by cluster, role or host. A cluster should have at least a default key.

```
# cd hpc-privatedata/files/hostkeys
# ssh-keygen  -t dsa -N '' -f ssh_host_dsa_key
# ssh-keygen  -t ecdsa -N '' -f ssh_host_ecdsa_key
# ssh-keygen  -t ed25519 -N '' -f ssh_host_ed25519_key
# ssh-keygen  -t rsa -N '' -f ssh_host_rsa_key
# for i in * ; mv $i $i.default ; done
```

Last step is to encrypt the keys with the cluster decrypt password. If there is a working `clara enc`, it is possible to use `clara enc encode` directly. Otherwise the following command will perform the encryption:

```
$ openssl aes-256-cbc -in ssh_host_ed25519_key.default -out ssh_host_ed25519_key.default.enc
    -k <cluster decrypt password>
```

6.3.5 Subsystems initializations

Some subsytems needs specific initialization steps that are not handled by the profiles, refer to the following sections for specific steps:

- MariaDB/Galera: Init/Start
- OpenLDAP: Replica
- SlurmDBD: Init

Chapter 7

Operations - MariaDB/Galera

7.1 Init/Start

You have to perform this operation anytime the cluster is completely down (first boot or full reboot).

```
# echo MYSQLD_ARGS=--wsrep-new-cluster > /etc/default/mysql
# systemctl start mysql
# rm /etc/default/mysql
```

7.2 Password management

To change a password for a database user, it only needs to be changed on one of the node of the cluster. Users are also replicated between nodes.

If the passwords are correct in Hiera during the node initial installation, the passwords should be set correctly.

```
[root-genbatch1-pts0] ~ # mysql mysql
Reading table information for completion of table and column names
You can turn off this feature to get a quicker startup with -A

Welcome to the MariaDB monitor.  Commands end with ; or \g.
Your MariaDB connection id is 20
Server version: 5.5.39-MariaDB-2-wsrep (Debian), wsrep_25.10.r4014

Copyright (c) 2000, 2014, Oracle, Monty Program Ab and others.

Type 'help;' or '\h' for help. Type ear the current input statement.

MariaDB [mysql]> SET PASSWORD FOR 'debian-sys-maint'@'localhost' PASSWORD('CLEARTEXT_PASSWORD
    ') ;
```

Chapter 8

Operations - OpenLDAP

8.1 Replica

When you initialize your cluster, if you wish to use a local OpenLDAP replica, you have to execute the script `make_ldap_replica` on your replica node. This script will use an ldif file that you must provide.

8.2 Logging

8.2.1 Changing log level

To change the log level on a running server you must define a logging modification ldif file:

```
dn: cn=config
changetype: modify
replace: olcLogLevel
olcLogLevel: stats
```

The new level is applied with this command:

```
# ldapmodify -Q -Y EXTERNAL -H ldapi:/// -f /tmp/logging.ldif
```

Chapter 9

Operations - SlurmDBD

9.1 Init

After the batch server have been installed, you must create the cluster in slurmdbd:

```
# sacctmgr add cluster <cluster_name>
```

Chapter 10

Debugging

10.1 Facts

10.1.1 Listing facts

```
# puppet facts find $(hostname) --render-as=yaml
```

It's possible to use the -d (debug) flag with this command.

10.1.2 Debugging facts

This small script can be used to obtain a more precise error when a fact is failing:

```
#!/bin/bash

hpc_puppet_modules_dir=/admin/restricted/puppet-hpc/puppet-config/modules

for dir in /admin/restricted/puppet-hpc/puppet-config/modules/*
do
  FACTERLIB="$FACTERLIB:${dir}/lib/facter"
done

export FACTERLIB

facter "${@}"
```

All traditional facter flags are working:

```
# ./hpc_facter -p --trace
undefined method ‘empty?’ for nil:NilClass
/admin/restricted/puppet-hpc/puppet-config/modules/hpclib/lib/facter/network.rb:91:in ‘<top (
    required)>’
/usr/lib/ruby/vendor_ruby/facter/util/loader.rb:130:in ‘load’
/usr/lib/ruby/vendor_ruby/facter/util/loader.rb:130:in ‘kernel_load’
/usr/lib/ruby/vendor_ruby/facter/util/loader.rb:115:in ‘load_file’
/usr/lib/ruby/vendor_ruby/facter/util/loader.rb:49:in ‘block (2 levels) in load_all’
/usr/lib/ruby/vendor_ruby/facter/util/loader.rb:47:in ‘each’
/usr/lib/ruby/vendor_ruby/facter/util/loader.rb:47:in ‘block in load_all’
/usr/lib/ruby/vendor_ruby/facter/util/loader.rb:45:in ‘each’
/usr/lib/ruby/vendor_ruby/facter/util/loader.rb:45:in ‘load_all’
/usr/lib/ruby/vendor_ruby/facter/util/collection.rb:104:in ‘load_all’
/usr/lib/ruby/vendor_ruby/facter.rb:126:in ‘to_hash’
/usr/lib/ruby/vendor_ruby/facter/application.rb:46:in ‘run’
/usr/bin/facter:9:in ‘<main>’
```

Chapter 11

Internals - Roles and Profiles

11.1 Overview

The Puppet configuration for HPC clusters has been designed following the **Roles and Profiles** pattern. The Puppet code is organized in different levels of abstraction :

- Roles, which represent the business logic. A node includes one role, and one only. Each role lists one or more profiles.
- Profiles, which represent the implementation of technical functionalities. A profile includes and manages modules to define a logical technical stack.
- Modules, which are the elementary technical blocks. Modules should only be responsible for managing aspects of the component they are written for and should be as generic as possible.

11.2 Background

Roles and Profiles is a common pattern in puppet code organisation. The pattern is explained in details by this presentation: https://puppet.com/presentations/designing-puppet-rolesprofiles-pattern

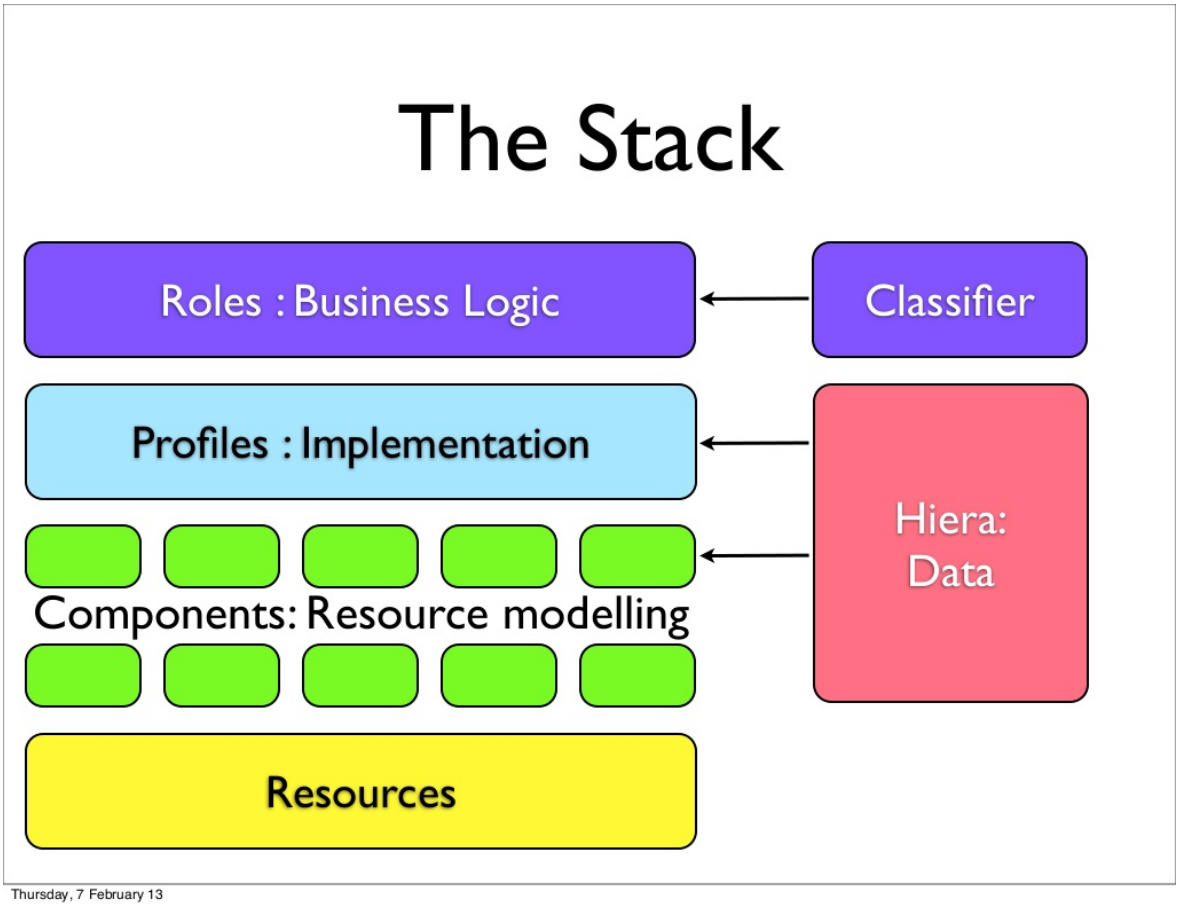

Figure 11.1: The stack (from *Designing Puppet: Roles/Profiles Pattern* by Craig Dunn)

11.3 Implementation

11.3.1 Node classifiers

A role must be associated to a node. In the Puppet HPC configuration, it is assumed the node is named after its role. The naming rule for a node follows the pattern :

- `<prefix><role name><numerical ID>` (Example: `gencritical1`)

  where `<prefix>` is the value of the variable `cluster_prefix` defined in Hiera.

  To retrieve the role of a node, its hostname is analyzed by searching for the pattern above. This is done by the Ruby script `role.rb` of the `hpclib` module which defines the `puppet_role` fact.

11.3.2 Roles

In the Puppet HPC configuration, each role is associated to the list of profiles it includes. The content of each role is specified in Hiera (`hieradata/default/roles`) in the corresponding YAML file `<role>.yaml`. This YAML file includes the definition of a `profiles` array, which is the list of profiles the role is composed of.

A simple module (`puppet-config/cluster/roles`) is loading all the profile classes defined in the Hiera `profiles` array. The configuration is providing some role definitions (critical, misc, batch, etc) but it is not mandatory to use them. Any site may define its own roles. That is why referencing a role by referencing its name directly should be avoided.

The module `hpclib` defines some facts that indicate which role includes a profile with some specific keywords in its name (`server`, `relay`, `mirror`, `replica`, `tracker`). Looking at the facts on any node of the cluster, it is easy to know which role implements the DNS server or the LDAP replica for example.

Example in `hieradata/common.yaml`:

```
conmand:     "%{hiera('cluster_prefix')}%{my_conman_server}"
```

The example above assumes an high-availability virtual IP address is defined for the role defining including the profile `profiles::conman::server`.

11.3.3 Profiles

Profiles are classes grouped inside the `profiles` module (`puppet-config/cluster/profiles`). Each profile is a feature for a host. That feature is built from technical components defined by modules. Profiles can rely heavily on Hiera to get the configuration for the site. The profiles do not use auto-lookup but explicit `hiera` (or `hiera_hash` and `hiera_array`) calls, this permits to control how data coming from multiple hierarchy levels are merged.

A profile should not need a hiera key if the value is just passed to the component class. This kind of parameters should be passed to the module class directly with *Auto Lookup*. Only parameters processed or retrieved with `hiera_array` and `hiera_hash` should be retrieved from the profile.

Doing this avoid redefining all module parameters in the profiles. Default values should be set in the `common.yaml` of the Puppet HPC configuration `hieradata` directory.

bad

`profiles/manifests/toto/server.pp`:

```
class profiles::toto::server {
  $peers = hiera_array('profiles::toto:server::peers')
  $port = hiera('profiles::toto:server::port')
  class { '::toto::server':
    port  => $port,
    peers => $peers,
  }
}
```

`cluster.yaml`:

```
profiles::toto::server::port: '42'
profiles::toto::server::peers:
  - servera
  - serverb
```

good

`profiles/manifests/toto/server.pp`:

```
class profiles::toto::server {
  $peers = hiera_array('profiles::toto:server::peers')
  class { '::toto::server':
    peers => $peers
  }
}
```

`cluster.yaml`:

```
toto::server::port: '42'
profiles::toto::server::peers:
  - servera
  - serverb
```

The `profiles` module only defines classes, no resources, methods or facts. It should not use templates or files sourced in the `profiles` module. Profiles are meant to rely on the rest of the Hiera/Puppet HPC configuration and may not work outside of it. Profiles can call any class except other profiles or roles.

11.3.4 HPC modules

HPC modules are modules that provide a lower level implementation of a technical solution than a profile, but they still rely on the Puppet HPC configuration as a whole. This includes modules that use facts defined by `hpclib`.

Those modules can use other modules and define resources, templates, files, facts, functions... Their name can be prefixed by `hpc_`.

11.3.5 Generic modules

Generic modules implement a specific solution directly in a manner that is not specific to the Puppet HPC configuration. These modules can not rely on data other than class parameters or standard facts.

Modules must be able to work without Hiera. Despite the fact that these modules can be used independently of the Puppet HPC configuration, most of them have been created to be used in the context of an HPC cluster. As a result, the features they support might seem incomplete for other uses. If you feel this way, please feel free to contribute and add features to make our modules as generic as possible.

HPC and Generic modules are not strictly separated. Whether data specific to the Puppet HPC configuration is used in the module or not is what makes a module falls into a category or the other.

11.3.6 Third Party modules

Usage of third party generic modules is encouraged when possible. The limitation is that it should come from a reliable source: distribution package or puppetlabs forge. For modules from the forge, support level should be **approved** at least. Exceptions to this rule is possible if the module has been properly reviewed.

Third party modules should be copied in the `puppet-config/modules_3rdparty` directory. This may change in the future.

Forking a third party module should be avoided as much as possible. If a suitable third party module does not provide all the necessary features, it is recommended to create another module that will **wrap** the third party module. Upstreaming the new feature when it makes sense is encouraged. In some cases, the wraping can be done directly in the profile (eg. `profiles::ntp::server`) to avoid creating a module.

Chapter 12

Internals - Writing Modules

12.1 Feature set

The Puppet HPC configuration uses the parser from Puppet < 4 (not the future parser). Modules must not use constructs that are only available with the parser from Puppet 4 (`foreach...`). Compatibility with the future parser is encouraged though.

You must assume that the manifest will be applied with the following setting in `puppet.conf`:

```
stringify_facts=false
```

The configuration provides a script `validate.sh` that checks the syntax with `puppet-lint`. You can use it or run `puppet-lint` yourself with the following arguments:

- `-no-class_inherits_from_params_class-check`, the configuration does not support puppet < 3.0, so this check is ignored
- `-no-80chars-check`, the limit in the style guide is 140 characters, but `puppet-lint` in Debian Jessie is not up to date

12.2 Style

Puppetlabs provides a style guide that should be respected by all modules: https://docs.puppet.com/guides/style_guide.h

12.3 Classes

A class with the same name as the service should exist, this class inherits a `params` class that defines default values for the software. Defaults should change when appropriate (eg. different package names for different OS).

Most modules follow a pattern where the following subclasses are defined:

- `install`, install the packages and other files
- `config`, setup the software (write configuration files)
- `service`, configure and launch the service

Classes `install`, `config` and `service` should inherit the main module class. They are included by the main module class with the proper requirements (typically `install` →`config` →`service`).

This pattern is used by the NTP module from the puppetlabs forge: https://github.com/puppetlabs/puppetlabs-ntp

If a module has multiple subclasses (a client and a server for example). This pattern can be replicated in a subdirectory of `manifests`. The `mysql` module from the puppetlabs forge does this: https://github.com/puppetlabs/pu mysql

12.4 Variables

To maintain consistency between modules, some recurring variables and/or parameters should use the following standardised names:

- `service`, the name of the service for this software (string)
- `service_ensure`, should the service be *running* or *stopped*
- `service_enable`, should the service be started on boot (boolean)
- `config_enc`, encrypted source of a configuration file (see the `hpclib::decrypt‘` function) (string)
- `config_file`, absolute file path of the configuration file (string)
- `config_options`, hash with the values to put in the `config_file` (hash)
- `packages`, list of packages to manage for this software (array of strings)
- `packages_ensure`, state of the packages: *latest* or *installed*
- `default_file`, configuration file path for init system (*/etc/default/xxx* or */etc/sysconfig/xxxx*)
- `default_options`, values to put in the `default_file` (hash)

If the module has multiple configuration files, the pattern *xxxx_enc*, *xxxx_file* and *xxxx_options* should be kept.

12.5 Documentation

All modules should have a file `README.md` at the root of the module. This file should follow the guidelines spelled out in the Puppetlabs Style Guide.

Even if the Puppet HPC configuration does not yet use the future parser, the class/resource documentation uses puppet strings: https://github.com/puppetlabs/puppetlabs-strings

The use of this tool makes the writing of the **reference** section in the README.md file redundant.

Code comments should use the markdown syntax (as recommended by the puppet strings documentation). The root of the module should include a `.yardopts` file with the following content:

```
--markup=markdown
--markup-provider=redcarpet
--verbose
```

12.6 Testing

Automatic module testing is based on rspec-puppet (http://rspec-puppet.com/) and puppetlabs-spec-helper (https://github.com/puppetlabs/puppetlabs_spec_helper). These two are packaged for Debian 8 in packages ruby-rspec-puppet and ruby-puppetlabs-spec-helper.

All modules should have a minimal batch of rspec-puppet tests to ensure the module can compile.

Indications on how to write test cases can be found here : * https://puppet.com/blog/next-generation-of-puppet-module-testing * http://rspec-puppet.com/tutorial/

12.6.1 Files

Some files are necessary at the root of the module to implement the tests : * Rakefile :

```
require 'rake'
require 'rspec/core/rake_task'
require 'puppetlabs_spec_helper/rake_tasks'
require 'rubygems'
require 'puppetlabs_spec_helper/module_spec_helper'
require 'puppet-lint/tasks/puppet-lint'


desc "Run all RSpec code examples"
RSpec::Core::RakeTask.new(:rspec) do |t|
  t.rspec_opts = File.read("spec/spec.opts").chomp || ""
end

SPEC_SUITES = (Dir.entries('spec') - ['.', '..','fixtures']).select {|e| File.directory? "
    spec/#{e}" }
namespace :rspec do
  SPEC_SUITES.each do |suite|
    desc "Run #{suite} RSpec code examples"
    RSpec::Core::RakeTask.new(suite) do |t|
      t.pattern = "spec/#{suite}/**/*_spec.rb"
```

```
      t.rspec_opts = File.read("spec/spec.opts").chomp || ""
    end
  end
end
task :default => :rspec

begin
  if Gem::Specification::find_by_name('puppet-lint')
    require 'puppet-lint/tasks/puppet-lint'
    PuppetLint.configuration.ignore_paths = ["spec/**/*.pp", "vendor/**/*.pp"]
    task :default => [:rspec, :lint]
  end
rescue Gem::LoadError
end
```

- .fixtures.yml : File used by the common rake tasks to automatically install dependencies for test runs. This file should indicate how to access to all the dependancies. For example, this a .fixture.yml file for a module called `ctorrent` using the `puppetlabs-stdlib` module installed via the standard debian package :

```
fixtures:
  symlinks:
    ctorrent: "#{source_dir}"
    stdlib: '/usr/share/puppet/modules/stdlib'
```

- spec/spec.opts : Used to indicate options to use with rspec

```
--format s --colour --backtrace
```

- spec/classes/*.rb Files containing rspec tests for classes of the module. It should contain at least one test for the main class of the module verifying it can compile. Exemple below :

```
# cat init_spec.rb
require 'spec_helper'
describe 'ctorrent' do
  context 'The following classes should be present in the catalog' do

    it { should compile }               # this is the test to check if it compiles.

  end
end
```

All the test cases concerning a class are written inside one file. The name of this file must be `[class_name]_spec.rb`.

- spec/defines/*.rb Files containing rspec tests for defines of the module.
- spec/functions/*.rb Files containing rspec tests for functions of the module.

12.6.2 Launching tests

Tests are launched via Rake tasks:

```
#rake help
rake help
rake build            # Build puppet module package
rake clean            # Clean a built module package
rake coverage         # Generate code coverage information
rake help             # Display the list of available rake tasks
rake lint             # Check puppet manifests with puppet-lint / Run puppet-lint
rake rspec            # Run all RSpec code examples
rake rspec:classes    # Run classes RSpec code examples
rake rspec:defines    # Run defines RSpec code examples
rake rspec:functions  # Run functions RSpec code examples
rake rspec:hosts      # Run hosts RSpec code examples
```

```
rake spec             # Run spec tests in a clean fixtures directory
rake spec_clean       # Clean up the fixtures directory
rake spec_prep        # Create the fixtures directory
rake spec_standalone  # Run spec tests on an existing fixtures directory
```

The main command to use is `rake spec`. It will create the dependancies in `spec/fixtures/` directory, run the rspec code and finally clean the test directory. Same result can be achieved by running successively `rake spec_prep`, `rake spec_standalone` and `rake spec_clean`.

When running the tests, results are clearly indicated :

```
#rake spec
/usr/bin/ruby2.1 -S rspec spec/classes/init_spec.rb --color
........

Finished in 2.07 seconds
8 examples, 0 failures
```

Chapter 13

Internals - Documentation

13.1 Main Documentation

13.1.1 Dependencies

-
- **inkscape** (for SVG to PNG conversion)

13.1.2 Generation

Main documentation (this document) uses `asciidoctor` to generate the documentation. From the `doc/` directory a call to `make` will generate the PDF and HTML version.

13.1.3 Modification

Chapters modification are be done in the `.asc` files in the `doc/src` directory. To add a chapter, you must edit the main documentation file: `doc/PuppetHPCConfiguration.asc` and add an include entry.

The included files should use the `.asc` extension, this permits GitHub to automatically detect the format and render the pages when browsing the git repository.

13.1.4 Figures

Figures and images should go into the `doc/src/img` directory. If there is an SVG file without a corresponding PNG file, make will call **inkscape** to generate it. Only the SVG source should be in the repository.

13.2 Puppet Documentation

13.2.1 Dependencies

- **puppet strings gem**
- **yard**

The puppet strings gem has been packaged for debian: https://github.com/edf-hpc/ruby-puppet-strings

13.2.2 Generation

In a module directory, call the command: `puppet strings`. It will generate a `doc/` directory with HTML pages.

13.3 GitHub Pages

The documentation is available in the GitHub pages: https://edf-hpc.github.io/puppet-hpc/

This set of file is in the `gh-pages` branch. This branch can be updated with a current documentation by calling the script: `build-gh-pages.sh`.

```
$ cd doc/
$ ./build-gh-pages.sh
```

This script will clone the current master branch of the local repository to a temporary location, generate the main documentation and puppet documentations for profiles and modules. The result will be commited in the gh-pages branch of the local repository.

To publish on GitHub, the branch should be pushed:

```
$ git push origin gh-pages:gh-pages
```

Last updated 2016-08-18 08:49:26 BST