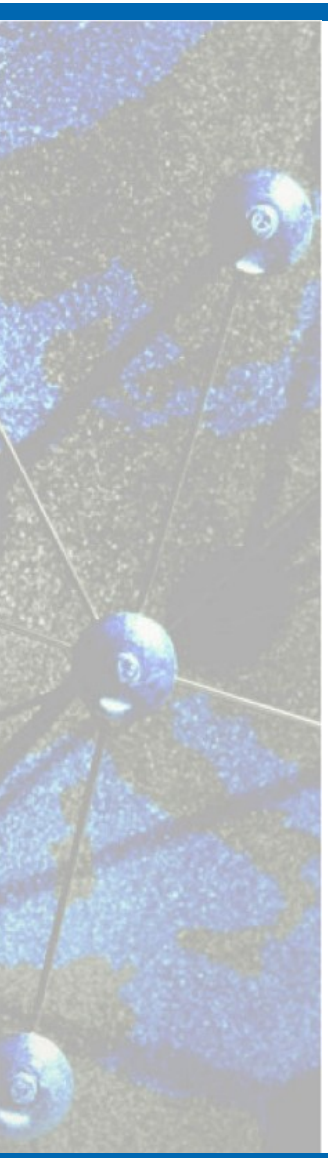


Capabilities and Best Practices for Cloud Computing

Björn Hagemeier, Boris Parak, Miroslav Ruda

Overview

- Motivation
- Questionnaire
- Use Cases
- Recommendations
- General Best Practices
- Future Work

Motivation

- Many use case images were not making use of cloud capacities
- Found number of use cases providing naive VM image layout for cloud usage
  - Does not scale, as some images were really large
  - Does not make use of cloud specific features, e. g.
    - Scalable storage (block/volume or object)
    - Scaling compute instances file systems at startup
- No common agreement on contextualization

Questionnaire

- Covering the areas of
  - Image preparation and management
  - Workload management
  - AAI and contextualization
  - (Big) Data
- Individualized questionnaire sent to seven user communities in May
- Received feedback from four of them
- Another one needs to define what they want in more detail, before we can ask technical questions

User Communities

- After evaluation of the questionnaires, we decided to directly support two user communities
  - BioVel
  - WeNMR
- An analysis of the images provided by the user communities has additionally been done for BNCweb

Findings

- 2/3 contained large numbers of unneeded packages and data
  - Desktop environment (auto-login)
  - even Games
  - Kernel images and modules for up to 13 kernel versions
  - Development environments
  - non-empty, unused Swapfile
  - Continuous integration server with data
- BioVel had a rather short list of packages
  - Data shipped with the image

WeNMR

- Ubuntu 11.04 (discontinued support)
- Long list of installed packages: 1666
- Continuous integration environment: Jenkins (1.2GB)
- Non-empty, but unused swapfile: 1GB, doesn't compress well
- /lib/modules for 13 Kernel versions: 1GB
- Full-blown desktop environment with auto-login
- no contextualization
  - Hard-coded Topos token server

BNCweb

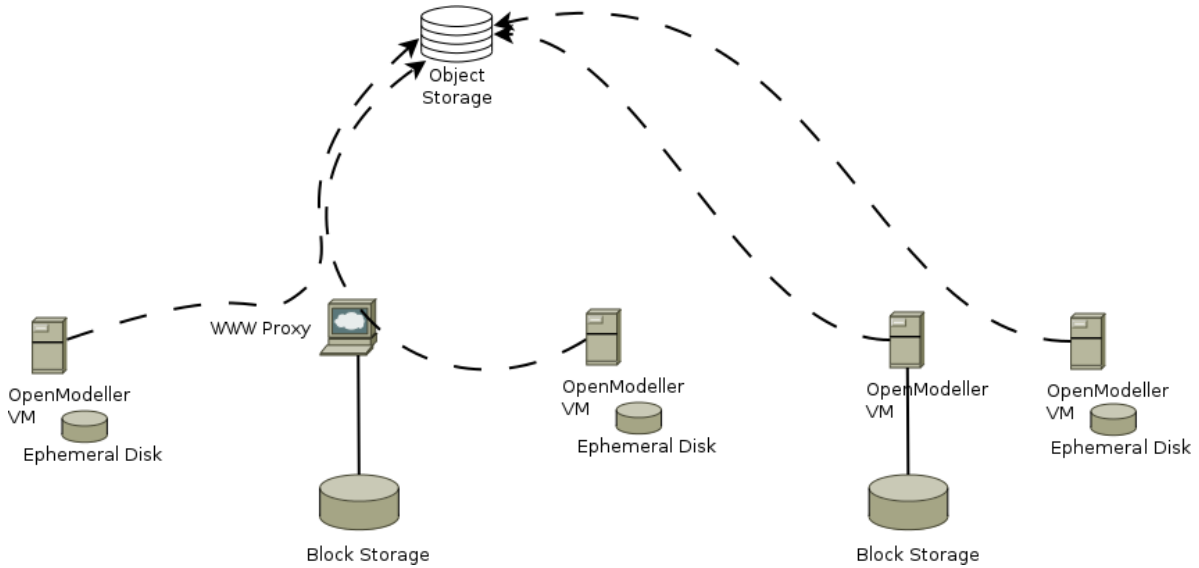
- Ubuntu 11.04 (discontinued support)
- Long list of installed packages: 1416
- Debian package cache not cleaned: 270MB
- Static data shipped with image: 340MB
- /usr/local/share/corpora: 4.7GB
- /usr/local/share/cwb: 2.5GB
- Kernel modules for 6 versions: 800MB

BioVel

- Smallest list of installed packages among images provided by use cases
  - These could be reduced, e.g. graphical and development environments may not be needed
- Shipping layer data with the image
  - 3.2G for test, could easily grow 10 times the size
  - We envision a scheme to provide this data from Cloud Object Storage
- Source: `http://bit.ly/19GFZSN`

BioVel

- Improve Data Handling
  - Do not ship huge amounts of data with the image
  - They're not even needed in every instance
- Options



- This scheme is generally applicable

Recommendations

Basic Image

- Provide recipes of how to create minimal OS installations
- Clearly state the pitfalls to avoid
- Further documentation about minimizing installations and images even more

Application

- Data management
- Make use of cloud block and object storage

General Best Practices

- Create minimal images
  - We have assembled information about how to achieve that and what are the common pitfalls
  - e. g. do not put a desktop environment into the image unless you really need it
  - avoid huge data payloads
- Make use of cloud features
  - Use block storage or object storage whenever applicable
  - Besides being able to keep images smaller, this will give you more reliable persistence
- Use contextualization
  - This is still the most vague field, although we recommend the use of cloud-init
  - We're continuously documenting our findings about cloud-init, as it is different among Linux distributions

Future Work

- Wrapping up the information we gathered in a concise, well structured report
- So far, information has been gathered in our Wiki
- Track our success by re-evaluating new images provided by user communities

Summary

- Questionnaire
- Image Evaluation
- Recommendations
- General Best Practices

- *We offer direct support!*