Cloud computing
What is it ?
Characteristics
Models
Solutions & ressources
StratusLab
Ressources
How to use a Cloud
MarketPlace
Monitoring
End-user client
Connexion
Persistent disk
Algorithm preservation in the Cloud
Requirements
Tools
Conclusion
References

Cloud technology for algorithm preservation

Cécile Cavet
cecile.cavet at apc.univ-paris7.fr

Centre François Arago (FACe), Laboratoire APC, Université Paris Diderot
LabEx UnivEarthS

November 5, 2014

Plan

1 Cloud computing

2 StratusLab

3 Algorithm preservation in the Cloud

4 Conclusion

Cloud computing
- What is it ?
- Characteristics
- Models
- Solutions & ressources

StratusLab
- Ressources
- How to use a Cloud
- MarketPlace
- Monitoring
- End-user client
- Connexion
- Persistent disk

Algorithm preservation in the Cloud
- Requirements
- Tools

Conclusion
- References

What is Cloud computing ?

➡ Provides IT resources on-demand.

Full definition of Cloud Computing: NIST report [1].

Cloud computing
What is it ?
Characteristics
Models
Solutions & ressources
StratusLab
Ressources
How to use a Cloud
MarketPlace
Monitoring
End-user client
Connexion
Persistent disk
Algorithm preservation in the Cloud
Requirements
Tools
Conclusion
References

What is Cloud computing ? A mature project

1. Hardware: virtualization of all ressources of commodity hardware.

Cloud computing
What is it ?
Characteristics
Models
Solutions & ressources
StratusLab
Ressources
How to use a Cloud
MarketPlace
Monitoring
End-user client
Connexion
Persistent disk
Algorithm preservation in the Cloud
Requirements
Tools
Conclusion
References

The virtualization revolution

Cloud computing
What is it ?
Characteristics
Models
Solutions & ressources
StratusLab
Ressources
How to use a Cloud
MarketPlace
Monitoring
End-user client
Connexion
Persistent disk
Algorithm preservation in the Cloud
Requirements
Tools
Conclusion
References

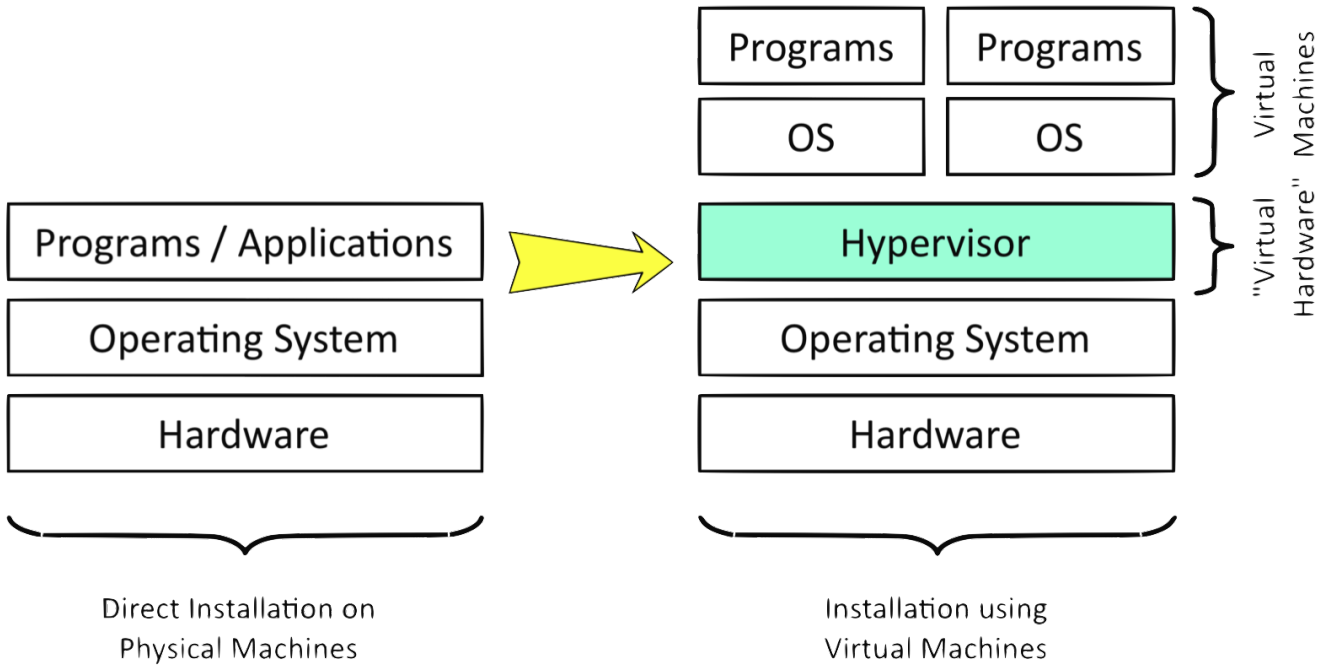


Figure: Principle of virtualization [2].

What is Cloud computing ? A mature project

1. Hardware: virtualization of all ressources of commodity hardware.
2. Software: simplified APIs (software interface) ➞ end-user client and Web interface.

Cloud computing
What is it ?
Characteristics
Models
Solutions & ressources
StratusLab
Ressources
How to use a Cloud
MarketPlace
Monitoring
End-user client
Connexion
Persistent disk
Algorithm preservation in the Cloud
Requirements
Tools
Conclusion
References

Simplified APIs

Cloud computing
- What is it ?
- Characteristics
- Models
- Solutions & ressources

StratusLab
- Ressources
- How to use a Cloud
- MarketPlace
- Monitoring
- End-user client
- Connexion
- Persistent disk

Algorithm preservation in the Cloud
- Requirements
- Tools

Conclusion
- References

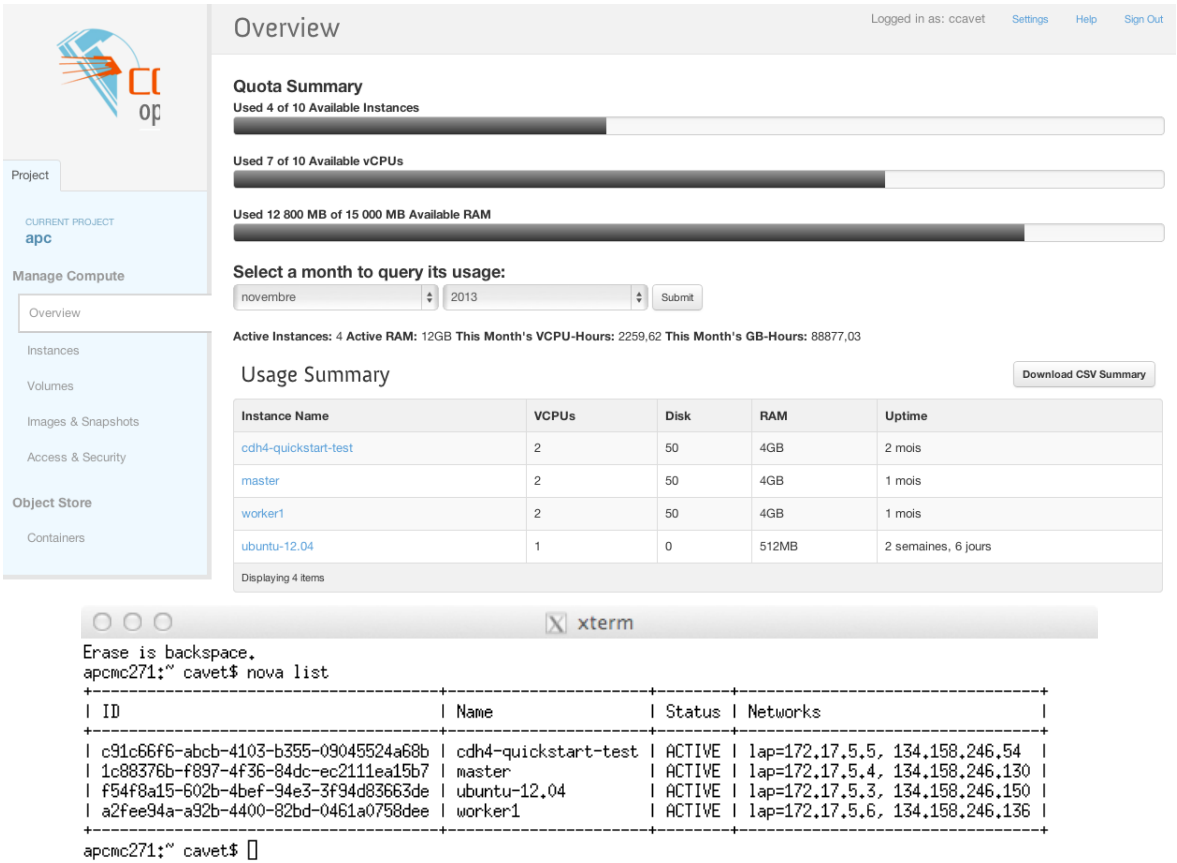


Figure: Horizon dashboard and Nova command-line client of OpenStack.

What is Cloud computing ? A mature project

Cloud computing
What is it ?
Characteristics
Models
Solutions & ressources
StratusLab
Ressources
How to use a Cloud
MarketPlace
Monitoring
End-user client
Connexion
Persistent disk
Algorithm preservation in the Cloud
Requirements
Tools
Conclusion
References

1. Hardware: virtualization of all ressources of commodity hardware.
2. Software: simplified APIs (software interface) ➟ end-user client and Web interface.
3. Resources: excess of **commercial** computing resources.
   ➟ In 2006: 50 % of Amazon's resources was not used.
   ➟ Now: growing infrastructures (academic and commercial).

What is Cloud computing ? Essential characteristics

Cloud computing
- What is it ?
- Characteristics
- Models
- Solutions & ressources

StratusLab
- Ressources
- How to use a Cloud
- MarketPlace
- Monitoring
- End-user client
- Connexion
- Persistent disk

Algorithm preservation in the Cloud
- Requirements
- Tools

Conclusion
- References

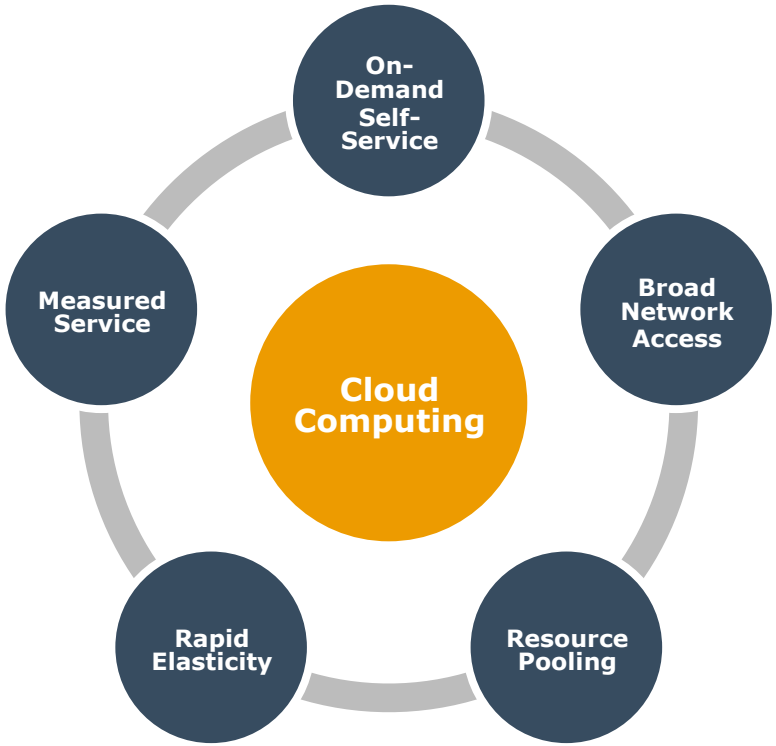


Figure: Attributes of Cloud computing.

➡ Huge flexibility for scientific applications.

What is Cloud computing ? A clump of Clouds...

Cloud computing
What is it ?
Characteristics
Models
Solutions & ressources
StratusLab
Ressources
How to use a Cloud
MarketPlace
Monitoring
End-user client
Connexion
Persistent disk
Algorithm preservation in the Cloud
Requirements
Tools
Conclusion
References

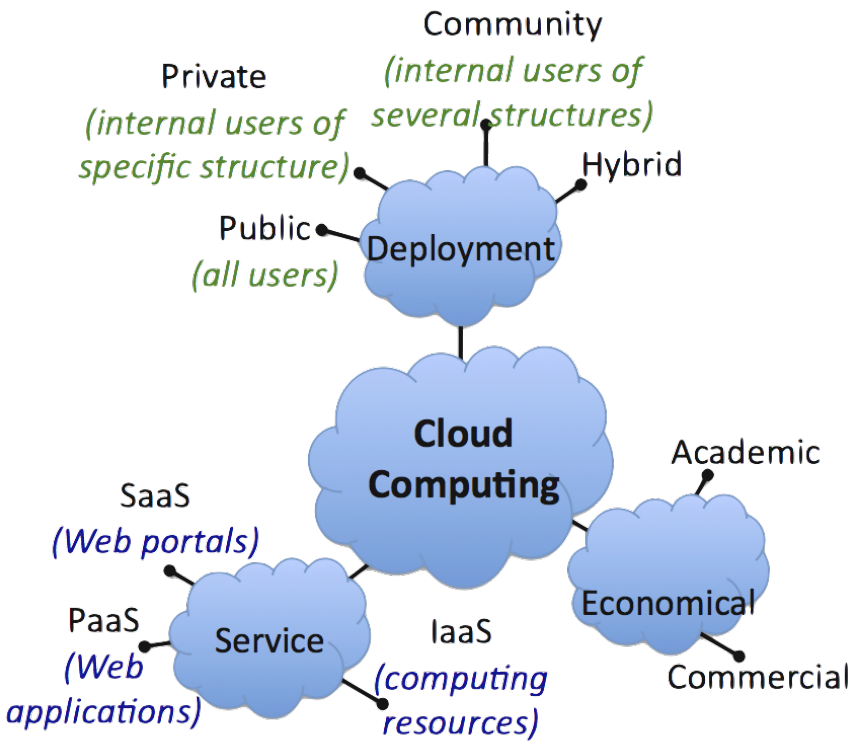


Figure: Different models of Cloud infrastructure (from [3]).

➡ We will focus on **academic public IaaS** Cloud.

Cloud computing
What is it ?
Characteristics
Models
Solutions & ressources
StratusLab
Ressources
How to use a Cloud
MarketPlace
Monitoring
End-user client
Connexion
Persistent disk
Algorithm preservation in the Cloud
Requirements
Tools
Conclusion
References

Cloud solution & ressources

Clouds in production phase:

- **StratusLab**@LAL [4] (EU, 2010):

  http://stratuslab.eu/index.html

- **OpenStack**@CC-IN2P3 (FR, 2012):

  http://ccwiki.in2p3.fr/infrastructure:cloud:start

➙ **Performance benchmarks, tests and scientific applications.**

Cloud computing
What is it ?
Characteristics
Models
Solutions & ressources
StratusLab
Ressources
How to use a Cloud
MarketPlace
Monitoring
End-user client
Connexion
Persistent disk
Algorithm preservation in the Cloud
Requirements
Tools
Conclusion
References

StratusLab Cloud :

- Cloud manager: **StratusLab** with libvirt (in dec. 2014, ~~OpenNebula~~).
- Resources **LAL@Orsay**: 16 nodes, 440 cores, 772 GB of memory, 15 TB of disk space, 1 GbE/s inter-node connexion.
- Utilization: $\sim$ 90%.
- Services:
  - End-user client (http://stratuslab.eu/index.html).
  - MarketPlace (http://marketplace.stratuslab.eu).
  - Persistent disk Web interface (https://pdisk.lal.stratuslab.eu/svc-pdisk.html).
  - Ressource monitoring (https://cloud.lal.stratuslab.eu/load.txt).

How to use the StratusLab Cloud

Cloud computing
- What is it ?
- Characteristics
- Models
- Solutions & ressources

StratusLab
- Ressources
- How to use a Cloud
- MarketPlace
- Monitoring
- End-user client
- Connexion
- Persistent disk

Algorithm preservation in the Cloud
- Requirements
- Tools

Conclusion
- References

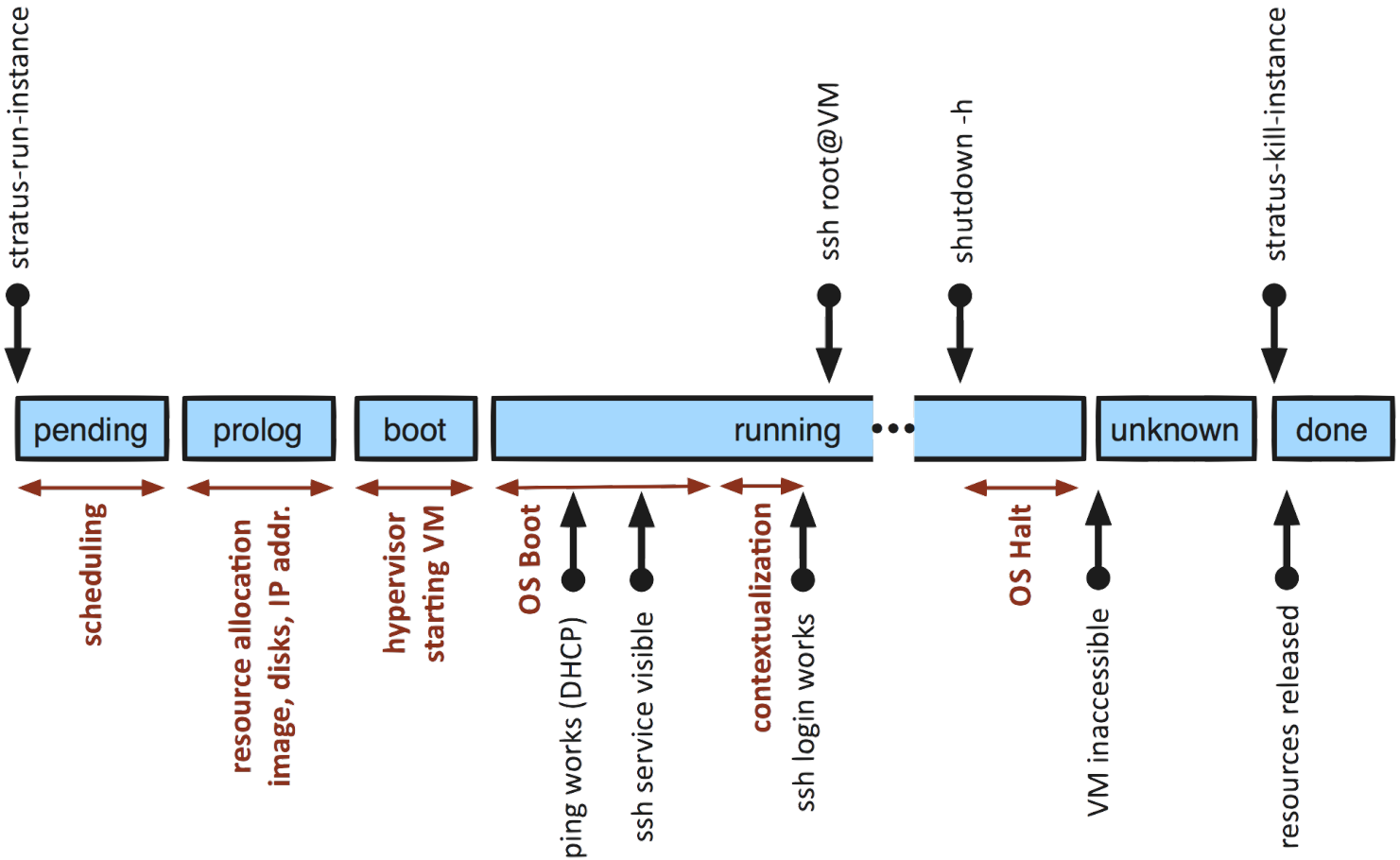


Figure: Virtual machine life-cycle @StratusLab.

Cloud computing
What is it ?
Characteristics
Models
Solutions & ressources
StratusLab
Ressources
How to use a Cloud
MarketPlace
Monitoring
End-user client
Connexion
Persistent disk
Algorithm preservation in the Cloud
Requirements
Tools
Conclusion
References

How to use the StratusLab Cloud

- Choice of the OS.

MarketPlace: share Disk Image

Cloud computing
- What is it ?
- Characteristics
- Models
- Solutions & ressources

StratusLab
- Ressources
- How to use a Cloud
- MarketPlace
- Monitoring
- End-user client
- Connexion
- Persistent disk

Algorithm preservation in the Cloud
- Requirements
- Tools

Conclusion
- References

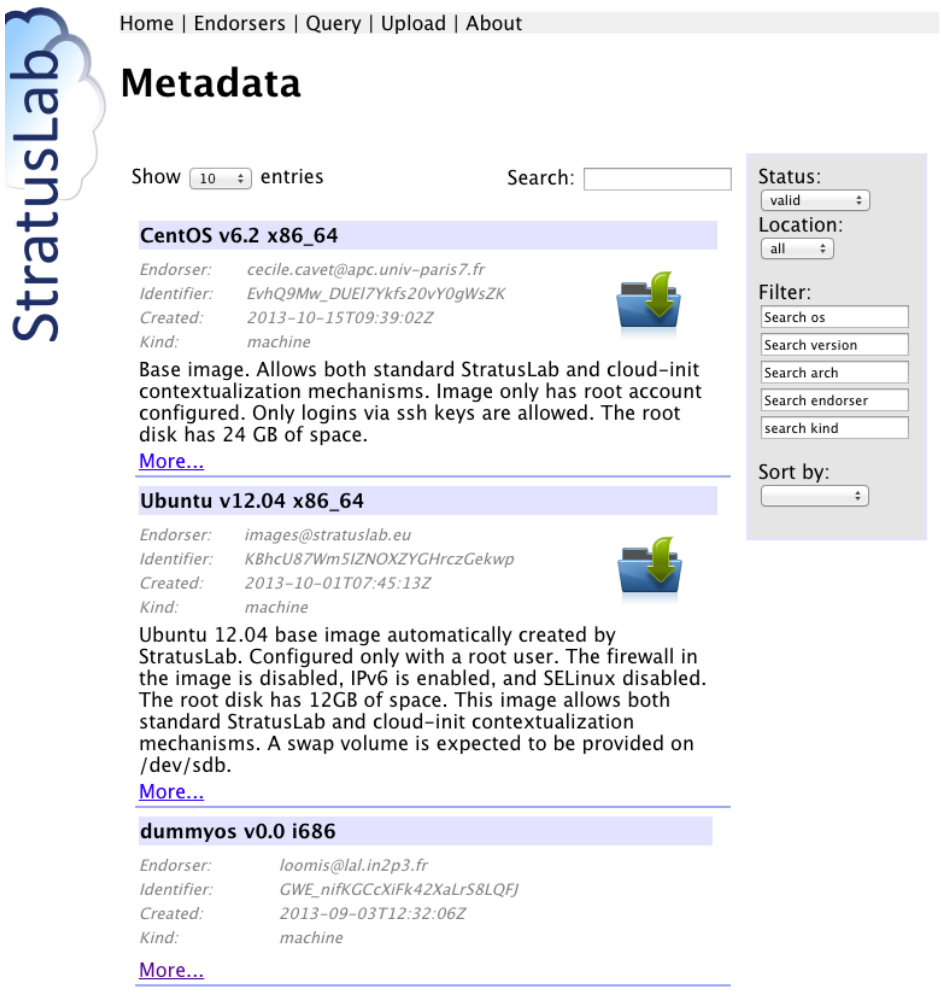


Figure: MarketPlace Website.

Cloud computing
What is it ?
Characteristics
Models
Solutions & ressources
StratusLab
Ressources
How to use a Cloud
MarketPlace
Monitoring
End-user client
Connexion
Persistent disk
Algorithm preservation in the Cloud
Requirements
Tools
Conclusion
References

MarketPlace: share Disk Image

- Disk image creation is time consuming:
  - Need to have access to a catalog of images.
  - Need to share disk images in project.
- To boot an image: identifier ➞ NYUUiZmum8Tj77kMhkxbPwxl5tA.
- Cloud dependent.
- Disk images have 6 months of validity ➞ OS update/upgrade for security.

Cloud computing
What is it ?
Characteristics
Models
Solutions & ressources
StratusLab
Ressources
How to use a Cloud
MarketPlace
Monitoring
End-user client
Connexion
Persistent disk
Algorithm preservation in the Cloud
Requirements
Tools
Conclusion
References

How to use the StratusLab Cloud

- Choice of the OS.
- Choice of the ressources.

Monitoring

Cloud computing
- What is it ?
- Characteristics
- Models
- Solutions & ressources

StratusLab
- Ressources
- How to use a Cloud
- MarketPlace
- Monitoring
- End-user client
- Connexion
- Persistent disk

Algorithm preservation in the Cloud
- Requirements
- Tools

Conclusion
- References

ID	NAME	RVM	TCPU	FCPU	ACPU	TMEM	FMEM	AMEM	STAT
20	onehost-31.lal.	7	3200	2896	0	63G	13.1G	3.6G	on
21	onehost-30.lal.	5	3200	3034	0	63G	9.2G	3.7G	on
23	onehost-15.lal.	6	2400	2396	400	35.4G	10.7G	3.1G	on
24	onehost-16.lal.	7	2400	2247	400	35.4G	16.1G	6.9G	on
25	onehost-17.lal.	6	2400	2396	600	35.4G	12.5G	9G	on
26	onehost-19.lal.	6	2400	2398	400	35.4G	28.4G	15.6G	on
27	onehost-20.lal.	4	2400	1299	0	35.4G	17.7G	18.3G	on
31	onehost-25.lal.	5	3200	2896	500	63G	17G	365.2M	on
32	onehost-26.lal.	6	3200	3197	1800	63G	23.5G	2.5G	on
33	onehost-27.lal.	6	3200	3095	0	63G	51.2G	2.5G	on
34	onehost-28.lal.	6	3200	3098	100	63G	56.5G	23.9G	on
36	onehost-21.lal.	6	2400	2396	1000	35.4G	11G	21.5G	on
37	onehost-14.lal.	7	2400	2396	1000	35.4G	19.1G	9.4G	on
38	onehost-23.lal.	3	2400	2398	600	35.4G	24.1G	77.9M	on
39	onehost-24.lal.	6	2400	1961	500	35.4G	9.1G	397.9M	on

Figure: Physical servers of the **StratusLab** Cloud: total and used resources (CPU and memory).

Cloud computing
What is it ?
Characteristics
Models
Solutions & ressources
StratusLab
Ressources
How to use a Cloud
MarketPlace
Monitoring
End-user client
Connexion
Persistent disk
Algorithm preservation in the Cloud
Requirements
Tools
Conclusion
References

How to use the StratusLab Cloud

- Choice of the OS.
- Choice of the ressources.
- Run VM.

End-user client: run Virtual Machine

Cloud computing
- What is it ?
- Characteristics
- Models
- Solutions & ressources

StratusLab
- Ressources
- How to use a Cloud
- MarketPlace
- Monitoring
- End-user client
- Connexion
- Persistent disk

Algorithm preservation in the Cloud
- Requirements
- Tools

Conclusion
- References

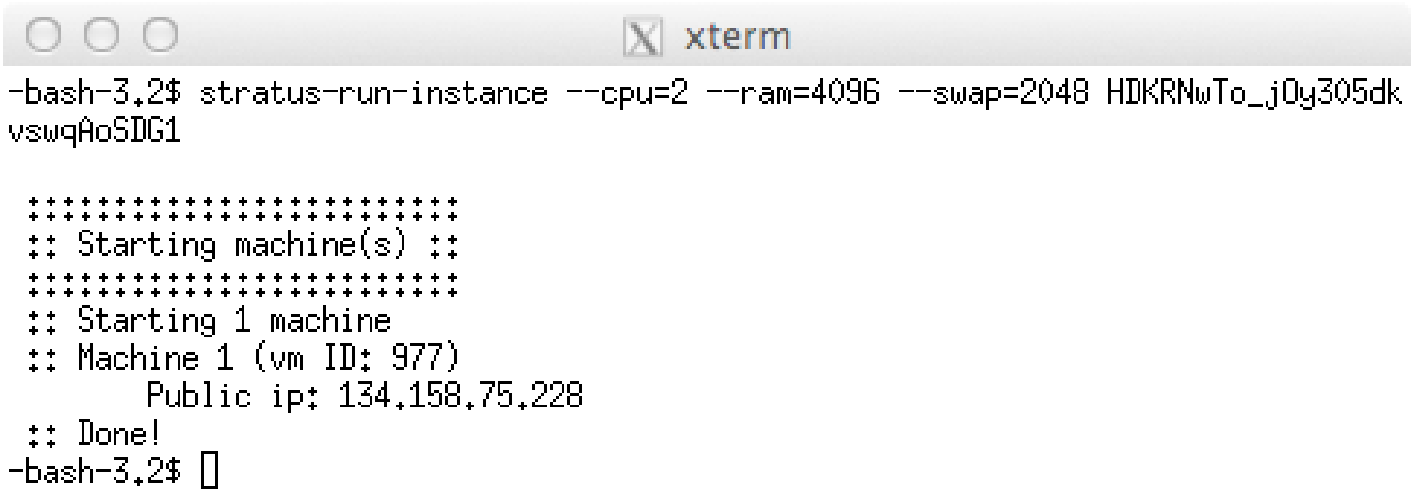


Figure: Command-line client: run a virtual machine.

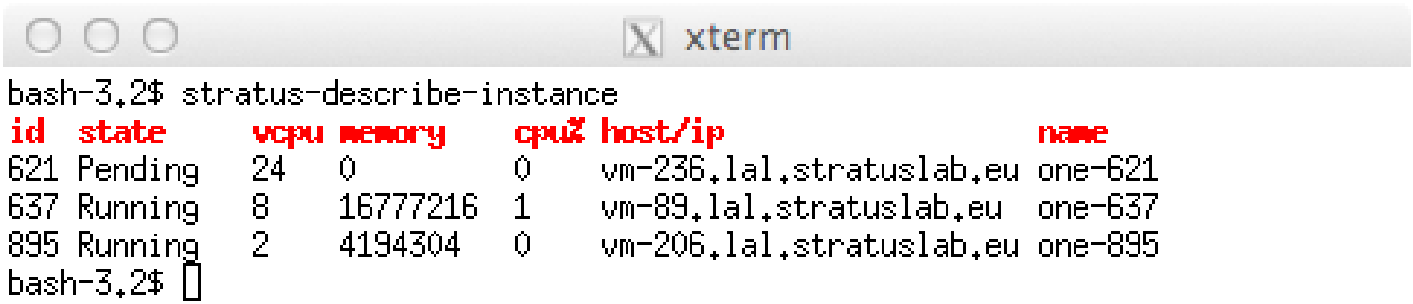


Figure: Command-line client: describe the state of virtual machines.

Cloud computing
What is it ?
Characteristics
Models
Solutions & ressources
StratusLab
Ressources
How to use a Cloud
MarketPlace
Monitoring
End-user client
Connexion
Persistent disk
Algorithm preservation in the Cloud
Requirements
Tools
Conclusion
References

How to use the StratusLab Cloud

- Choice of the OS.
- Choice of the ressources.
- Run VM.
- Connect to VM and run the application.

Cloud computing
What is it ?
Characteristics
Models
Solutions & ressources
StratusLab
Ressources
How to use a Cloud
MarketPlace
Monitoring
End-user client
Connexion
Persistent disk
Algorithm preservation in the Cloud
Requirements
Tools
Conclusion
References

Connexion

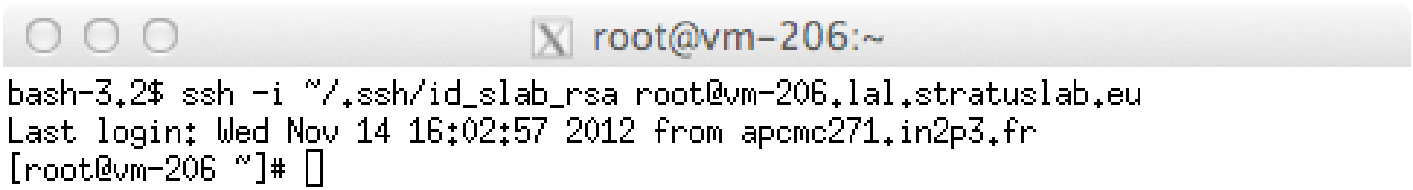


Figure: Connexion to the virtual machine: by SSH without password (key identification) and as the Root user.

- Virtual machine ready in 5min !
- Install all scientific/administration Linux packages as Root.
- Launch scientific applications as User.
- Run your jobs.

How to use the StratusLab Cloud

- Choice of the OS.
- Choice of the ressources.
- Run VM.
- Connect to VM and run the application.
- Save/Use the data.

Cloud computing
What is it ?
Characteristics
Models
Solutions & ressources
StratusLab
Ressources
How to use a Cloud
MarketPlace
Monitoring
End-user client
Connexion
Persistent disk
Algorithm preservation in the Cloud
Requirements
Tools
Conclusion
References

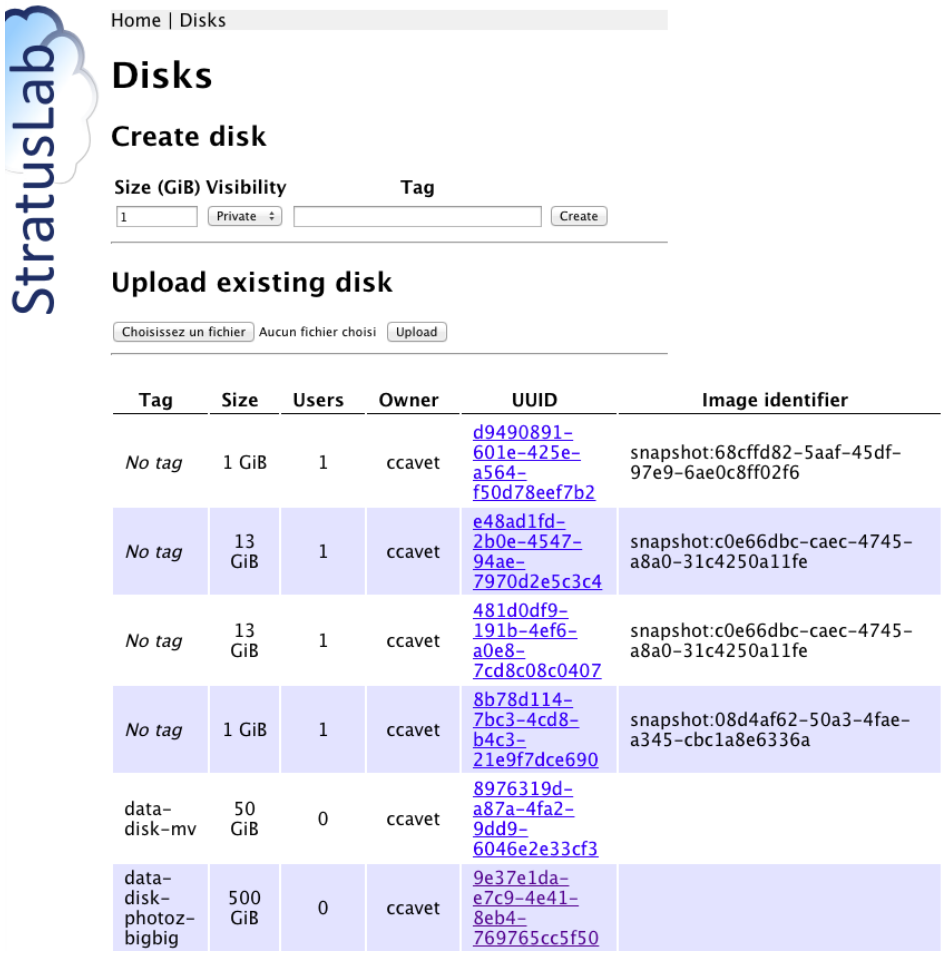
Persistent disk: manage volumes

Cloud computing
- What is it ?
- Characteristics
- Models
- Solutions & ressources

StratusLab
- Ressources
- How to use a Cloud
- MarketPlace
- Monitoring
- End-user client
- Connexion
- **Persistent disk**

Algorithm preservation in the Cloud
- Requirements
- Tools

Conclusion
- References



Tag	Size	Users	Owner	UUID	Image identifier
No tag	1 GiB	1	ccavet	d9490891-601e-425e-a564-f50d78eef7b2	snapshot:68cffd82-5aaf-45df-97e9-6ae0c8ff02f6
No tag	13 GiB	1	ccavet	e48ad1fd-2b0e-4547-94ae-7970d2e5c3c4	snapshot:c0e66dbc-caec-4745-a8a0-31c4250a11fe
No tag	13 GiB	1	ccavet	481d0df9-191b-4ef6-a0e8-7cd8c08c0407	snapshot:c0e66dbc-caec-4745-a8a0-31c4250a11fe
No tag	1 GiB	1	ccavet	8b78d114-7bc3-4cd8-b4c3-21e9f7dce690	snapshot:08d4af62-50a3-4fae-a345-cbc1a8e6336a
data-disk-mv	50 GiB	0	ccavet	8976319d-a87a-4fa2-9dd9-6046e2e33cf3	
data-disk-photoz-bigbig	500 GiB	0	ccavet	9e37e1da-e7c9-4e41-8eb4-769765cc5f50	

Figure: Persistent disk Web interface.

Cloud computing
What is it ?
Characteristics
Models
Solutions & ressources
StratusLab
Ressources
How to use a Cloud
MarketPlace
Monitoring
End-user client
Connexion
Persistent disk
Algorithm preservation in the Cloud
Requirements
Tools
Conclusion
References

Persistent disk: manage volumes

- Persistent storage (it survives to the VM).
- Not made for long term storage.
  - Not true for **commercial Cloud** (AWS, OVH...) but expensive and security issues.
  - Exploration of connection with non virtual storage for **academic Cloud** (OpenStack@CC-IN2P3).
- Cloud dependent:
  - If Cloud computing part (MV) is down/full = unreachable data.
  - Need to transfer data on each infrastructures.

Algorithm preservation in the Cloud: requirements

Cloud computing
What is it ?
Characteristics
Models
Solutions & ressources
StratusLab
Ressources
How to use a Cloud
MarketPlace
Monitoring
End-user client
Connexion
Persistent disk
Algorithm preservation in the Cloud
Requirements
Tools
Conclusion
References

Data preservation : challenge is also preserving software and algorithm.

We need to:

- be able to run the old software avoiding:
  - old computer dependency.
  - compiler errors, missing libraries...
  - re-writing of the code.
- be able to re-run analysis:
  - to re-obtain results.
  - by updating the code to compare new results with previous studies.

Tools from the Cloud

- Virtualization:
  - Encapsulation of the algorithms within the environnement (OS + packages + libraries).
  - Huge portability.
  - But... no long term studies (hypervisors will change, disk image standards will evolve...) ➙ Need to fix standards.
- Cloud infrastructure:
  - Running algorithms on-demand.
  - Huge flexibility.
  - But... solutions are always evolving (and connected tools also).

Cloud computing
What is it ?
Characteristics
Models
Solutions & ressources
StratusLab
Ressources
How to use a Cloud
MarketPlace
Monitoring
End-user client
Connexion
Persistent disk
Algorithm preservation in the Cloud
Requirements
Tools
Conclusion
References

Tools from the Cloud

Cloud computing
What is it ?
Characteristics
Models
Solutions & ressources
StratusLab
Ressources
How to use a Cloud
MarketPlace
Monitoring
End-user client
Connexion
Persistent disk
Algorithm preservation in the Cloud
Requirements
Tools
Conclusion
References

- MarketPlace:
  - Supermarket for disk images.
  - Ideal for sharing projects.
  - But... only 6 months of validity ➡ We can imagine a MarketPlace only for **long term preservation** of disque images (not for every day applications).
- Persistent disk:
  - Possibility to store data in the Cloud.
  - But... no long term preservation ➡ Specific storage in the Cloud ? Storage outside the Cloud ?

Cloud computing
What is it ?
Characteristics
Models
Solutions & ressources
StratusLab
Ressources
How to use a Cloud
MarketPlace
Monitoring
End-user client
Connexion
Persistent disk
Algorithm preservation in the Cloud
Requirements
Tools
Conclusion
References

Conclusion

- **New subject**: everything has to be done.
- Cloud is very adaptative to the algorithm preservation problem:
  - Easy and quick access.
  - Flexibility.
  - Based on virtualization.
  - Already developed tools (MarketPlace, Persistent disk...).
  - Future:
    - Growing infrastructures and community.
    - Federated Clouds.
    - Mutli-cloud solution.
    - Algorithm preservation ?

Thank you for your attention.

Cloud computing
- What is it ?
- Characteristics
- Models
- Solutions & ressources

StratusLab
- Ressources
- How to use a Cloud
- MarketPlace
- Monitoring
- End-user client
- Connexion
- Persistent disk

Algorithm preservation in the Cloud
- Requirements
- Tools

Conclusion

- References

Cloud computing
What is it ?
Characteristics
Models
Solutions & ressources
StratusLab
Ressources
How to use a Cloud
MarketPlace
Monitoring
End-user client
Connexion
Persistent disk
Algorithm preservation in the Cloud
Requirements
Tools
Conclusion
References

References

End-user client installation and full command description: FACe Wiki

- [1] NIST report: http://csrc.nist.gov/publications/nistpubs/800-145/SP800-145.pdf
- [2] Loomis, Présentation aux Journées Cloud de Paris Sud (2012) : `http://indico2.lal.in2p3.fr/indico/getFile.py/access?contribId=0&resId=0&materialId=slides&confId=1897`
- [3] Rapport sur le Cloud : `http://cordis.europa.eu/fp7/ict/ssai/docs/cloud-report-final.pdf`
- [4] Loomis et al., Chapître de « European Research Activities in Cloud Computing », Cambridge Scholars Publishing, p. 271 (hal-00676252) (2012)
- [5] Cavet et al. (2012): http://mesogrilles2012.sciencesconf.org/file/18139
- [6] Berriman et al., Phil. Trans. R. Soc. A, 371 (2013)