Architecture comparison of an in-situ data processing application applied to glacier mass balance analyse

Elisabeth RANISAVLJEVIĆ, Florent DEVIN,
Yannick LE NIR, Dominique LAFFLY

9th september 2014

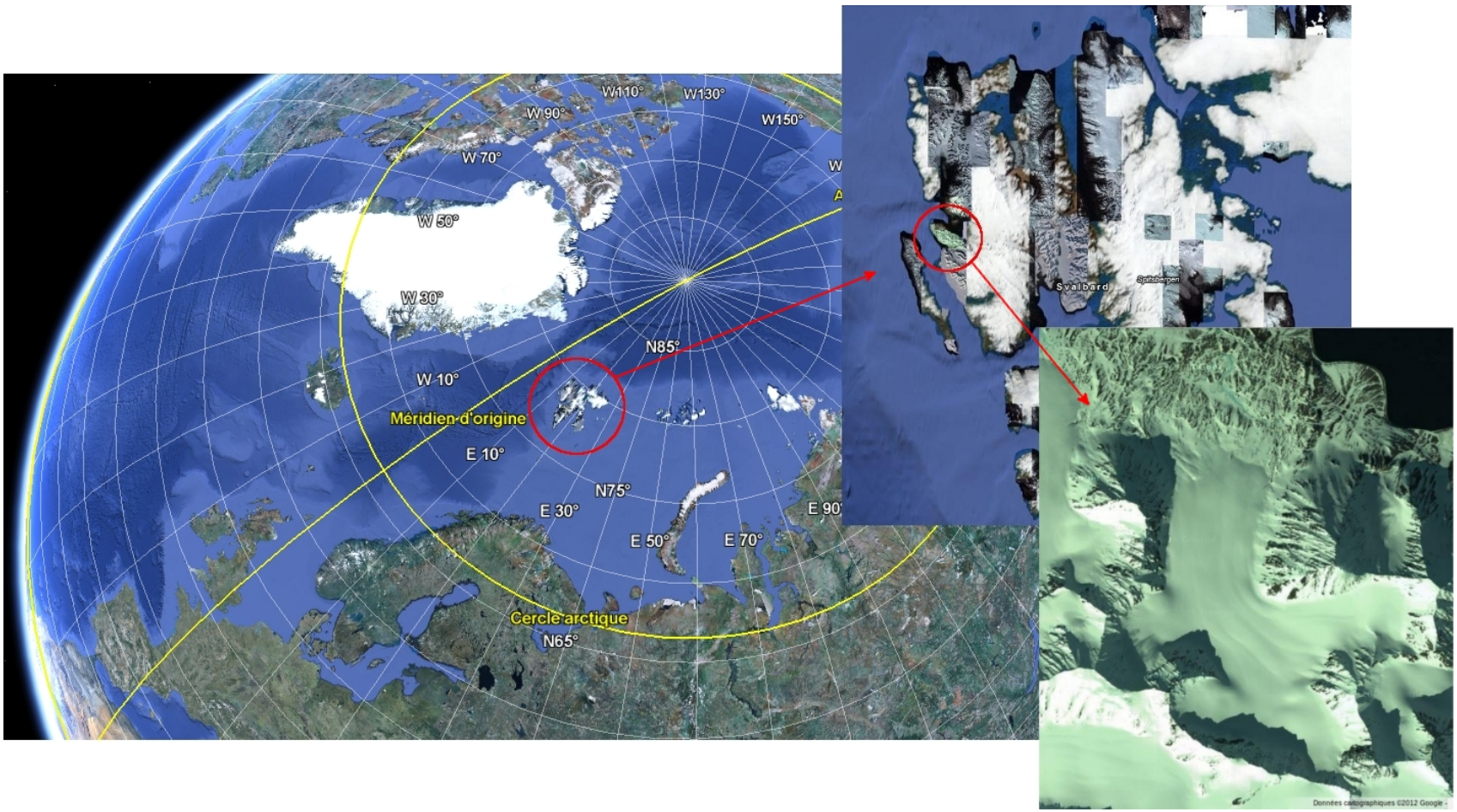
Introduction

- Analysing landscape require many and various data
- Process in order to collect information

Hydro Sensor FLOWS project (M. Griselin, C. Martin and D. Laffly)

- Map the temporal evolution of the snow cover
- Couple it with a hydrologic model
- East Loven glacier, Spitsberg

East Loven glacier



Snow cover and glacier evolution

- *In situ* sensing in order to complete remote sensing
- *In situ* constraints :
  - atmospheric disturbances
  - electronic deficiency
  - geometry variations

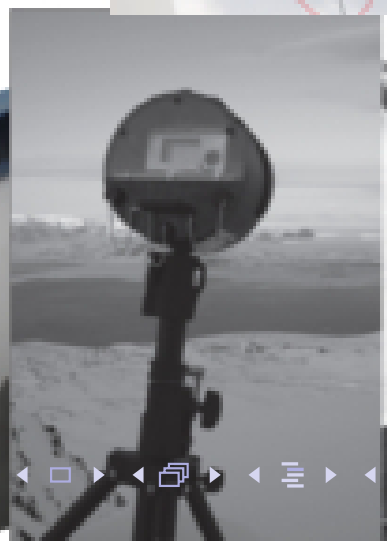


Image processing

- Many data to process (over 30 000)
- Heavy and complex tasks
  - classification
  - projection
  - reconstruction of the satelite view
  - ...
- Variation of the workflow

- How easily process all the data with all the specificities ?
- What is the advantage of a cloud architecture ?

Cloud Computing

Set of resources, servers and applications, offered *"as a service"* over a network

Advantages :

- Easiness of access
- Large storage capacity
- Lightness of application
- Modular : add, remove, modify services
- Scalable
  - Increase of users connected to the service
  - Increase the computing capacity according to the needs
  - Fault tolerance

Cloud Computing

Deployment models

Private cloud

- provider $=$ client
- part of the workstation dedicated for the cloud, hosting services and data
- control of the machines and data

Public cloud

- provider $\neq$ client
- client get an access
- no information on where the data is stored, where the services are, and who else is using the cloud

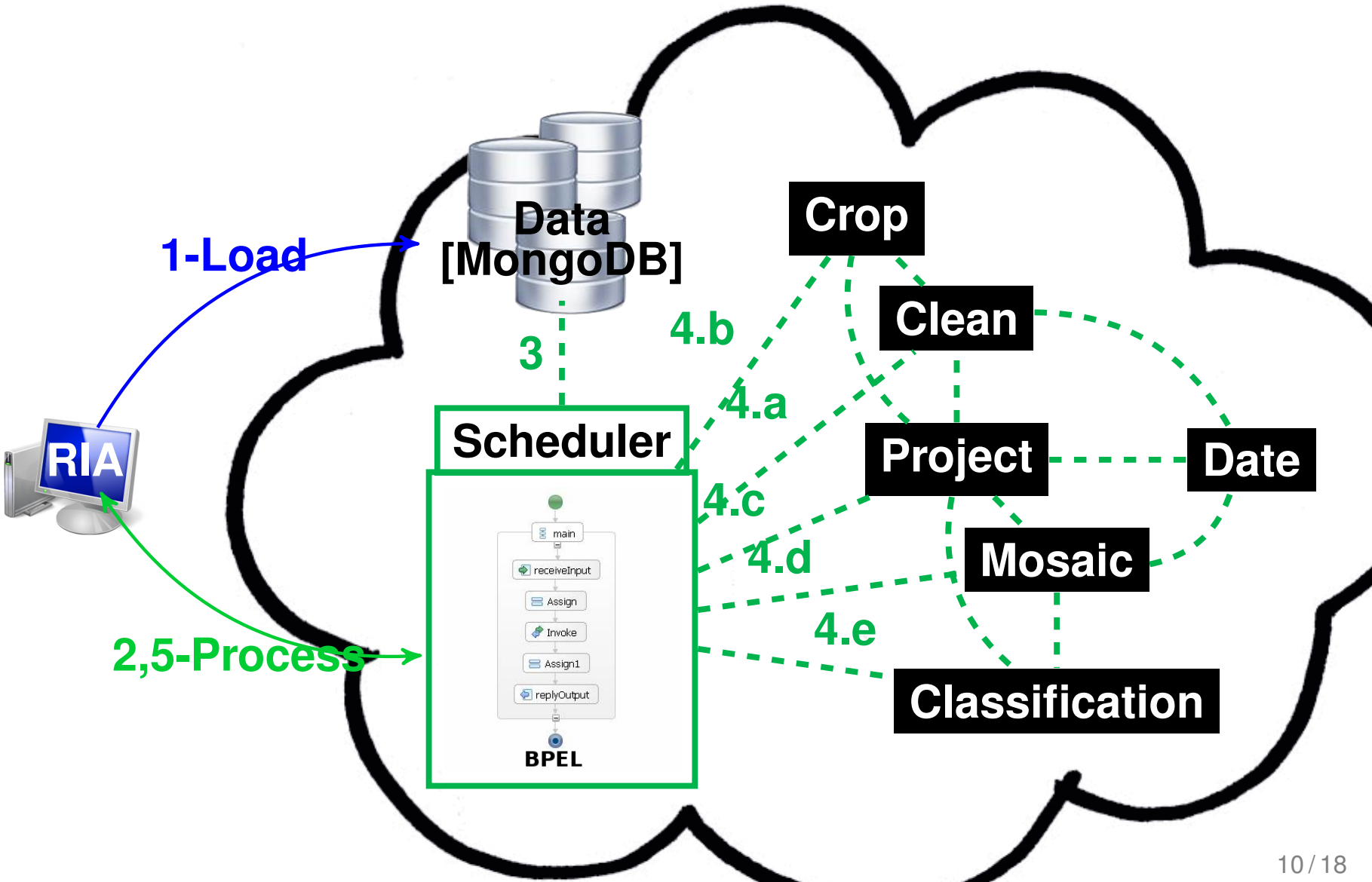
Hybrid cloud

- composition of clouds

Image processing

- Split the process into several tasks
  - Classification, Project, Clean, ...
  - Web services
- Schedule a workflow to fully process the image
  - modify the order of services to generate new workflows

Model


1-Load
Data
[MongoDB]
Crop
Clean
4.b
3
4.a
Scheduler
RIA
Project
Date
4.c
4.d
Mosaic
4.e
2,5-Process
main
receiveInput
Assign
Invoke
Assign1
replyOutput
BPEL
Classification

Architectures

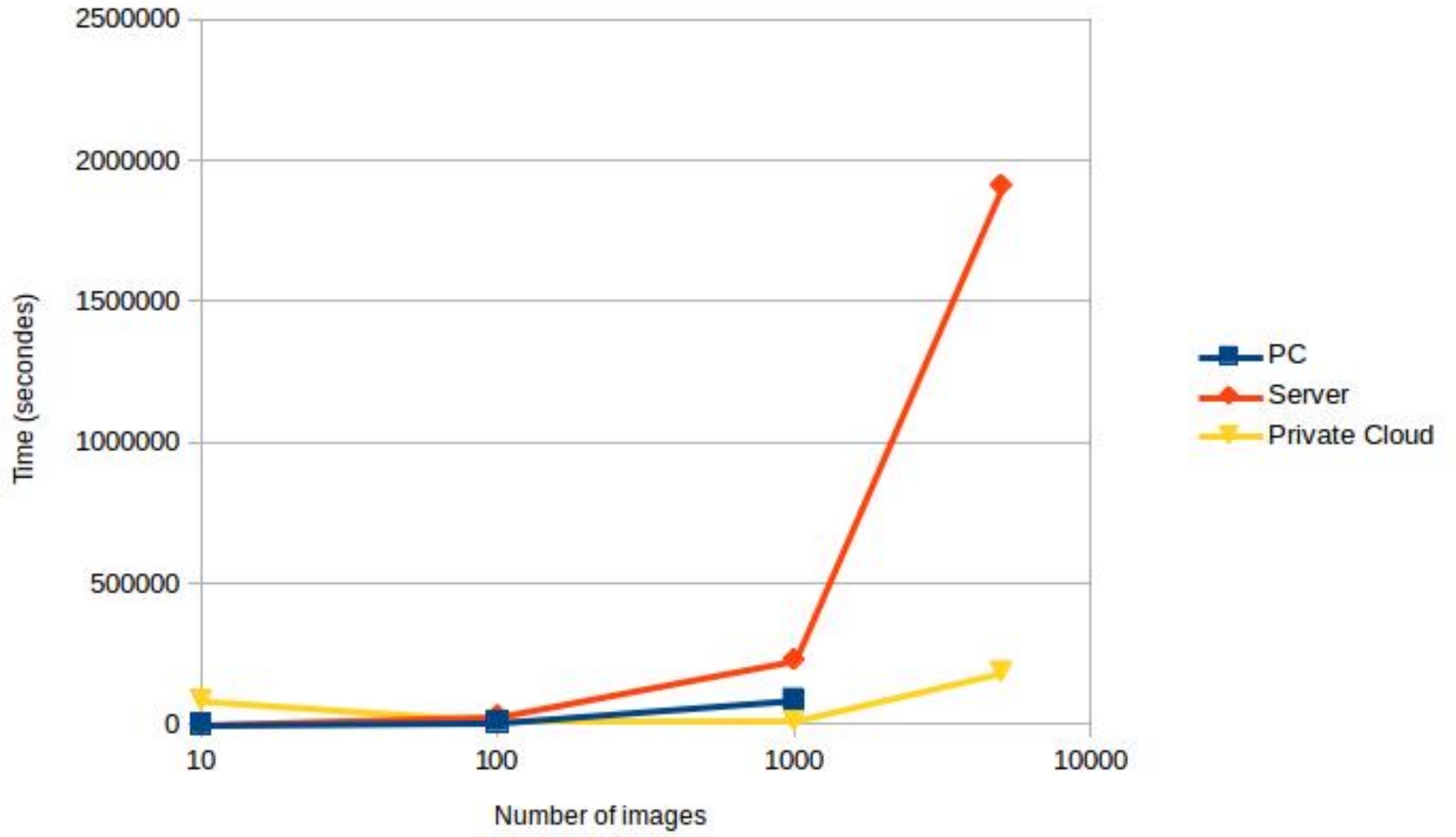
- Personal computer
  - 4 CPU @ 2.50 GHz, 8 GB RAM
- Server
  - 8 CPU @ 2.00 GHz, 4 GB RAM
- Private cloud
  - 40 machines, not dedicated
  - 4 CPU @ 3.10 GHz, 2 GB RAM

Experiment

- Parameters
  - number of images : 10, 100, 1000, 5000
  - architectures : personal computer, server, private cloud
  - workflow : vary the number of services called
- Measures
  - time
  - CPU usage (average)
  - bandwidth usage (average)
  - memory usage (average)

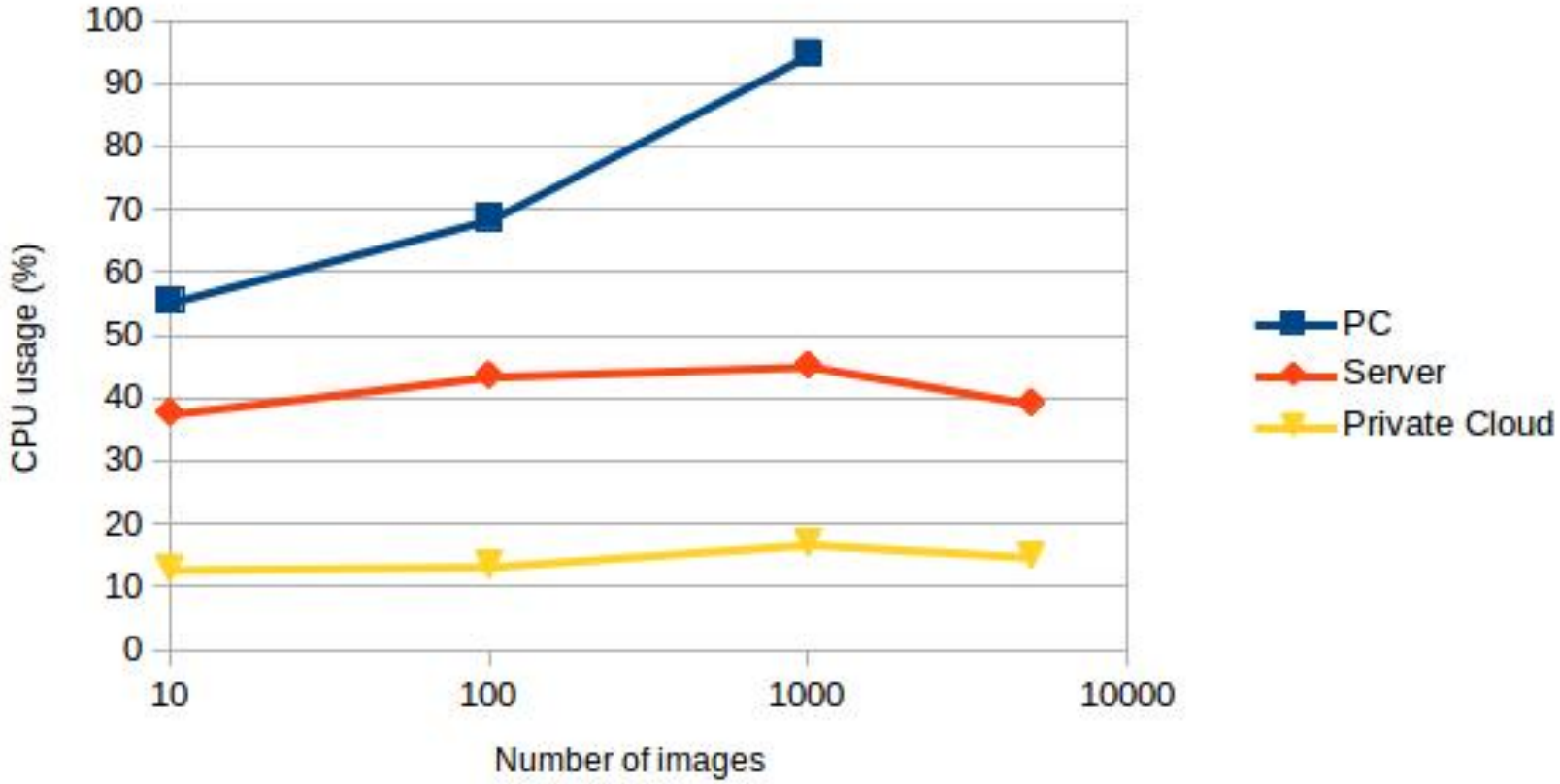
Results

Time



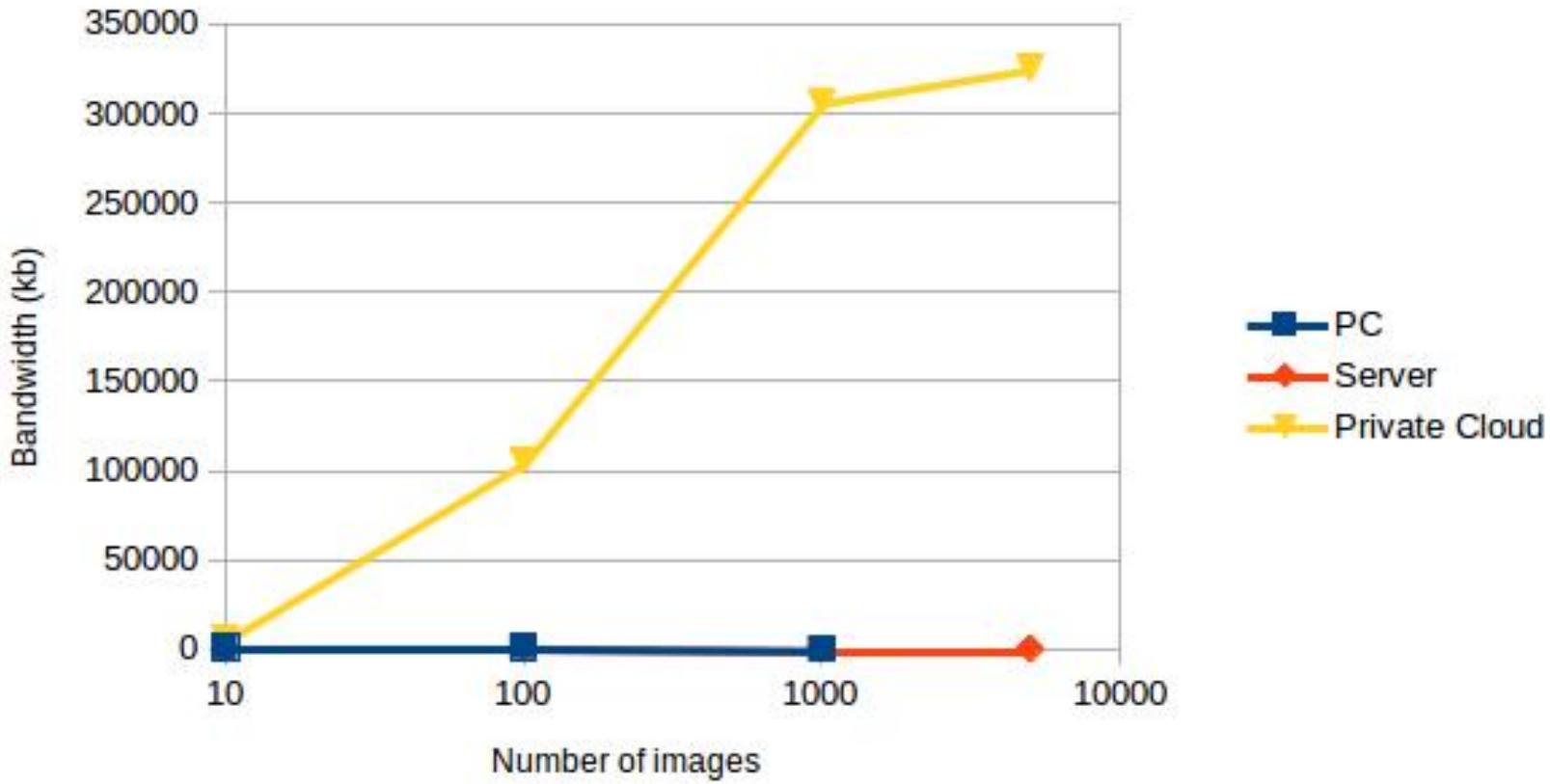
Results

CPU usage (average)



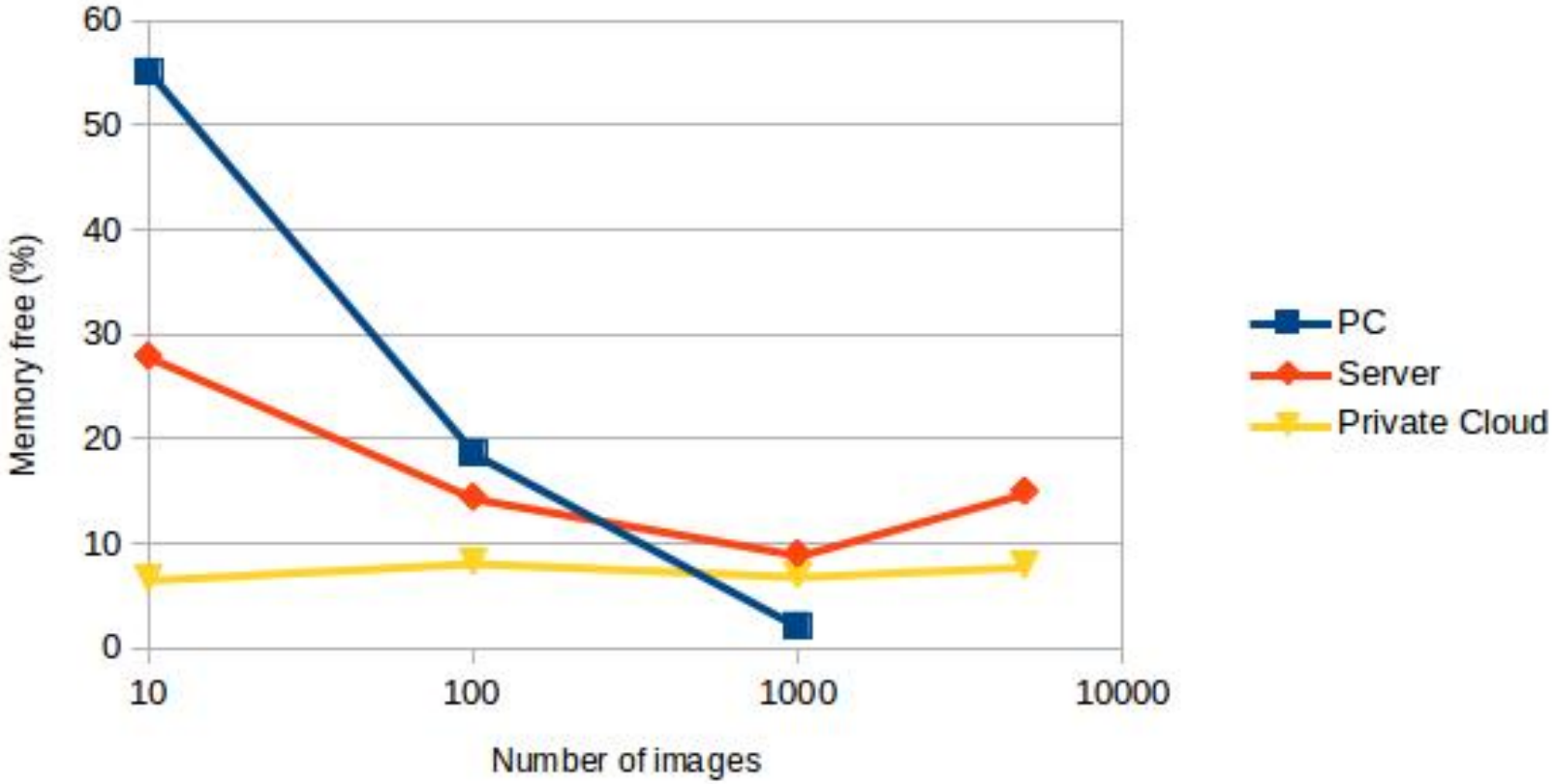
Results

Bandwidth (average)



Results

Memory usage (average)



Conclusion

- Public cloud :
  - not limited by the number of machines
  - cost
- Bring out a metric
  - choose the best architecture according to
    - inputs (number of images, ...)
    - constraints (time, price, ...)

Thank you