HPC vs. Cloud Mindset

Peter Kandolf

Engineering Kiosk Alps - Meetup Innsbruck

2024-10-24

About the author and motivation

Author

- PhD in Numerical Linear Algebra
- HPC experience for Scientific Computing
- Work, research and a pinch of administration experience with HPC and Cloud services

About the author and motivation

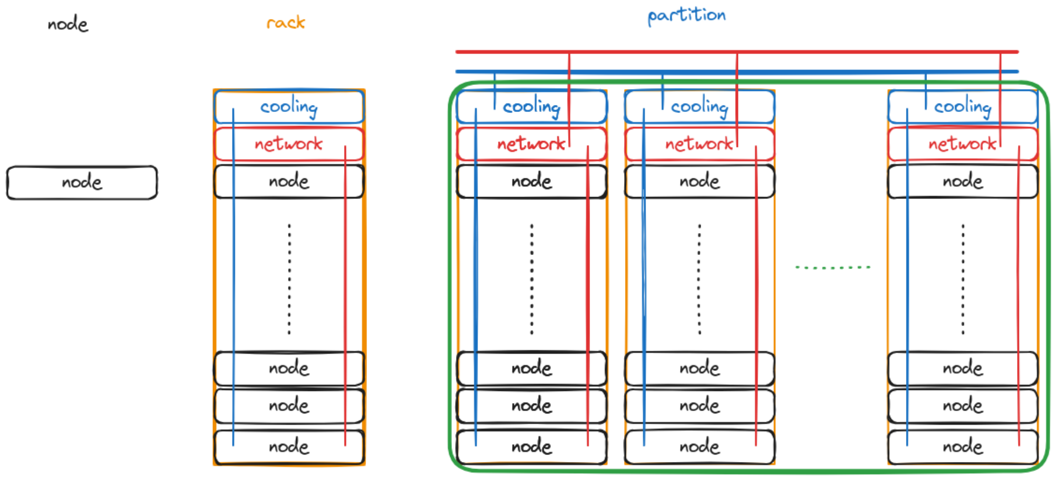
Author

- PhD in Numerical Linear Algebra
- HPC experience for Scientific Computing
- Work, research and a pinch of administration experience with HPC and Cloud services

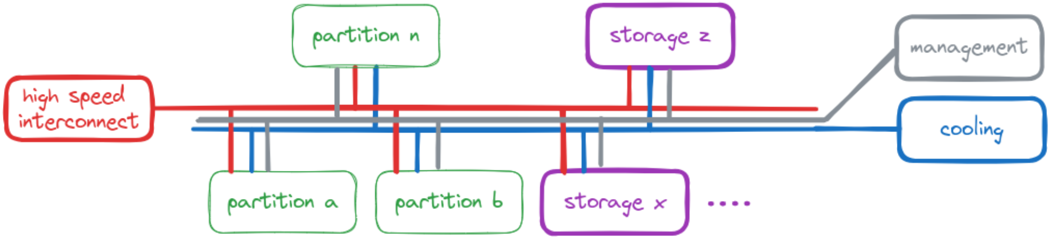
Motivation

- Co-lead in the Austrian DataLAB and Services Project
  - Project to foster collaborative approaches between 8 Universities
  - Increase usability, lowering the learning curve
  - Bring in cloud services
- I found myself between *HPC* and *Cloud data center* admins and had to learn how to moderate discussions

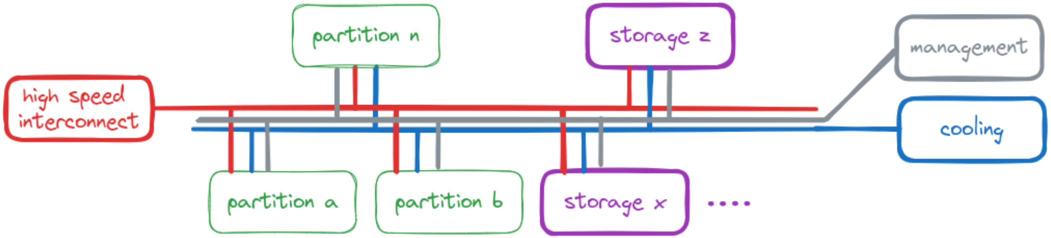
Basics - from node to partition



Basics - from partition to data center



Basics - from partition to data center

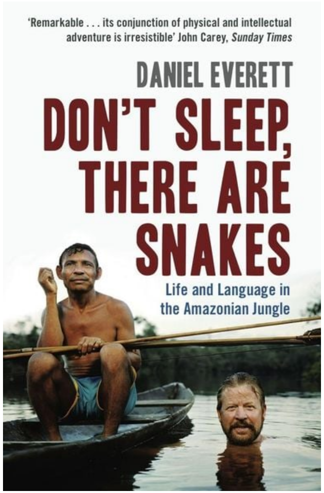


Is this HPC or Cloud infrastructure?

Approach for this talk

> *I have removed most of the technical linguistic details so that it flows better. Talking to people from very different cultures, as this story shows, involves much more than merely getting the word meanings right. One can translate every word well and still have a hard time understanding the story. This is because our stories include unstated assumptions about the world that are made by our culture.*

Daniel L. Everett: Don't sleep, there are snakes


'Remarkable . . . its conjunction of physical and intellectual adventure is irresistible' John Carey, Sunday Times
DANIEL EVERETT
DON'T SLEEP, THERE ARE SNAKES
Life and Language in the Amazonian Jungle

Purpose of the system

HPC

Squeeze out the most FLOPS i.e. most performance[a] for an application that knowns and cares about the cluster architecture. The focus is on providing as much raw computing power as needed and to not stand in the way to utilize it.

- performance first - hunger for FLOPS

[a]Performance measures is a tricky business and FLOPS alone are not always the optimal measure.

Purpose of the system

HPC

Squeeze out the most FLOPS i.e. most performance[a] for an application that knowns and cares about the cluster architecture. The focus is on providing as much raw computing power as needed and to not stand in the way to utilize it.

- performance first - hunger for FLOPS

[a] Performance measures is a tricky business and FLOPS alone are not always the optimal measure.

Cloud

Squeeze out the most applications i.e. most performance[a] where each application hardly cares about the (abstracted) hardware. The focus is on keeping the applications running and separate so nobody influences each other or even knows about anybody else.

- security first - hunger for abstraction

[a] Performance measures is a tricky business and FLOPS alone are not always the optimal measure.

(Admin) Mantra - Users should

HPC

- worry about:
  - CPU architecture and the toolchain,
  - which CPU is closer to the GPU and NUMA
  - where the other nodes in the job are,
  - software availability,
  - (compile flags)
- not worry about:
  - access: network, nodes, data, (security)

(Admin) Mantra - Users should

HPC

- worry about:
    - CPU architecture and the toolchain,
    - which CPU is closer to the GPU and NUMA
    - where the other nodes in the job are,
    - software availability,
    - (compile flags)
- not worry about:
    - access: network, nodes, data, (security)

Cloud

- worry about:
    - secure service interconnection,
    - secure authenticate,
    - automation,
    - state,
    - security settings
    - (yaml syntax)
- not worry about:
    - architecture: physical location, compute architecture, (performance)

Assumptions - User

HPC

- You need to apply to qualify
- We **trust** you:
  - respect the job scheduler
  - install programs in user space
  - move between nodes on OS level
  - you can and should max out CPU/IO if it helps
  - be British (queue like a pro)
- You can/should not:
  - run something for a long time
  - assume your application restarts
  - interfere with other users on purpose

Assumptions - User

HPC

- You need to apply to qualify
- We **trust** you:
  - respect the job scheduler
  - install programs in user space
  - move between nodes on OS level
  - you can and should max out CPU/IO if it helps
  - be British (queue like a pro)
- You can/should not:
  - run something for a long time
  - assume your application restarts
  - interfere with other users on purpose

Cloud

- Just create an account
- We **never** trust you:
  - you get restricted access to our API
  - you have no access to any hardware
  - you can *deploy* an application
  - don't use more as assigned or we kill you(r application)
  - we restart your application if it dies
- You can/should not:
  - assume a resource belongs to you alone
  - assume your applications stays on one server

Assumptions - Run your application 2

HPC

You get the best performance by adapting your code for the compute environment. You describe the resources you need and run your code, but you don't care if it runs somewhere else.

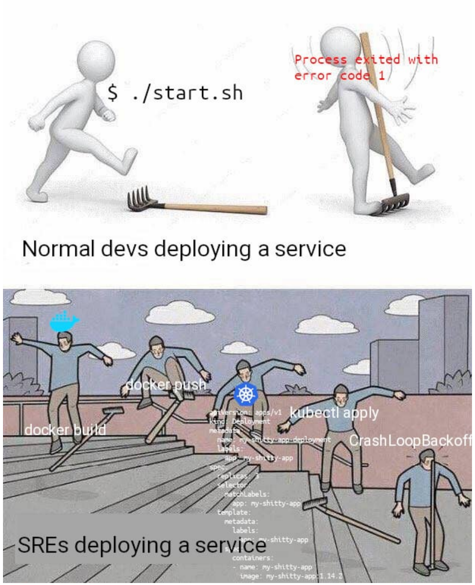
- you know your hardware

Cloud

You get the best performance by making everything as abstract and independent as possible. You describe the state you want to achieve, but you don't care how it is achieved.

- everything is a file

Assumptions - Run an application



```
apiVersion: apps/v1
kind: Deployment
metadata:
  name: my-shitty-app-deployment
  labels:
    app: my-shitty-app
spec:
  replicas: 3
  selector:
    matchLabels:
      app: my-shitty-app
  template:
    metadata:
      labels:
        app: my-shitty-app
    spec:
      containers:
      - name: my-shitty-app
        image: my-shitty-app:1.14.2
```

Figure 1

Assumptions - Scaling

HPC

Scaling is going from one thread to multiple threads, from one node to multiple nodes, from one tier to the next.

Cloud

Scaling is to provide the service for a hand full of accesses, for millions of accesses, distributing your service over the world.

Fin

Thank you!

Do you have any questions?