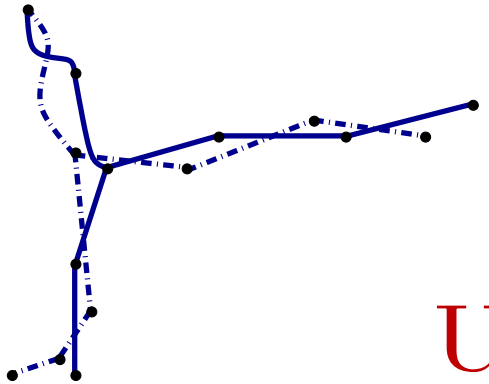
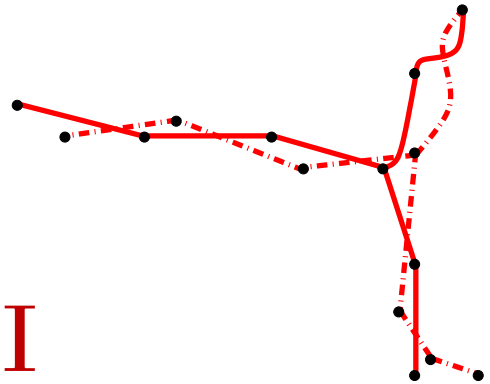


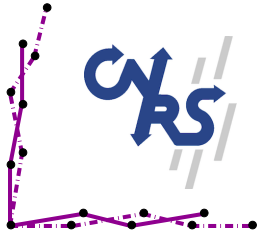


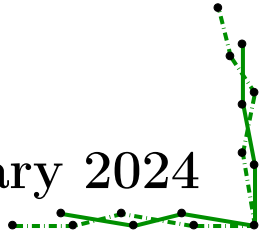
UE projet : OpsCI

Operations and Continuous Integration

https://www-npa.lip6.fr/~buixuan/opsci2023

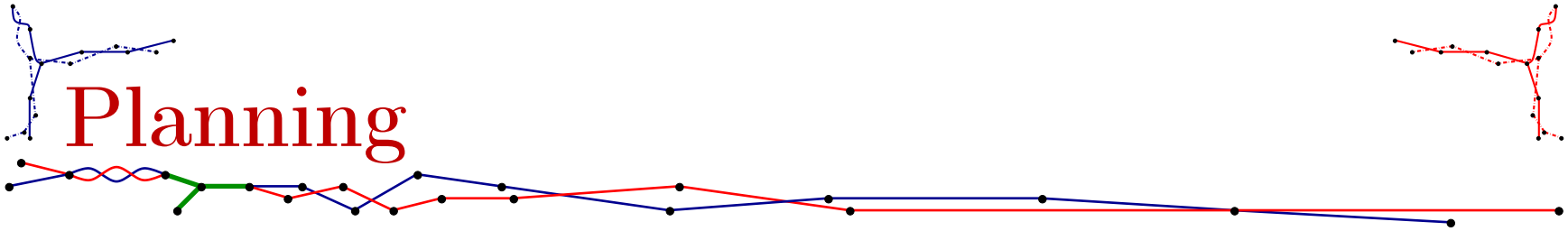
Binh-Minh Bui-Xuan





PARIS, January 2024

Planning

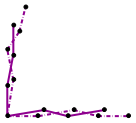


Teaching team :

- Katia Amichi
- Binh-Minh Bui-Xuan
- Alfred Deivassayagame
- Arthur Escriou

Assessment :

- session 1 = projects (3 total) + written exam
- session 2 = written exam



Highlights

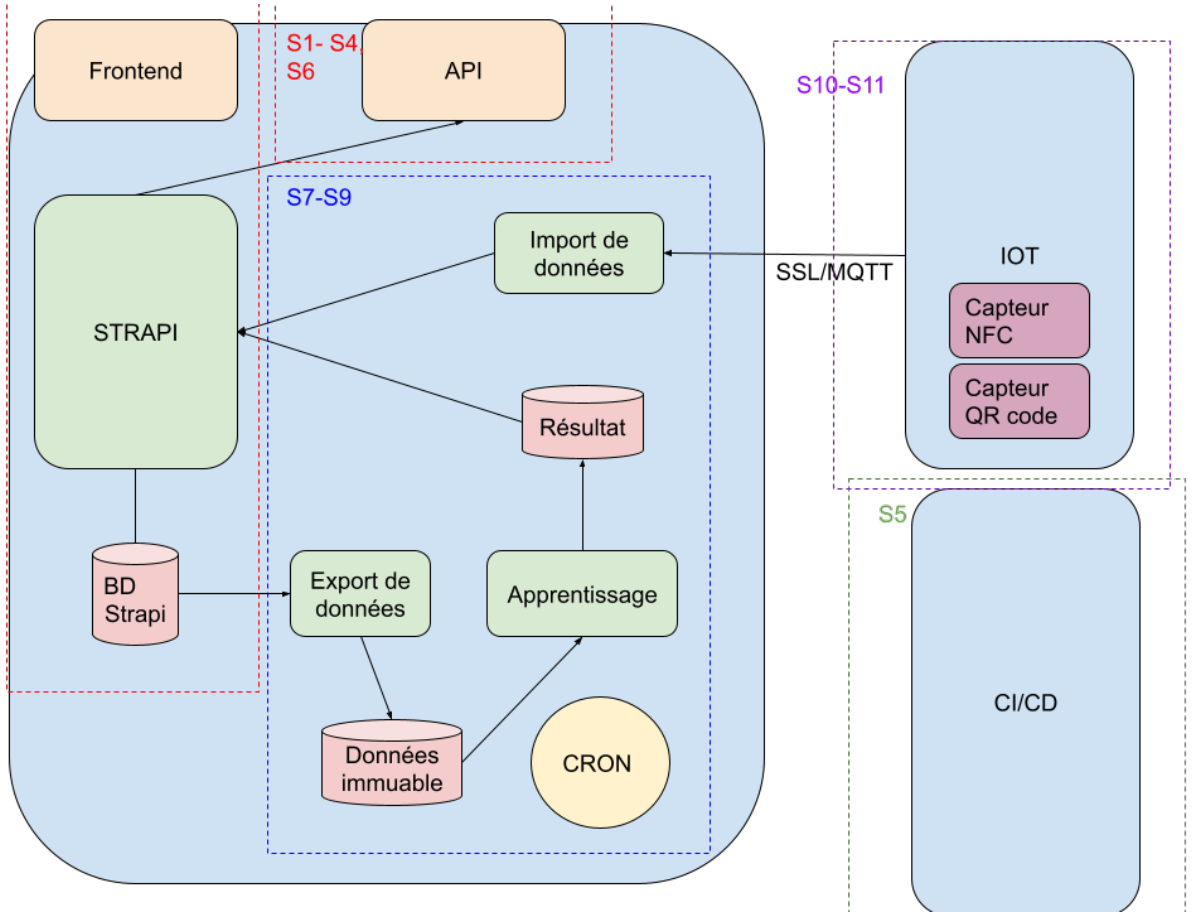
Ops & CI/CD :

- Ops : DevOps, DataOps, MLOps, DevSecOps, ...
- CI/CD : code, build, test, deploy, monitoring, ...

Project teams :

- lectures 1-6 : DevOps project
- lectures 7-9 : DataOps project
- lectures 10-11 : IoT project

Architecture overview



Why ?

Professional choices :

- Ingénieur DevOps, Intégrateur
- Data Engineer
- Ingénieur SRE, Ingénieur Système & Réseaux, DSI

Study choices :

- M1 : MU4IN019 CRV (Cloud et Réseaux Virtuels)
- M2 : MU5IN552 DAAR (Moteurs de Recherche)

Scientific choices :

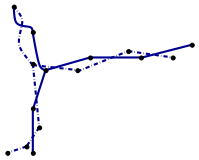
- EDITE
- ANR/ERC PI & partners'
- CIFRE

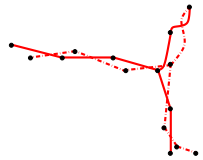
Today's work

1. Local Infrastructure

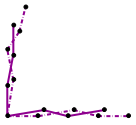
2. LAN/WAN lookups

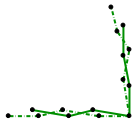
3. Scheduler



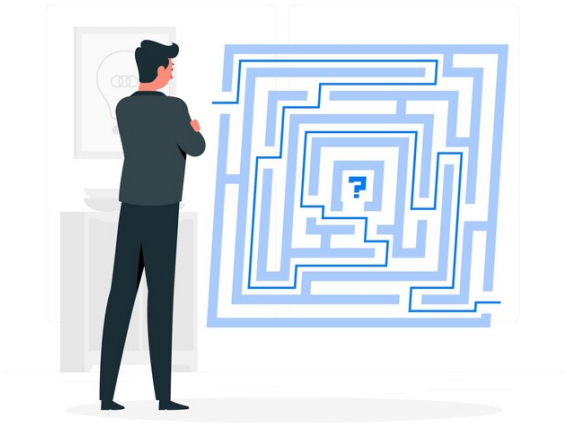


Part I : Local Infrastructure



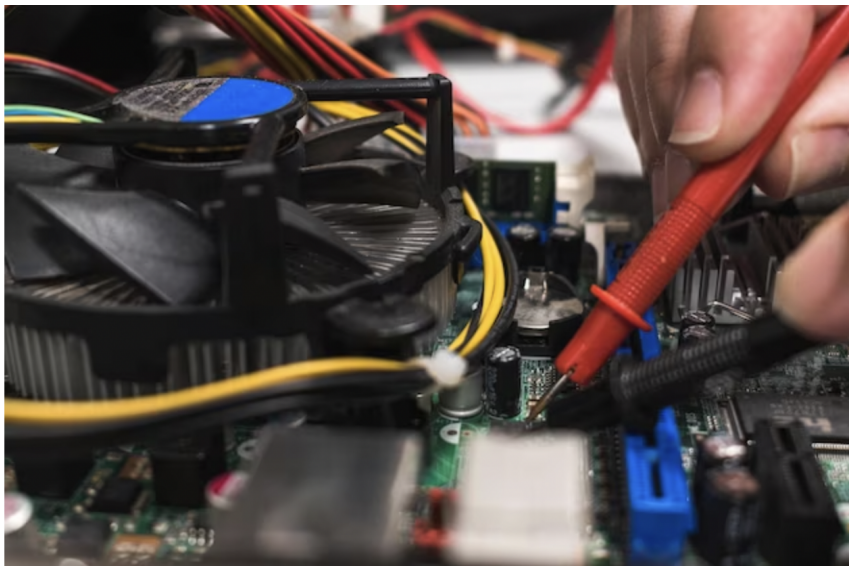


Linux basics



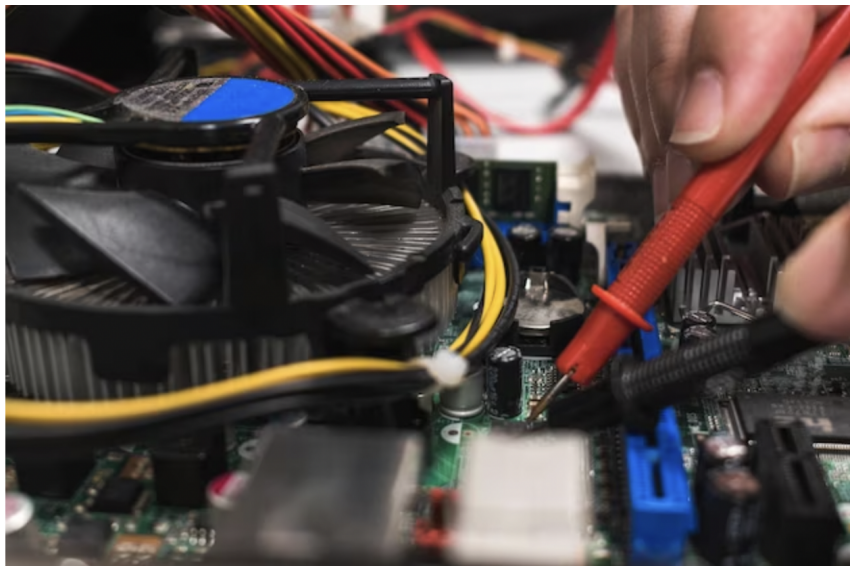
- file system : directory ? user rights ? quota ?
- lookups : filename vs. file contents ; RegEx
- threads : suspend, background, pid listing, kill/stop

Exercise : localhost



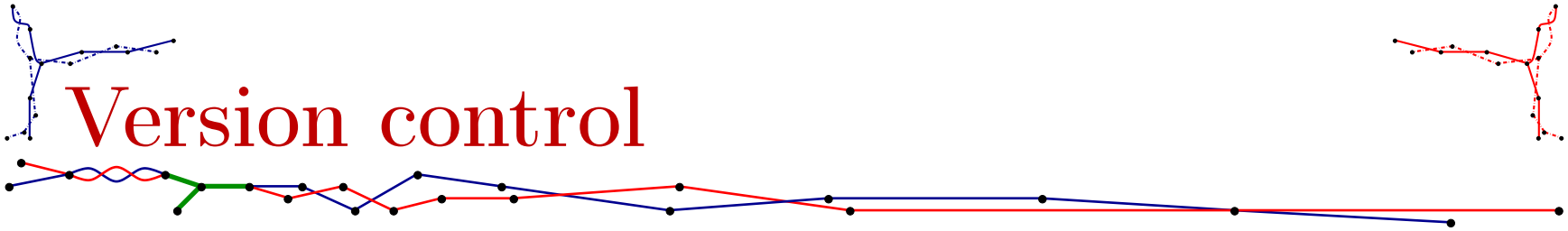
– Exercise : report characteristics of one's machine.

Exercise : localhost



– Exercise : report characteristics of one's machine.

⇒ disks ? devices ? cpu ? gpu ? mem ? vendor ? (!)



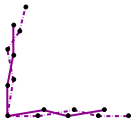
Version control

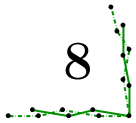
Pros :

- coordination, sharing, collaboration, ...
- XP
- human factor

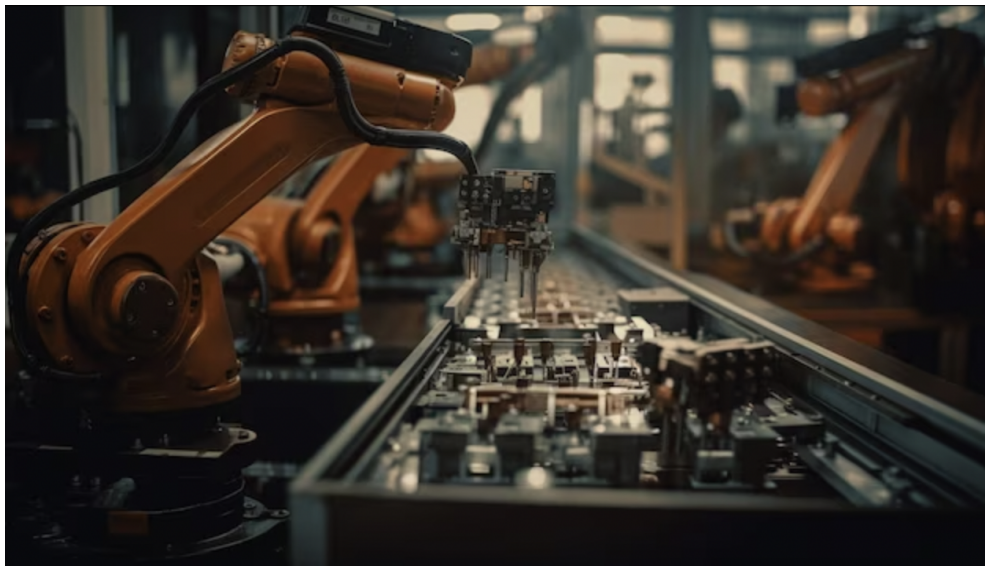
Cons :

- version conficts ⇒ risk of data loss
- sensitive information exposure
- needs Craft (“Pros”, not “Cons” ?)





Automation

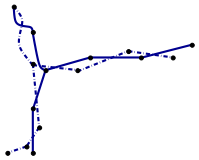
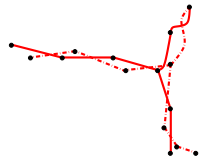


– clicks vs. scripts
– cornerstone for data jobs
– (!) spoiler (!) increases your M2 grade

Recap

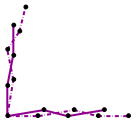
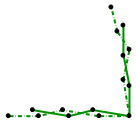
What to do with an isolated machine :

- navigate with file systems
- check devices
- scripts ⇒ background jobs

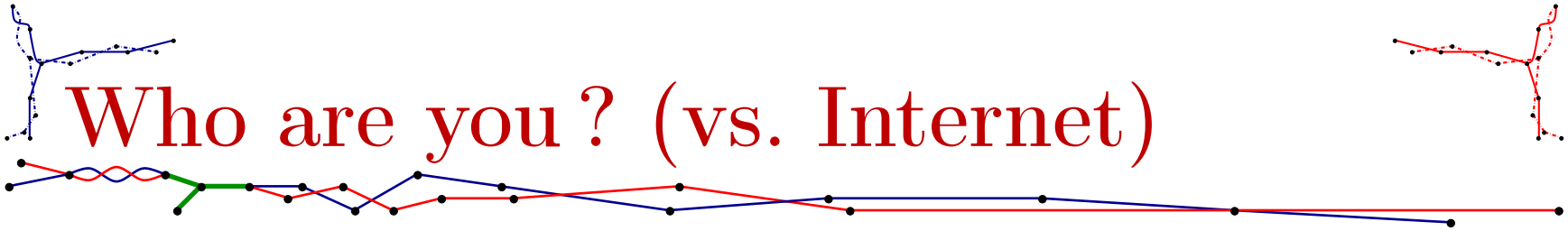
Part II : Local/Wide Area Networks

(LAN/WAN) lookups

Who am I ? (not only localhost)

– 06/07 ? ⇒ IPv4, IPv6 !



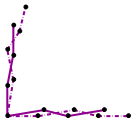
Who are you ? (vs. Internet)

Pros :

- measurement, cartography
- deep web
- debug

Cons :

- scan vs. attacks
- scrapping policies



Are we bloated ? (vs. bandwidth)

– unnecessary trafic → $. Ex : datadog
– vivre heureux, vivre caché ?

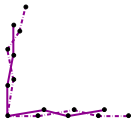
ssh, console, best practices recap

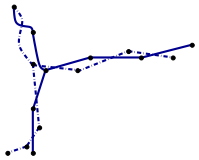
Dos :

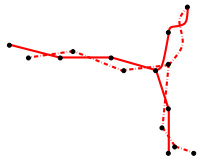
- ssh key login ; add passphrase (encrypt even the ssh private key)
- think text
- save logs/scripts/redirects
- log out (may require to move processes to background)

Don'ts :

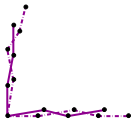
- ssh password login
- expect X11, bitmap images, videostreaming, ...
- leave sessions unattended

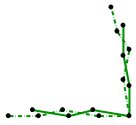


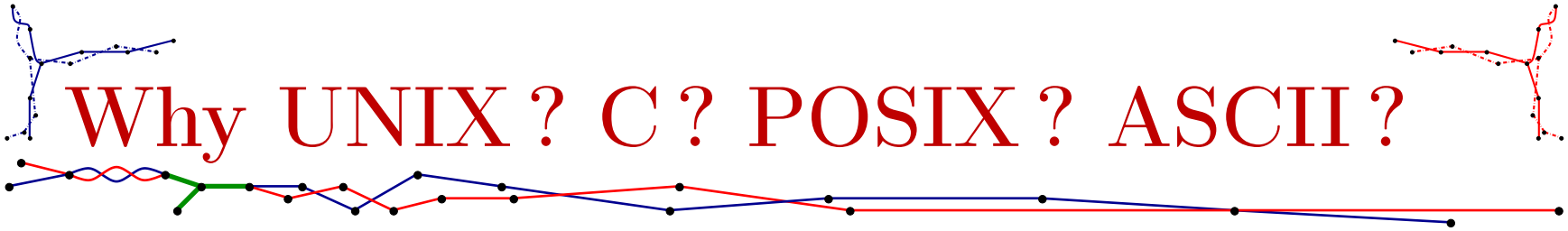




Part III : Scheduler

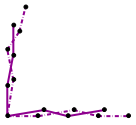






Why UNIX ? C ? POSIX ? ASCII ?

– UNIX ≈ BSD, Linux, Darwin, ...
– think machine = think in C
– text pipelining : local, URI, commands, ...
– RegEx, P-tries, SEO, other pragmatical matters ⇒ ASCII



Simple UNIX scheduler : CRON

```
# Example of job definition:
# .---------------- minute (0 - 59)
# |  .------------- hour (0 - 23)
# |  |  .---------- day of month (1 - 31)
# |  |  |  .------- month (1 - 12) OR jan,feb,mar,apr ...
# |  |  |  |  .---- day of week (0 - 6) OR sun,mon,tue,wed,thu,fri,sat
# |  |  |  |  |
# *  *  *  *  *  user command to be executed
```

Conclusion

- Infra : UNIX is good for you (when Ops & CI).
- LAN/WAN lookups : fast & lightweight uses with console.
- Scheduler : CRON ok ; other tools exist.

⇒ enjoy practical works !