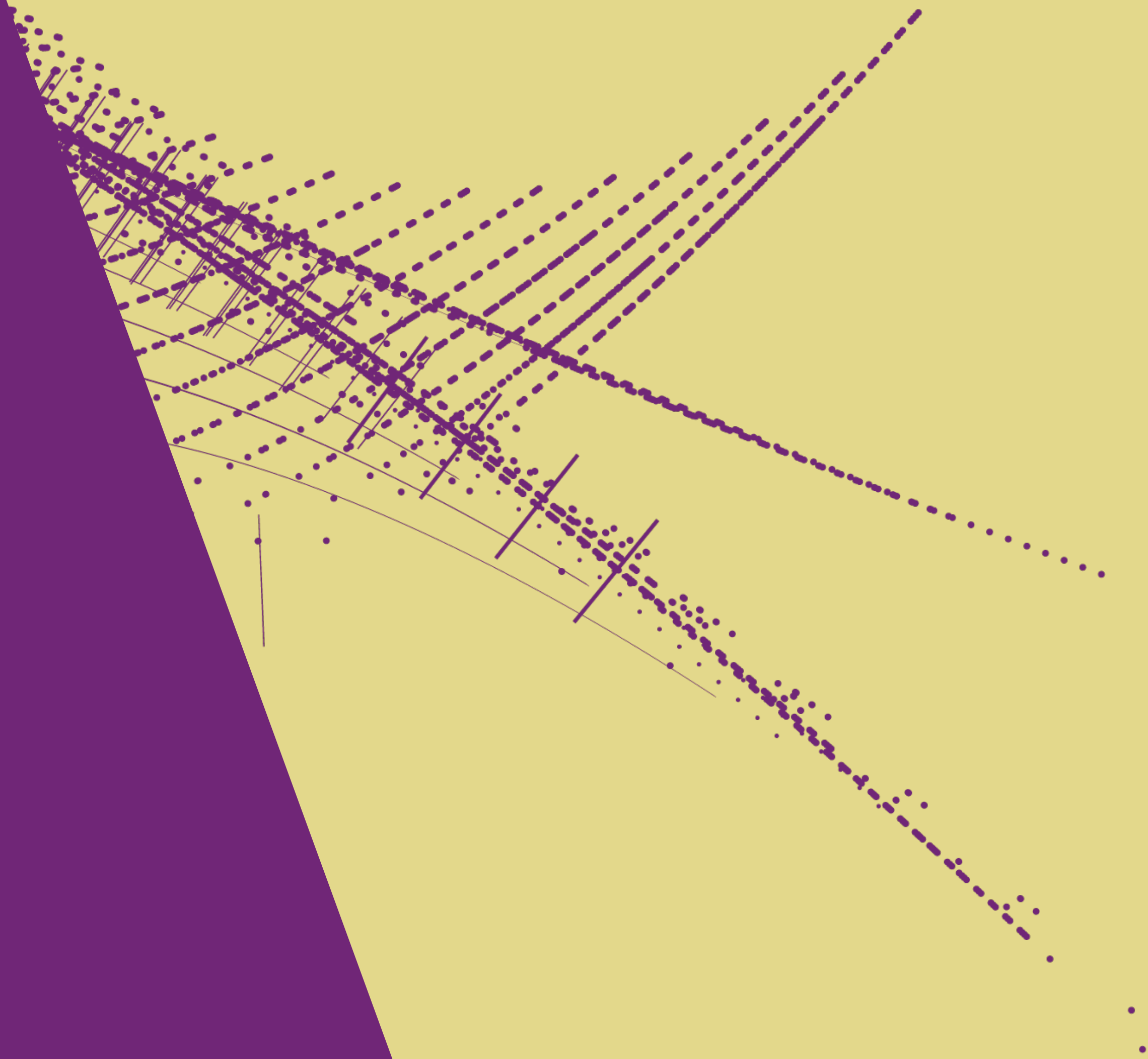# Nikhef

COMPUTING COURSE
INTRODUCTION

CONTEXT

Working as Nikhef scientist; (possibly) new here.

- Email / Mobile / Laptop / Desktop(?)
- Storage & Backup
- Authentication
- Teamwork / Project Management
- Large Scale Storage (no backup)
- Repositories for Data and Software*†
- Research Computing (local)
- Research Computing (distributed / collaboration)

"CT"

"PDP"

DATA AND SOFTWARE

"Data Stewardship"

Archive "your data"

- Choices on what to archive and where
- may not be practical to archive everything! References?
- what can you easily regenerate (MC code + versions + input file)

Archive your analysis

- Code is what you did, maybe not what you think you did
- Dependecies on other code (eg numpy): record versions too!

FAIR

Findable, Accessible, Interoperable, Reusable

Nikhef

DATA AND SOFTWARE

Intellectual Honesty

"Scientist even have formal code of conduct to promote intellectual honesty. They call it "scientific method". You are supposed to look critically at your results, pro-actively try to find problems, not massage your data until they prove what you want them to prove, submit your work for peer-review, welcome replication efforts and calmly accept if you were proven wrong." Link
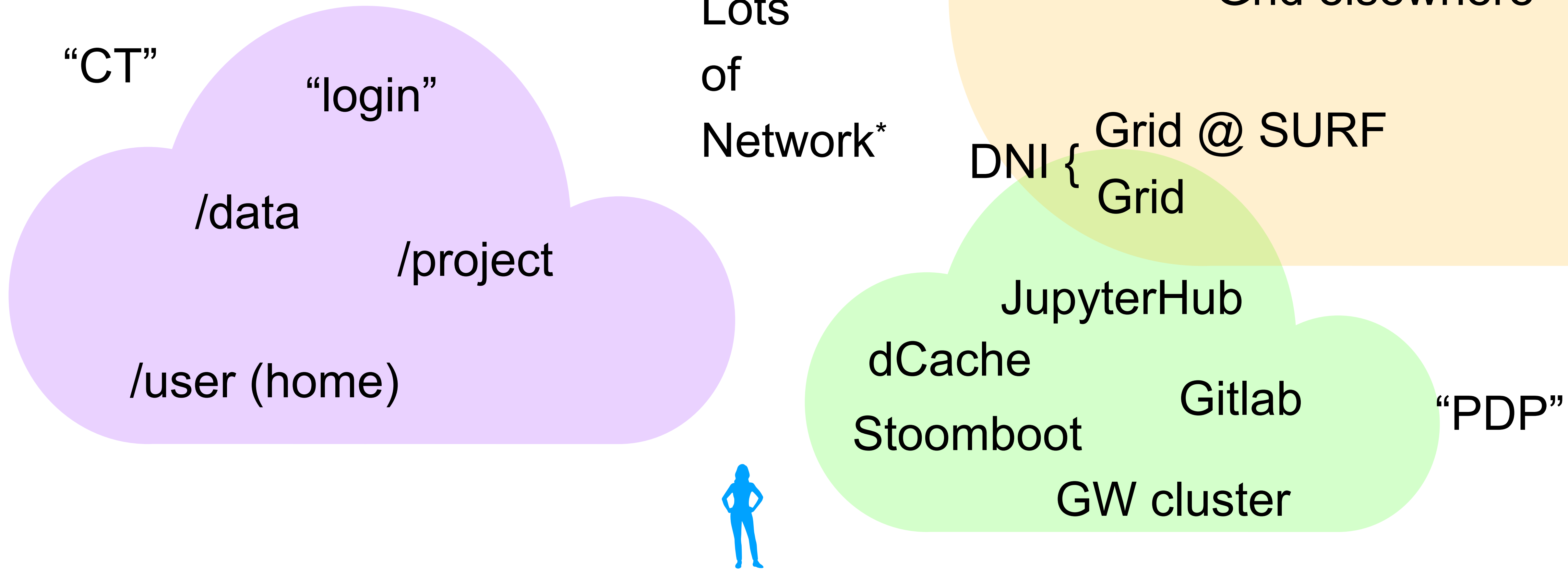
Nikhef

COMPUTE / DATA ECOSYSTEM

"Elsewhere"

Grid elsewhere

"CT"

"login"

Lots
of
Network*

DNI { Grid @ SURF

Grid

/data

/project

JupyterHub

dCache

/user (home)

Stoomboot

Gitlab

"PDP"

GW cluster

Nik|hef

TYPICAL USES

"Login" : jump host, editing files

Stoomboot interactive nodes : (stbc-i & stbc-g)

- Test runs
- Short / small real runs (< hrs / < 8 cores)

Stoomboot / Ganymede (GW) :

- Hours to days, up to 64 cores
- large memory (up to 256 GB)
- Thousands of jobs: stoomboot 2k cores, Ganymede 576 cores
- FAST access to dCache (hundreds of terabytes)

```
V100        0      1 (  100 %)
GTX1080     0      2 (  100 %)
MI50        0      6 (  100 %)
```

Useful to who?

TYPICAL MISTAKES

- Forgetting the ecosystem
  Student submits thousands of jobs, reading files from /data. Result : electronics engineers cannot access their files.

- Scaled-up run automated, leave immediately for vacation

- Giant Laptop Logic

  - Test run was okay, scaling up by 100x will be OK

    - Forgot that the 10 MB log files now become 1 GB

  - Job takes one hour on laptop. Try to speed it up.

GIANT LAPTOP LOGIC

One hour job on laptop, ported to stoomboot:

- 30 pieces : 131 sec (120 sec)
- 60 pieces : 73 sec (60 sec)
- 120 pieces : 47 sec (30 sec)
- 360 pieces : 51 sec (10 sec)
- 3600 pieces : 5 min 20 (1 sec)

Batch Overhead
Prefer jobs > 5 min
Avoid jobs < 1 min

Analogy:
Amsterdam -> Brussel

- 45 min airplane
- 2,5 hr by car

But:

- City to/from airport 30 min at each end
- 3 hrs beforehand at Schiphol
- True duration almost 5 hr by plane

HELP

- helpdesk@nikhef.nl
- Physical helpdesk : SP110 / D1.11
- Stoomboot / grid admins next door
- Stoomboot general questions: `stbc-users@nikhef.nl`
- Stoomboot something broken: `stbc-admin@nikhef.nl`
- When in doubt : `users` list … admins read it too.

GRID

- At Nikhef, Petabytes of storage, ~10k cpu
- Normally no direct submission
- Your collaboration has interfaces to use
- Can sometimes be an advantage to specify "Nikhef"
- Will need a certificate

CERTIFICATES, AUTHENTICATION, AUTHORIZATION

CAREFUL WITH PASSWORD: NIKHEF ACCOUNT IS POWERFUL!

GRID

- At Nikhef, Petabytes of storage, ~10k cpu
- Normally no direct submission
- Your collaboration has interfaces to use
- Can sometimes be an advantage to specify "Nikhef"

Come see us if you need more than stoomboot
And can't figure out the collaboration interface
Or there is no collaboration interface
Or you get stuck

TAKEAWAYS

- Remember : shared ecosystem. Be kind to your colleagues
- Think (recommend "Power Tool Fails" on YouTube)
- If you don't have time to think, come see us
- Come see us anyway. There is cake at office hours.