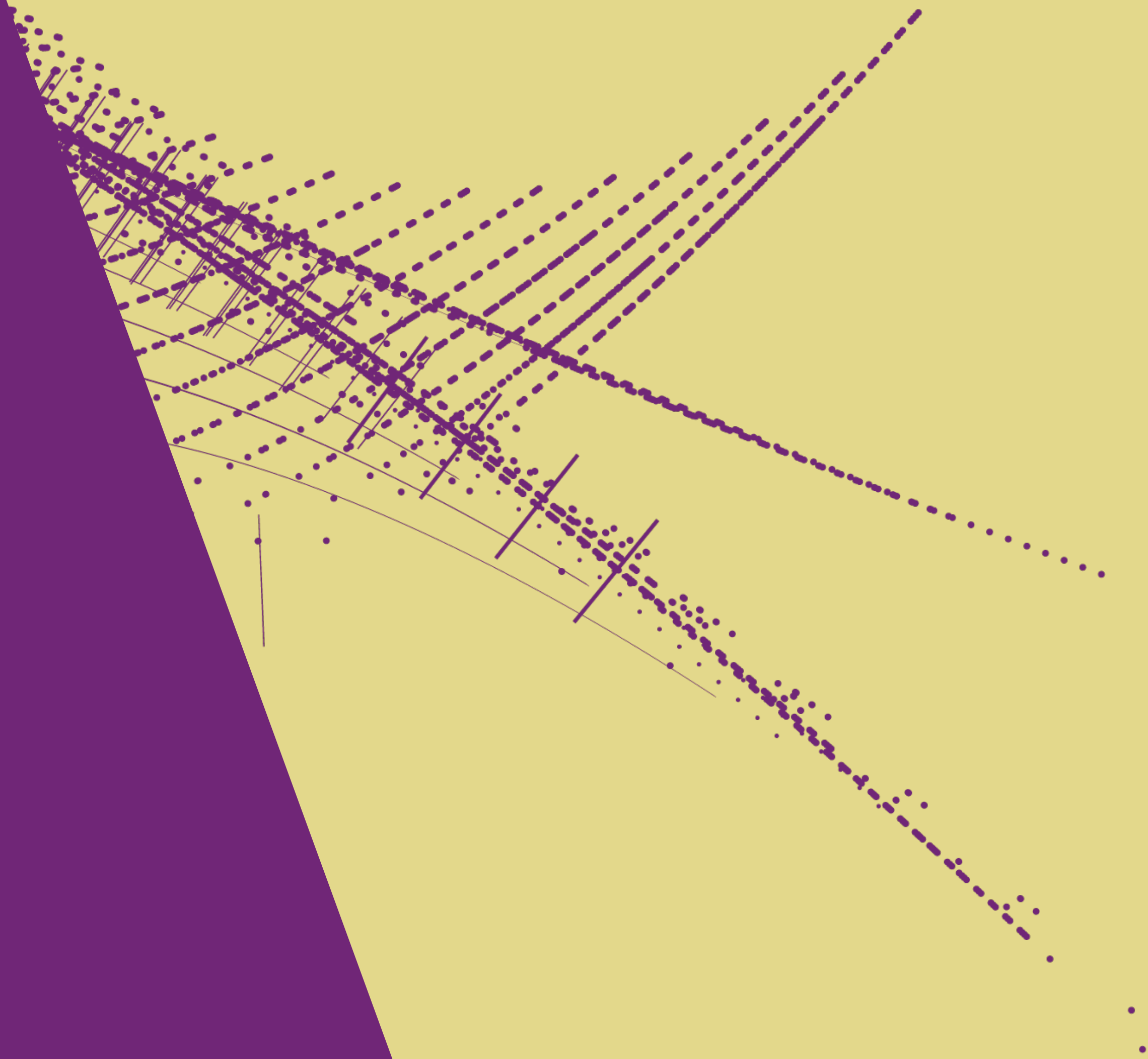




COMPUTING COURSE INTRODUCTION



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
- Dependencies on other code (eg numpy): record versions too!

FAIR

Findable, Accessible, Interoperable, Reusable

Nikhef



Jeff Templon

Research Integrity, Nikhef Research Computing Course

DATA AND SOFTWARE

Intellectual Honesty

"Scientists 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](#)

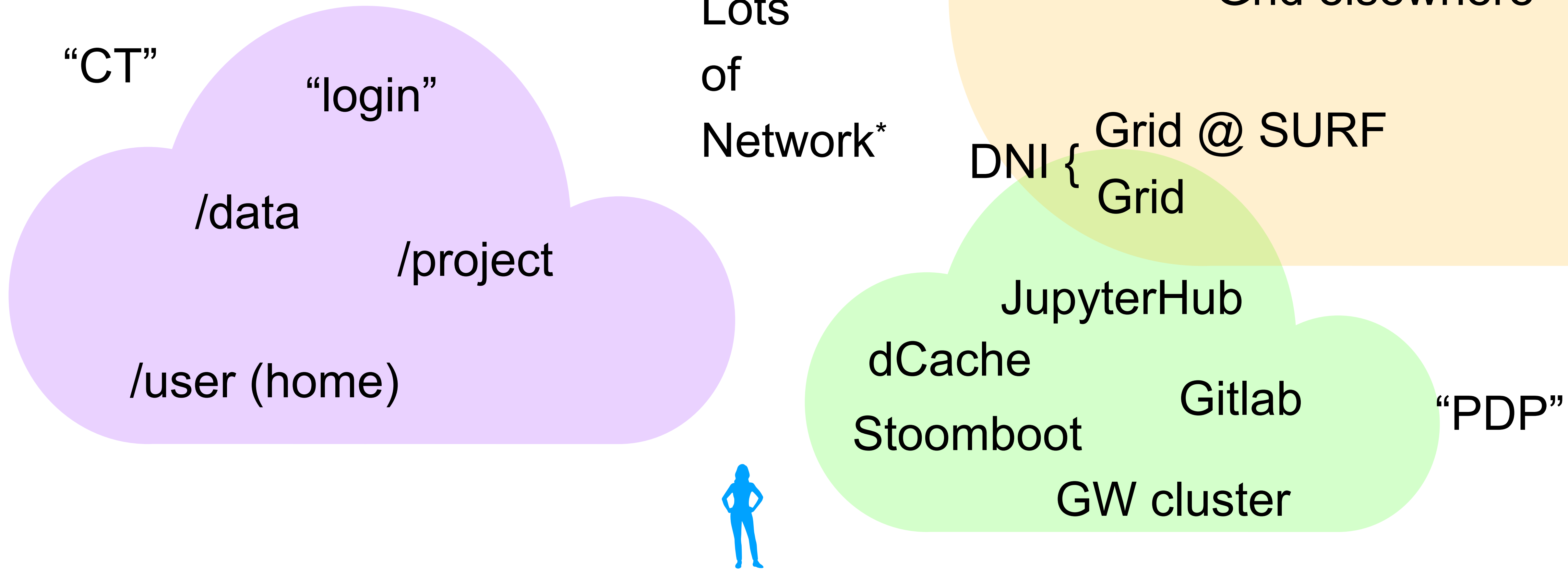
Nikhef



Jeff Templon

Research Integrity, Nikhef Research Computing Course

COMPUTE / DATA ECOSYSTEM



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

The Nikhef Identity Management and SSO Service

Nikhef and the Nikhef CT group are introducing centralized user management and a single-sign-on service, *NikIDM*: the Nikhef Identity Management system. The Nikhef Identity Management system is the organisational ICT system of Nikhef that implements the interface with regards to generic authentication and authorization for its Users.

If you have any problems logging in to the NikIdM services, look at the [User Documentation](#) or contact the helpdesk.

- **[Review your account](#)** to see whether you are allowed to access the service you want (and whether your password is correct)
- **[Change your password](#)** before it expires (your password must be changed every 13 months)
- **[Check the information that you give out](#)** to external service providers - to make sensible decisions about your privacy
- use *local and federated services* like SURFspot and eduGroepen, request eduroam visitor accounts, obtain your grid certificate and auto-login to CERN, register your networked devices like laptops, and much more ... see the left-hand menu for a selection of options.

Most Nikhef ICT services now use the NikIdM for user data and authentication: email, Linux and Windows desktop login, SUFconext services, grid certificates, travel requests, &c. Most of our services redirect you to this web site to log in, so that you can recognise the login page as a trusted brand. The list of services that use the NikIDM single sign-on service is given on the [CT User Documentation pages](#). In addition, all [NDPF services](#) (Grid and the Nikhef Data Processing Facility) use the same directory and single sign-on password.

What if my NikIdM data is no longer current?

Your NikIdM account is the source of identity and access control for CT services and all federated (SURFconext or eduGAIN) services. In case you need to update or augment your access rights or entitlements, please contact the [helpdesk](#).

Your Privacy

Your personal data is kept and processed in accordance with the [NikIdM Policy](#). Please see specifically chapter 8 of the policy for data protection information. You can review, augment and correct your data any time by contacting the help desk. When you use Nikhef services, your [attributes are released and may be processed](#) in order to provide you the service. In case you use federated services, you are explicitly informed regarding the attributes that will be released. You can then choose to

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.