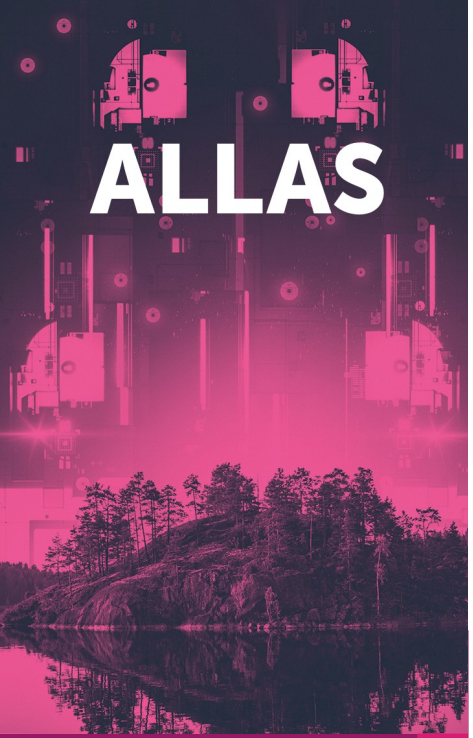
CSC

ICT Solutions for Brilliant Minds

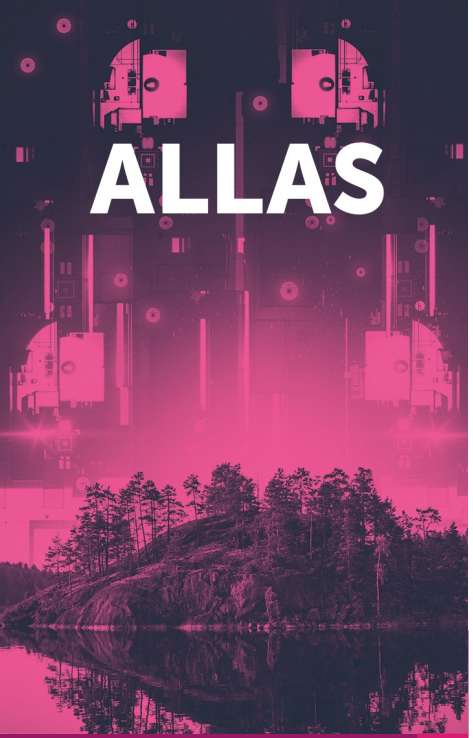
Allas Object Storage service

15.11.2019



Allas – object storage: what it is for?

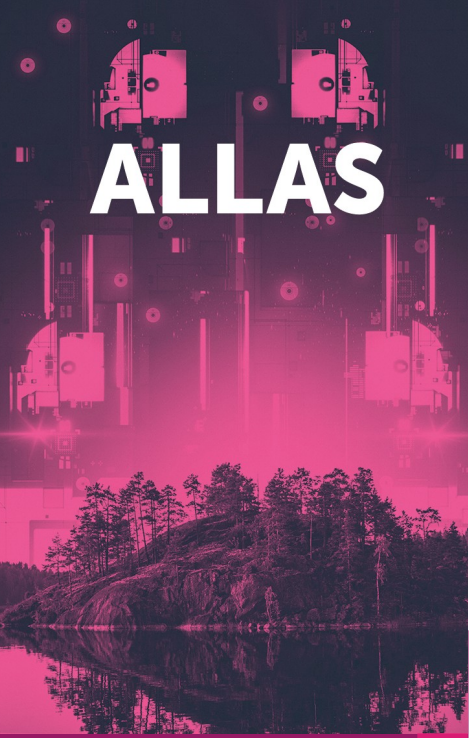
- Allas is new storage service for all computing and cloud services
  - Meant for data during project lifetime
  - Default quota 10 TB / Project.
  - Possible to upload data from personal laptops or organizational storage systems into Allas
  - Available in Taito, Puhti and Mahti
  - Data can also be shared via Internet



In customer use end of Q3 / 2019

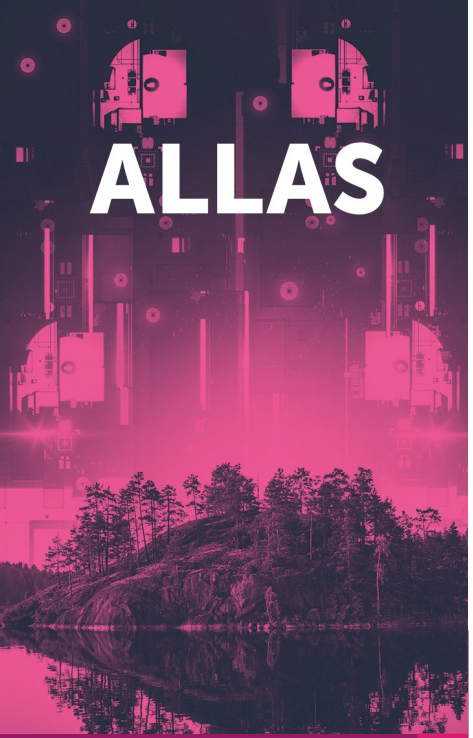
Allas - storage

- **12 Petabytes** of storage space for data stored over the life-time of a project
- Object storage based on open source CEPH
- An object is stored in multiple servers so a disk or server break does not cause data loss.
  - **There is no backup** i.e. if a file is deleted, it cannot be recovered
  - Data cannot be modified while it is in the object storage – data is immutable.
- Rich set of data management features to be built on top of it, initially S3 and Swift APIs supported



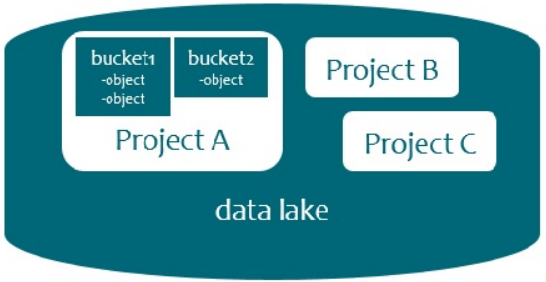
Allas – object storage: what it is for?

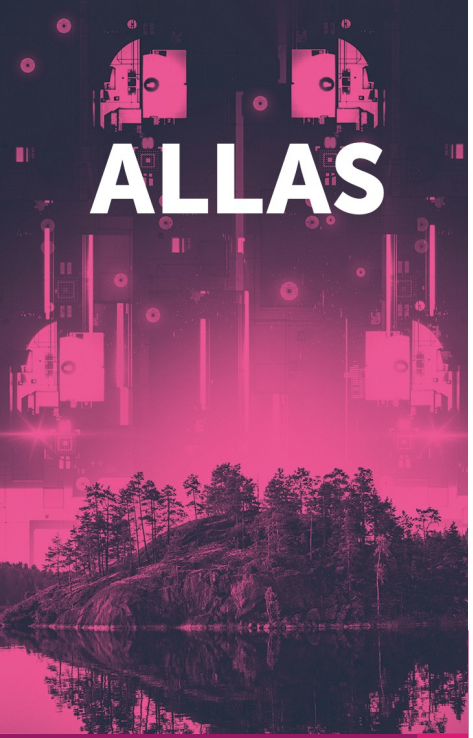
- Data can be moved to and from Allas directly without using supercomputer.
- For the computation the data has to be typically copied to a file system in some computer
- Data can be shared publicly to Internet, which is otherwise not easily possible at CSC.



Allas – object storage: terminology

- Storage space in Allas is provided **per CSC project**
- Project space can have multiple **buckets**
  - Only one level of hierarchy of buckets (no **buckets** within **buckets**)
- Data is stored as **objects** within a **bucket**
  - Blobs of data, can be anything (generally, **object** = file)
- Name of the **bucket** must be unique within Allas
- **Objects** can have metadata
  - Metadata are a key-value pairs, e.g. ”content=shoe”
  - Pseudofolders inside a **bucket** via metadata

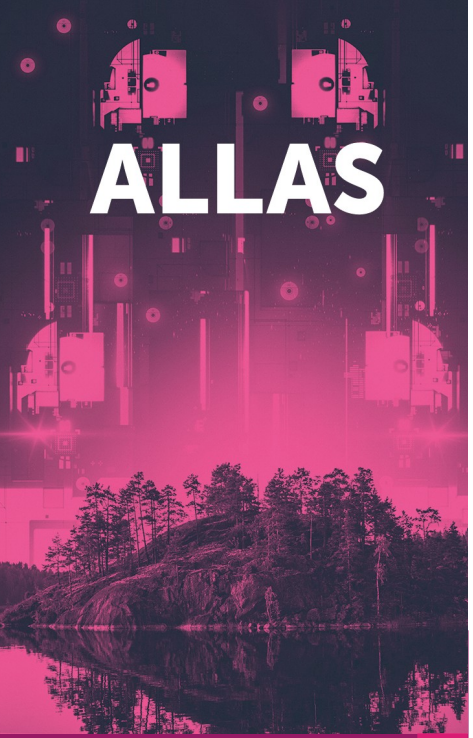




Allas – How to get access

Use https://my.csc.fi to
1) Register to CSC (haka)
2) Set up a project at CSC (Principal Investigator)
3) Apply for Allas service, quota and billing units for your project
4) Add other registered users to your project

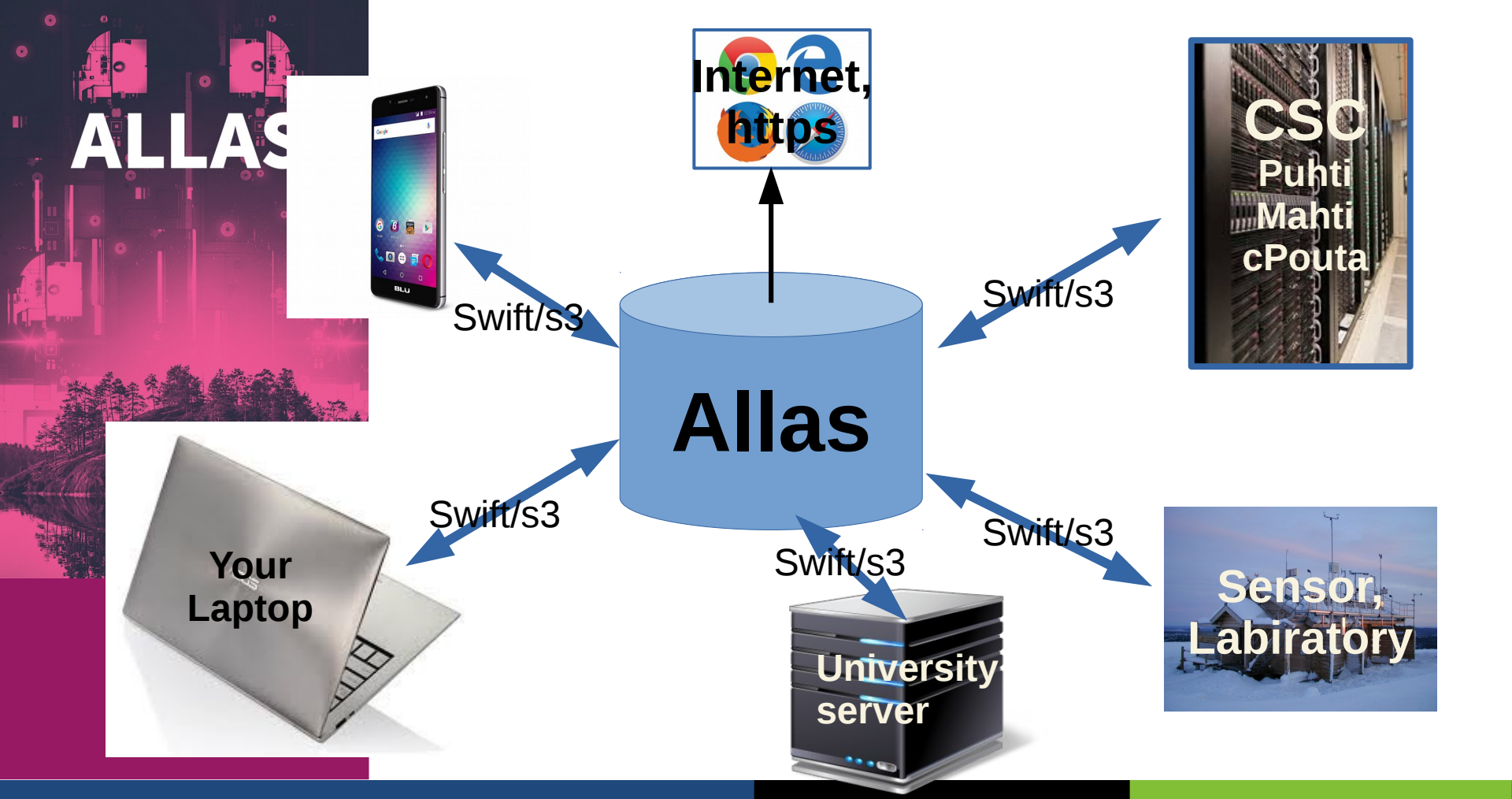
All project members have equal access to the data in Allas.

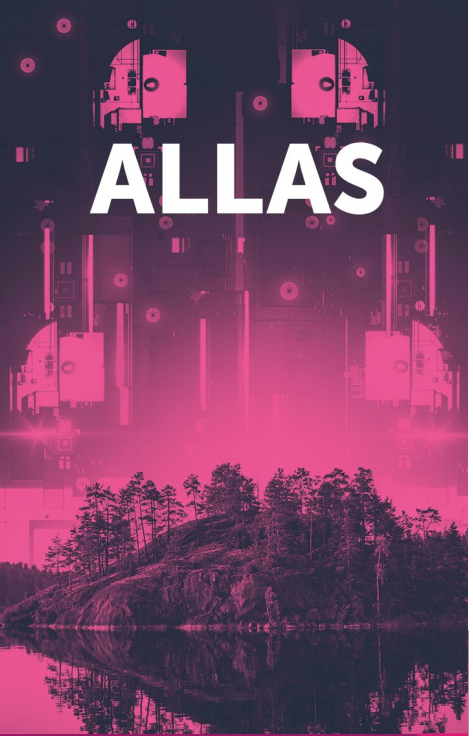


Allas supports Two Protocols

- **S3** (used by: s3cmd)
- **Swift** (used by: swift, rclone, a-tools, cyberduck)

- Authentication is different
  - S3: permanent key based authentication – nice, easy and unsecure
  - Swift: authentication based on temporary tokens – more secure, requires authentication every 8 hours
- Metadata is handled in different ways
- Over 5G files are split in different ways
- → **Don’t cross-use Swift and S3 based objects!**


ALLAS
Internet, https
CSC
Puhti
Mahti
cPouta
Swift/s3
Swift/s3
Allas
Swift/s3
Swift/s3
Swift/s3
Your Laptop
University server
Sensor, Labiratory



Allas Clients: read, write, delete

- Puhti, Taito, Linux servers:

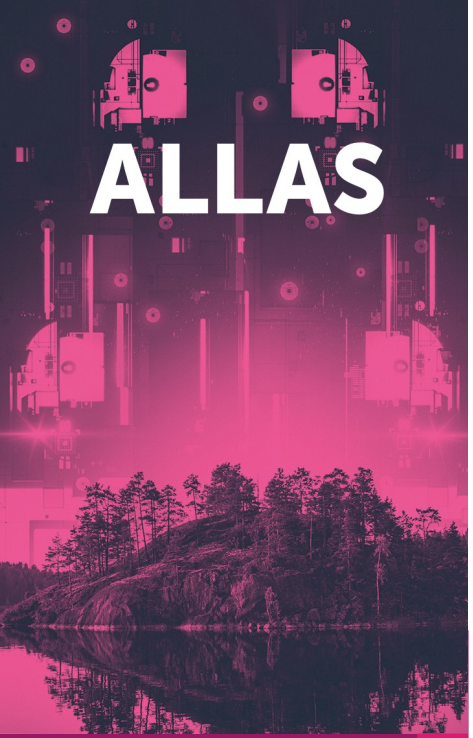
  - rclone, switft, s3cdm, a-tools

- Virtual machines, small servers:

  - FUSE based virtual mounts

- Laptops:

  - Cyberduck, FileZilla(pro), pouta-www interface



Allas – first steps for Puhti and Taito

1) Use https://my.csc.fi to apply Allas access for your project – Allas is not automatically available

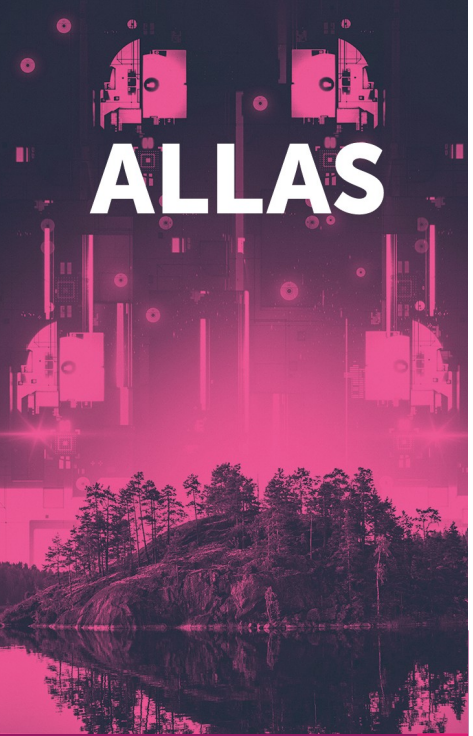
2) In Puhti and Taito, setup connection to Allas with commands:

```
module load allas

allas-conf
```

Study the manual and Start using Allas with rclone or a-tools:

https://docs.csc.fi/#data/Allas/

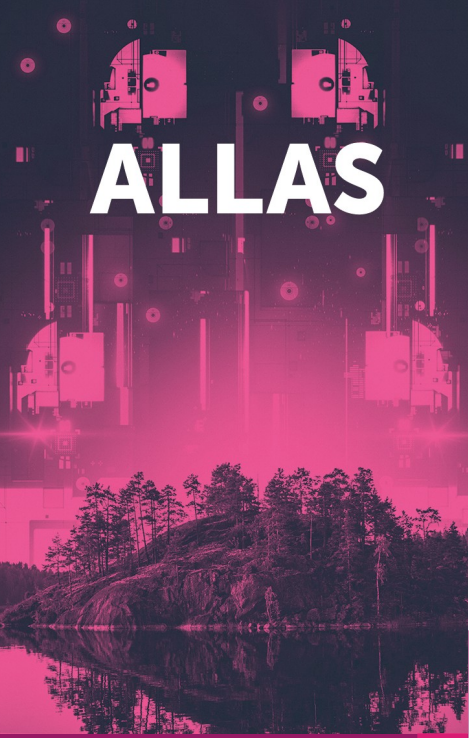


In customer use end of Q3 / 2019

Allas – rclone

- Straight-forward power-user tool with wide range of features
- Fast and effective.
- Available for Linux, mac and windows.
- Overwrites and removes data without asking!
- The default configuration at CSC uses swift-protocol but S3 can be used too.
- Use with care:

https://docs.csc.fi/#data/Allas/using_allas/rclone/



In customer use end of Q3 / 2019

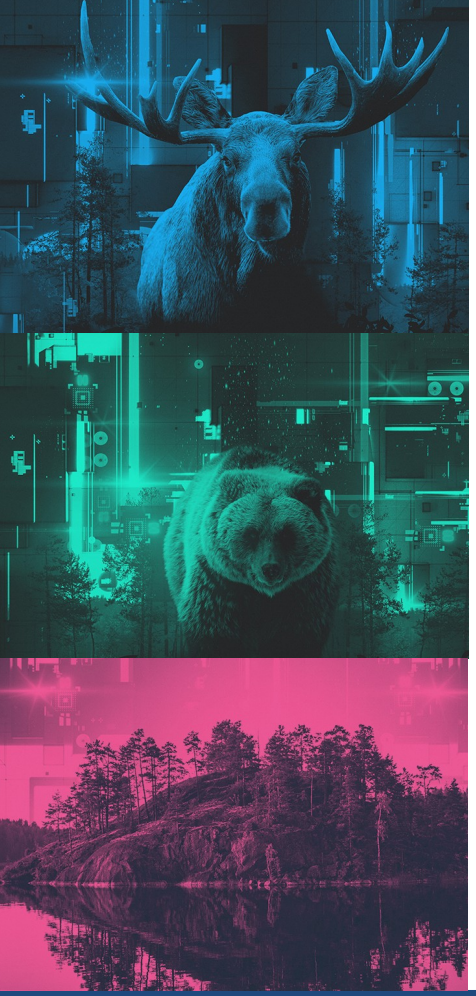
Allas – a-tools

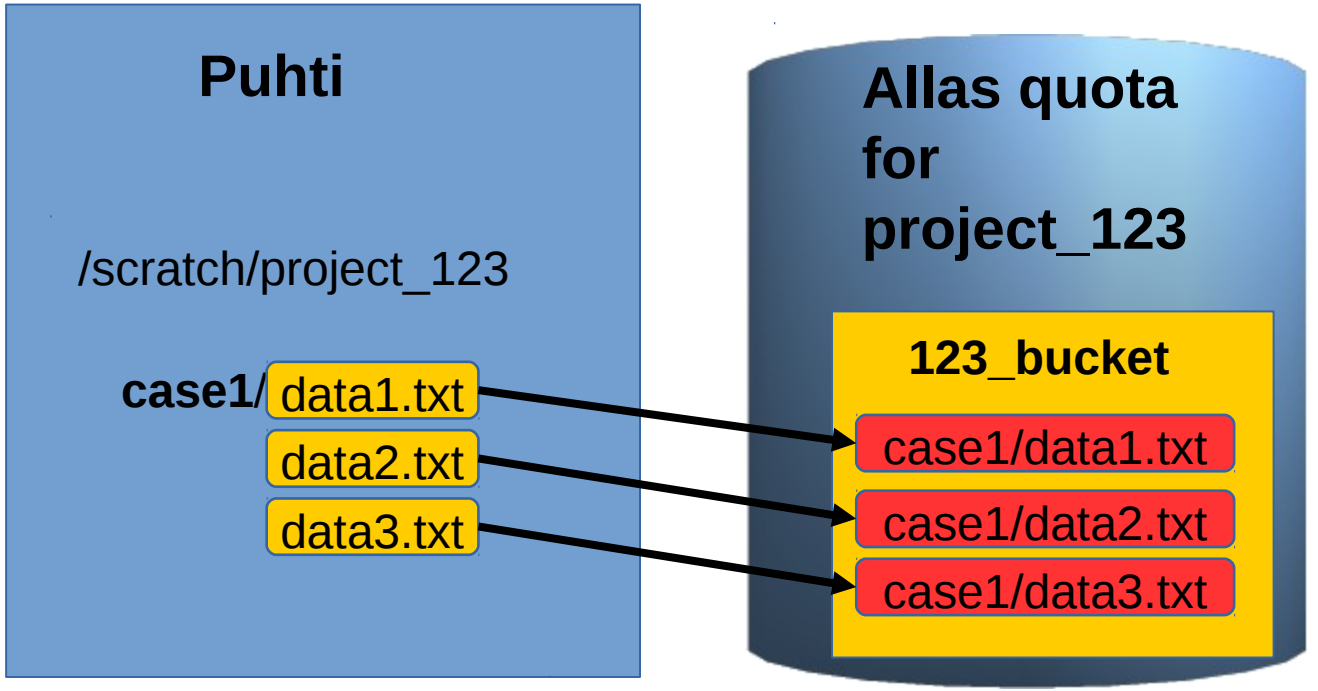
- Rclone based scripts for using Allas in Taito and puhti
- A-tools try to provide easier and safer way to use Allas for
- occasional Allas user users.

Developed for Taito and Puhti but you can install the tools in other linux, mac machines.

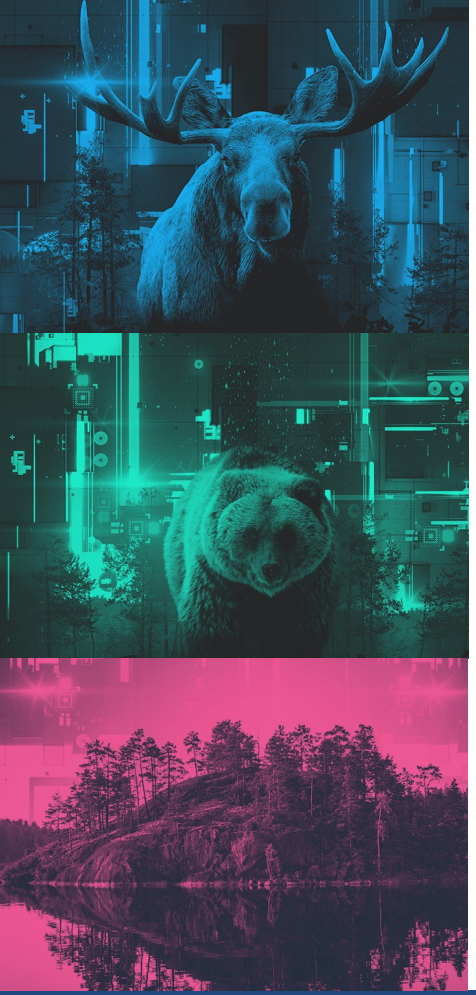
- Do not overwrite and remove data without asking!
- Automatic packing and compression.

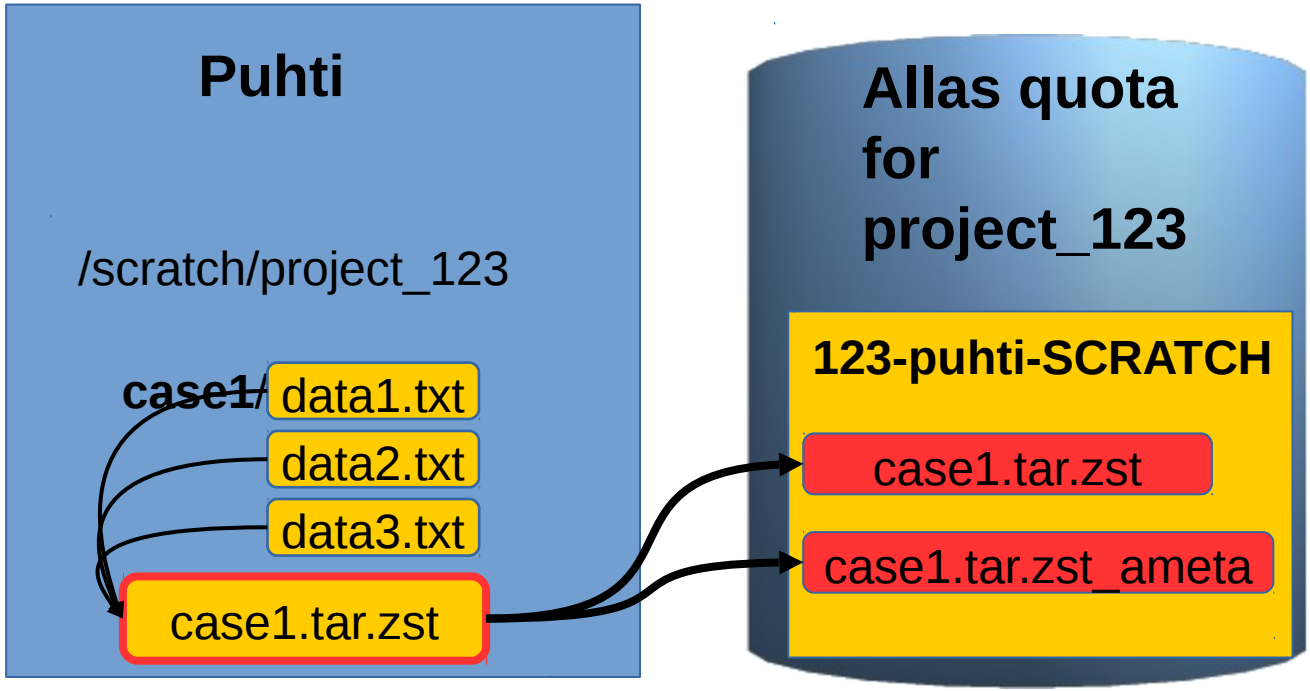
https://docs.csc.fi/#data/Allas/using_allas/a_commands/






```
rclone copyto case1/ allas:123_bucket/case1/
```




Puhti
/scratch/project_123
case1/
data1.txt
data2.txt
data3.txt
case1.tar.zst
Allas quota for project_123
123-puhti-SCRATCH
case1.tar.zst
case1.tar.zst_ameta

```
a-put case1
```

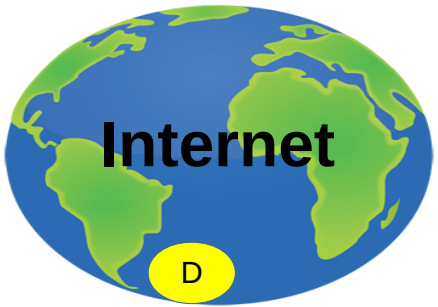
Things that users should consider

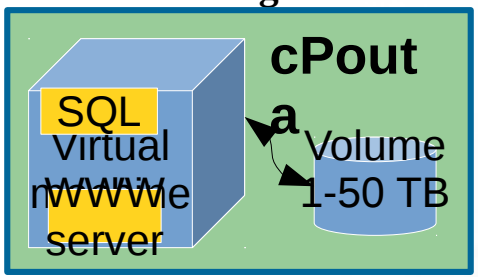
- Should I store files as one object or as bigger chunks?
- Should I use compression?
- Who can use the data: Projects and accession permissions ?
- What will happen to my data later on?
- How to keep track of all the data I have in Allas?

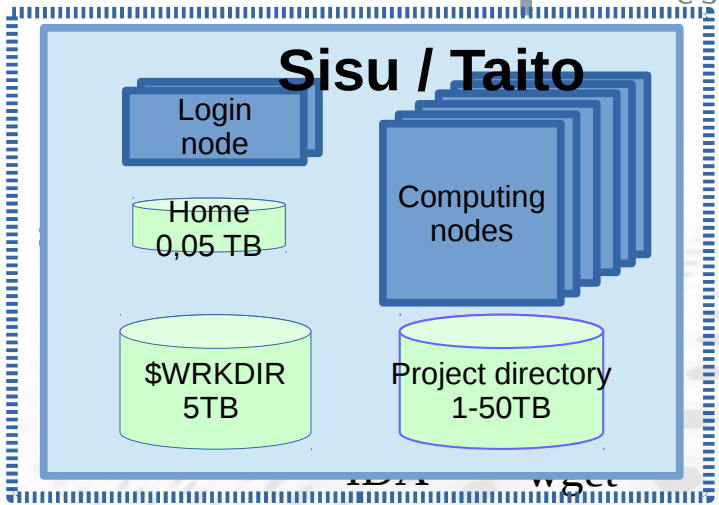
Installation and decommission schedule

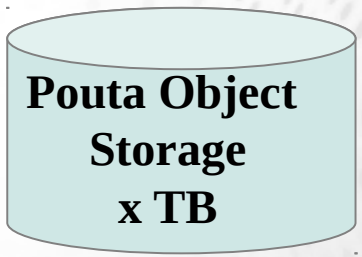
- April – July 2019: Puhti and Allas installation and acceptance testing
- August 2019: Puhti and Allas available for customers
  - Both CPU and GPU partitions of Puhti
- August 2019: Sisu decommissioned
- Late 2019 – early 2020: Mahti installation and acceptance testing
- End of 2019: Taito decommissioned
- Early 2020: Mahti available for customers

Reasearch project dataflow example

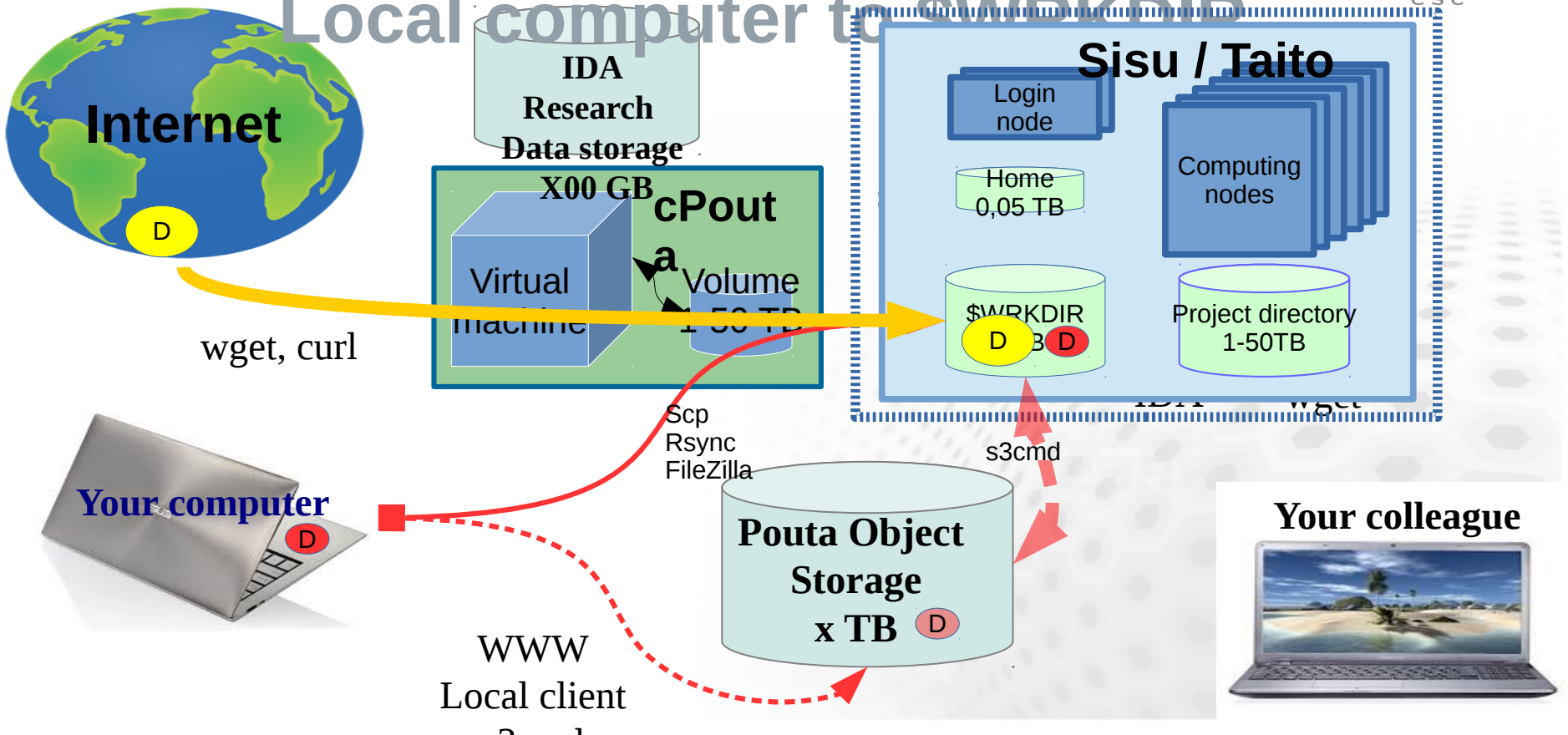




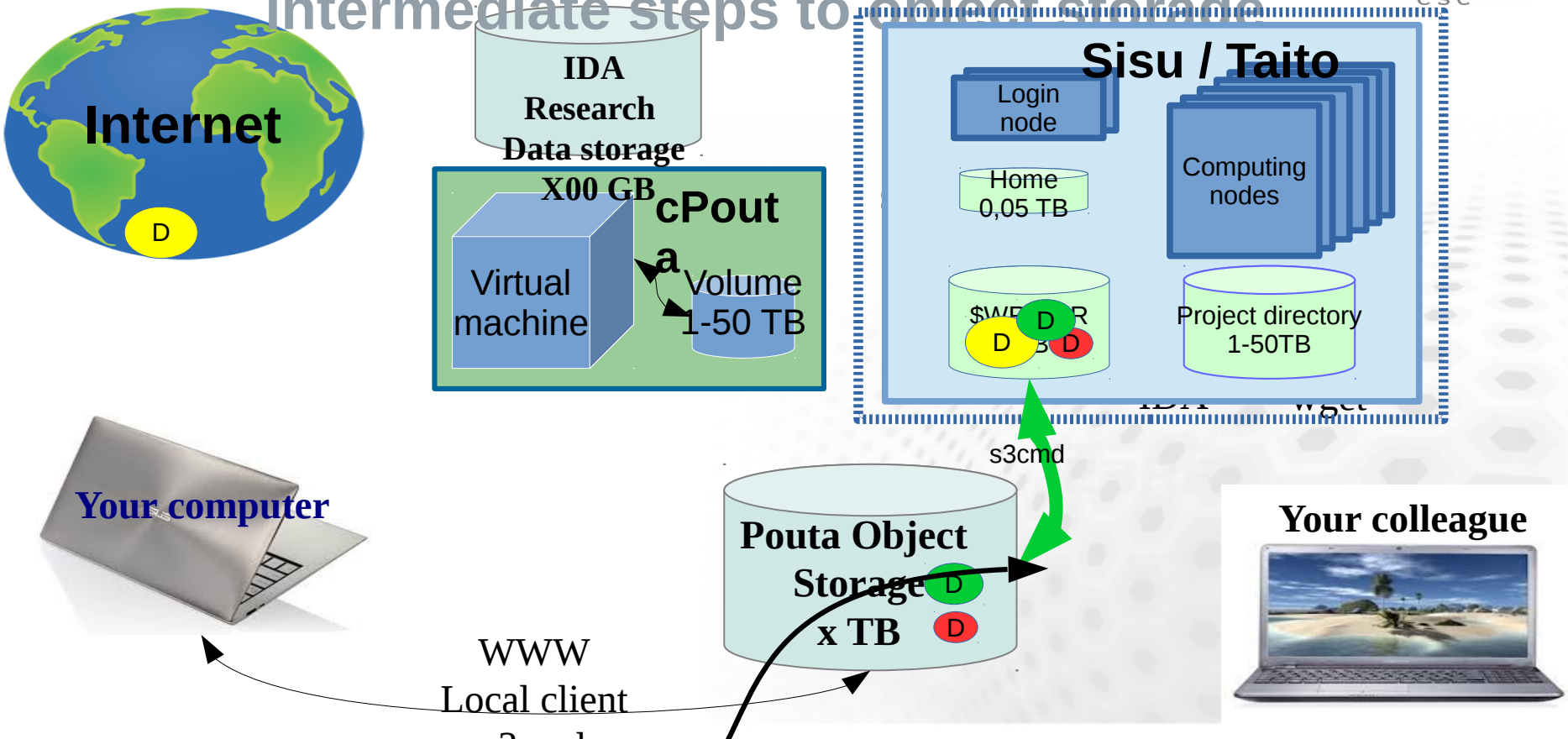




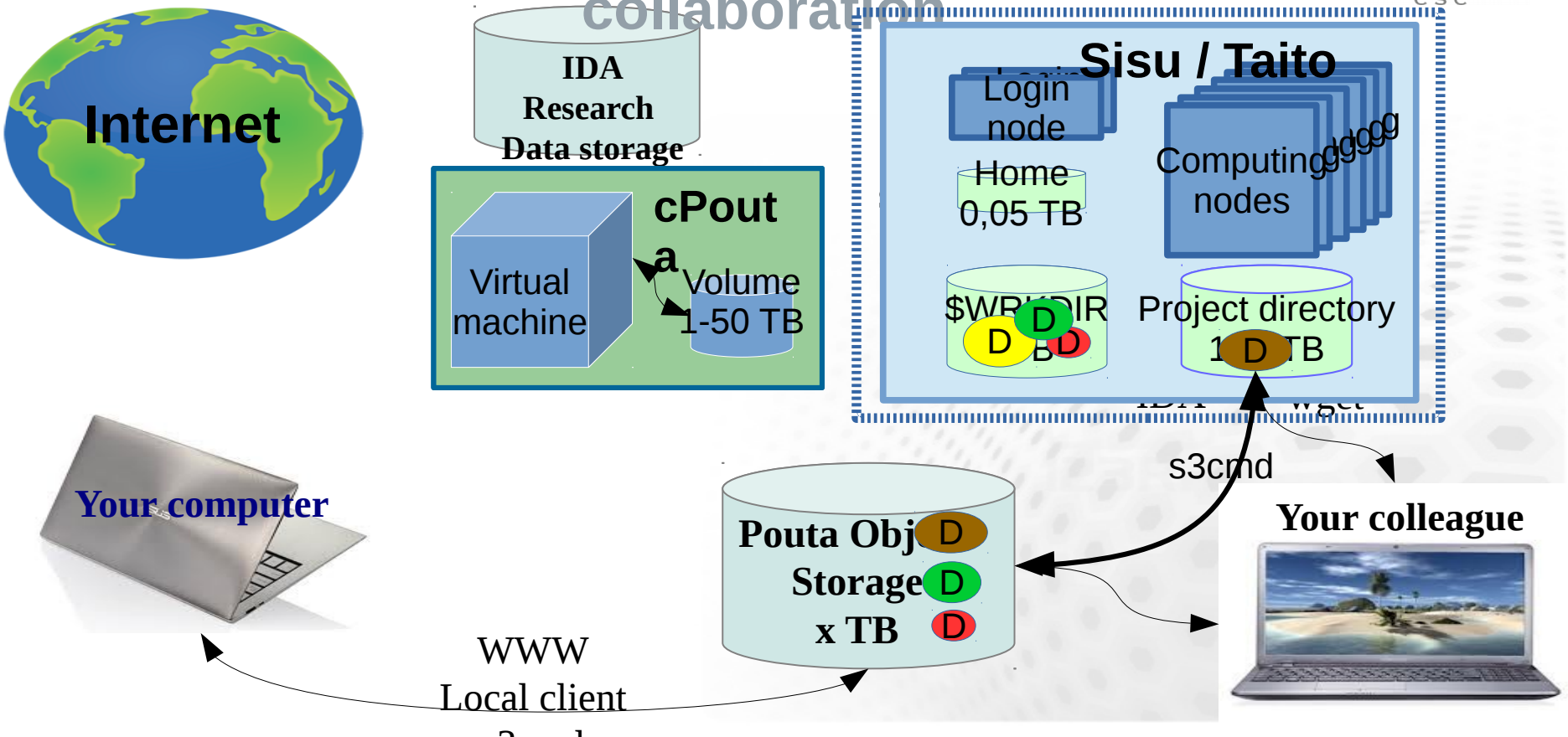
1. Copy the data from internet and your Local computer to $WRKDIR



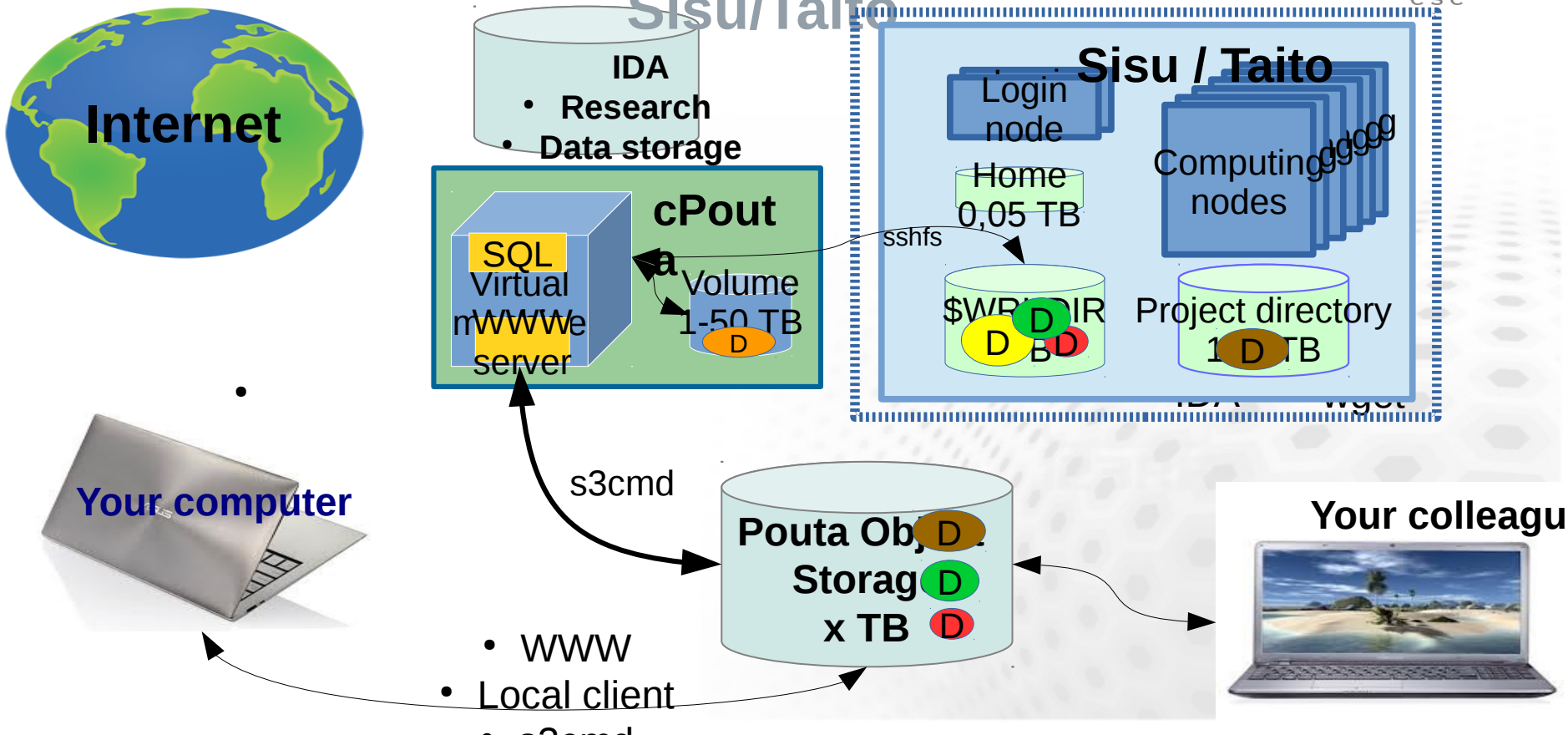
2. Process your data at CSC and backup intermediate steps to object storage



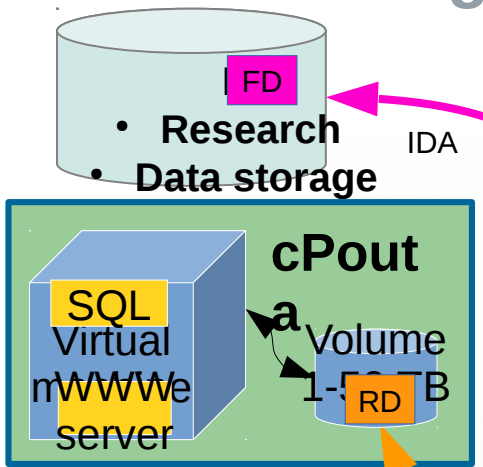
3. Use Object Storage and project directory for collaboration

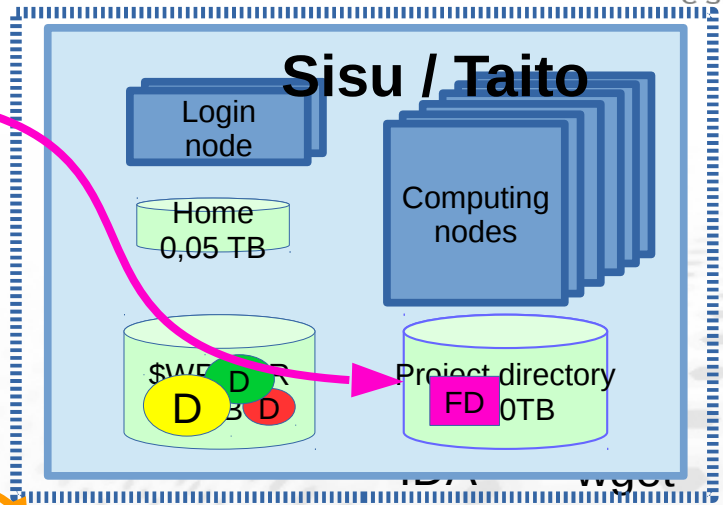


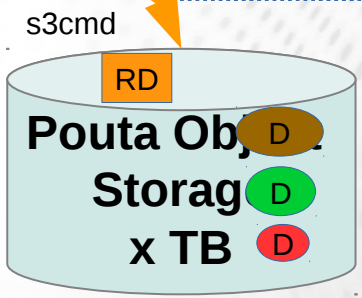
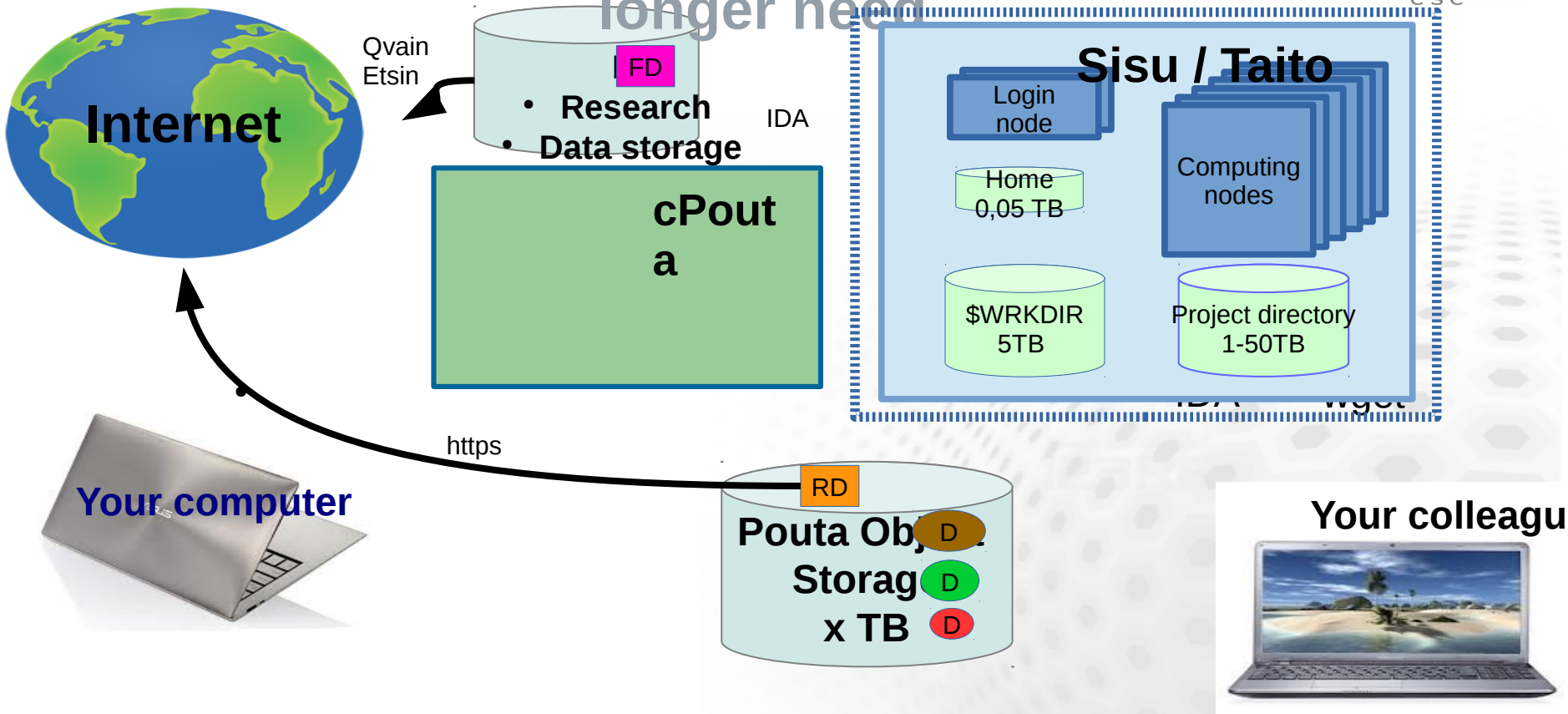
4. Use cPouta for tasks that are not possible in Sisu/Taito



5. Store your final results to IDA or object storage







6. Publish your data and clean the files you no longer need