Computational resources

What is needed? Where do I get it?

Marcus Lundberg

SciLifeLab Compute & Storage
marcus.lundberg@nbis.se

NBIS
NATIONAL BIOINFORMATICS INFRASTRUCTURE SWEDEN

SNIC

Computational resources

- What do you need
- What options exist
- How do I access SNIC resources
- What next?

Computers as scientific equipment

- For a given project, you'll have to evaluate your needs
- Consider what fits your needs:
  - a small self-maintained unit
  - a centrally-funded infrastructure
  - a combination or third option
- As with any equipment, training is beneficial or even necessary

Possible needs

- **Compute** needs: CPU cycles to run programs
- **Storage** needs:
  - Long-term space to store your raw data
  - Shorter-term space for analysis (raw data + results + temporary/intermediate files)
- Archiving (and Publication?)

Your laptop

- Good for:
  - Pre- and post-processing, scripting, etc.
  - A window into more powerful systems
- Limited by:
  - Disk space
  - Don’t want to run long-running programs
  - Sensitive personal data not OK
  - Managed by you/local IT

Powerful workstation

- Good for:
  - Running longer programs (can do some genome analysis)
  - Visualisation
- Limited by:
  - Disk space limited to a few TB
  - Managed by yourself/local IT
  - Tricky to prepare for sensitive personal data
  - What happens when system grows old or the main user moves on?

Local IT services

- University and department IT support can often provide various types of storage servers and small clusters

“Shadow IT”

- Buy your own systems for analysis and storage.
- Put disk drives with important data in locked drawer.
- Upside: quick & dirty
- Downside:
  - disorganised
  - unreliable
  - waste of time and resources in the long run

Swedish HPC Resources

- **SNIC** — Swedish National Infrastructure for Computing
  - Free to use
  - Collaboration
- **UPPMAX** — HPC center with large storage system
- **SNIC-SENS** — Secure system for sensitive personal data (WGS of living humans)
- **Compute and Storage** — SciLifeLab facility, part of Bioinformatics Platform, supports Life Science users at SNIC resources

SNIC Centra

- **HPC2N** - Very large memory, GPU's
- **UPPMAX** - Large storage
- **PDC** - Highly parallel processing
- **NSC** - Large general-purpose system
- **C3SE** - Göteborg
- **Lunarc** - Lund

Cloud

- **SNIC Science Cloud** (Free! Open to provisioning Software-as-a-Service and Platform-as-a-Service)
- **SUNet** SafeSpring (not free, ok for sensitive personal data)
- International science clouds
- Commercial providers:
  - Amazon web services (AWS)
  - Microsoft Azure
  - Google Cloud Platform

What SNIC can/can’t be used for

- **Do**:
  - Analyse data
  - Store data while you’re working with it
- Do **not**:
  - Use to (only) store/backup/move a dataset
  - Do patient diagnosis
  - Run a business

SNIC

Applying for SNIC project

- At UPPMAX and elsewhere, the procedure is the same
- Projects are managed in SUPR (supr.snic.se)
- Resources are granted to projects
- Project membership determines which files are readable and how much compute power is available
- https://www.uppmax.uu.se/support/getting-started/applying-for-projects/

Types of projects

- Compute project
- Storage project
- SNIC SENS
- Cloud

Scale of project

	Compute (core-hours/ month)	Storage (TB)	SENS
Small	< 10,000	< 10	< 10 kch/m < 20 TB
Medium	< 100,000	< 50	< 100 kch/m < 150 TB
Large	> 200,000	> 50	> 100 kch/m > 150 TB

The cost?

- Free to you but...
- Rackham's extension cost (very roughly, only initial investment):
  - 375 kr/TB per year
  - 0.1 kr/core-hour
- A typical PhD-student's project (1 TB, 1000 core-hours/month for four years): 6,300 kr.
- An ongoing potato genomics project (20 TB, 1000 core-hours/month): 8,700 kr each year.
- A large shotgun genomics project (5 TB, 30,000 core-hours/month) for half a year:19,000 kr.
- The creation of a new reference genome, requiring 100 TB and 50,000 core-hours/month for one year: almost 100,000 kronor.
- Large molecular dynamics simulations, using 200,000 core-hours/month: 240,000 kr/year.

So what do I need?

- If in doubt: apply for a **SNIC Small Compute** (at UPPMAX if you want a large software library)
- Apply for **storage** for:
  - Your sequencing data (NGI can tell you how much)
  - Plus an expansion factor (2-5x)
- Apply for both ideally about 1 month before NGI is ready to deliver data

Using SNIC/UPPMAX

- Linux command-line
- Graphical remote Linux desktop
- Training available via SNIC, UPPMAX, and NBIS

Getting data to UPPMAX

- From **NGI**: Data is placed on the Grus server and a “delivery project” is created on SUPR. You can then move the data to Rackham or Bianca by clicking a button in SUPR.
- From **elsewhere**: upload/download as normal using scp, rsync, sftp, wget, git, etc

Software at UPPMAX

- A huge software library is installed and maintained
- Additional software can be requested, but be aware this can take a while
- Most software can be installed by yourself

Sensitive personal data

- What is (or might be) sensitive personal data?
- Why do you need additional technical measures?

Bianca (SNIC SENS)

- “Lätt att göra rätt”
- A secure HPC environment
- Collaboration-friendly

transit.uppmax.uu.se

- A server is set up that allows easy “mounting” of Bianca wharfs.
- Useful for:
  - Copying data between SENS projects
  - Downloading data from internet to a SENS project

When does a project end?

- SNIC project lasts 12 months (typically).
- Must be extended or replaced, or access will be revoked.
- SNIC Common User Agreement states that data may be deleted after 6 months.

What next?

- After analysis, your data continues to live. Think about:
  - Publishing
  - Depositing into a repository
  - Hand it over to a university archive
  - Storing it yourself
    - Cloud (e.g. SUNet SafeSpring)
    - UPPMAX Offload Storage (Lutra)
    - SNIC Swestore
    - USB drive or HDD on a shelf in your office

Lutra (UPPMAX Offload Storage)

- For larger projects or groups
- 50 TB increments, 4 years
- Annual cost is 25,000 SEK/increment

Compute and Storage Resources

- The landscape is always changing
- Expertise is available
- Just ask if you have any questions:
  - support@uppmax.uu.se
  - marcus.lundberg@nbis.se
  - Anyone else you've heard from today

Thank you for listening