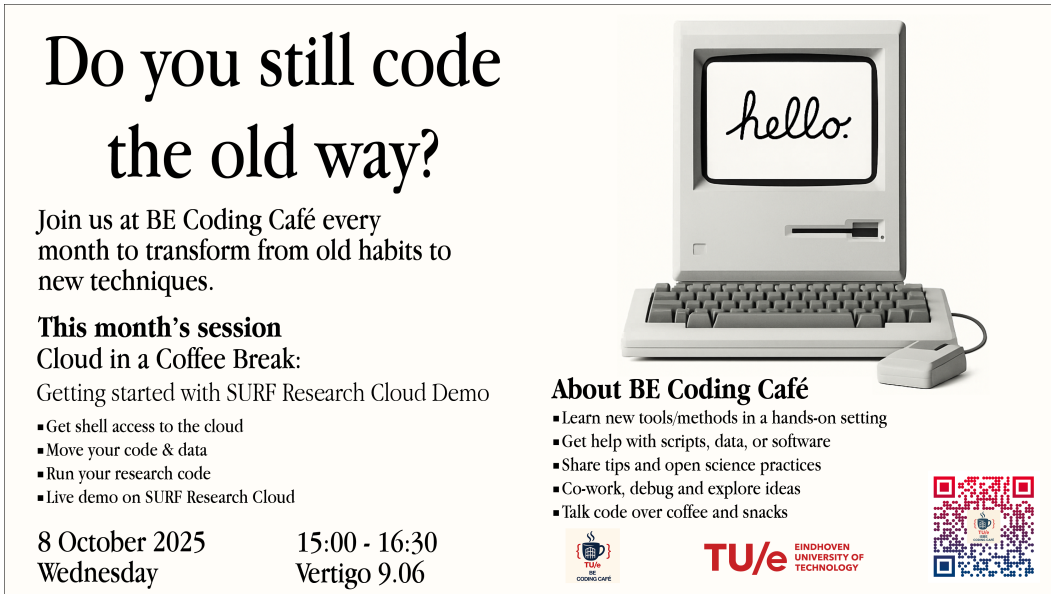
Cloud in a Coffee Break

Getting started with SURF Research Cloud Demo

Workshop context:
Coding Café @ Built Environment
September session
Instructor: Özgün Balaban

becodingcafe.com



parts of the content is adapted from Data Carpentries content:
https://datacarpentry.github.io/cloud-genomics/

What is Coding Cafe?

Originating in places like Netherlands eScience Center, Programming Cafés are low-threshold, peer-driven sessions where researchers:

- 👩‍💻 Present a short demo of a programming method or tool
- 🛠️ Ask or offer help with code or workflows
- Create networking between people working on same issues / software
- Co-work informally, with coffee or pizza
- Focus on community over curriculum

Session Goals

- Understand what “cloud computing” means in research.
- Learn when and why to use it.
- Experience remote access concepts (SSH, workspaces).
- **SURF Research Cloud**.

Outline

1. Why Cloud Computing
2. Logging into the Cloud
3. Verifying Your Setup
4. Choosing a Platform
5. Mapping to SURF Research Cloud
6. Wrap-up and Reflection

Why Cloud Computing?

- Big data needs **more RAM / CPU / storage** than laptops provide.
- Cloud = **on-demand remote compute**.
- Pay for what you use, no hardware purchase.

When Do You Need the Cloud?

- Your local system runs out of memory or storage.
- Analyses take hours/days to complete.
- You need specific software that’s hard to install locally.
- You want reproducible or travel-friendly environments.

Advantages

Benefit	Description
Scalability	Use 1 to 64 cores instantly
Flexibility	Install anything you need
Accessibility	SSH from any device
Reproducibility	Snapshot entire environments

Disadvantages / Trade-offs

- Upload/download times for big data.
- Costs accumulate while instance is running.
- Less institutional IT support.
- Risk of “orphaned” instances left running.

Which Cloud for my data?

🔷 Public Clouds

- **AWS** – largest ecosystem; broadest range of services (EC2, S3, SageMaker).
- **Google Cloud Platform (GCP)** – strong in analytics & AI (BigQuery, Vertex AI).
- **Microsoft Azure** – popular in enterprise and universities; integrates with Office 365 & AD.
- **SURF Research Cloud (SRC)** – privacy-compliant, SRAM-based access, community pricing.

Other Compute Options

Type	Description	Ideal For
HPC / Cluster Computing	Dedicated high-performance nodes managed by schedulers (Slurm, PBS).	Simulation, modeling, long batch jobs.
PaaS (Platform-as-a-Service)	Managed runtime environments (JupyterHub, Databricks, Azure ML).	Collaborative analysis, data science.
SaaS (Software-as-a-Service)	Fully hosted applications (Zenodo, Overleaf, ArcGIS Online).	Sharing, publication, specific workflows.
Hybrid / On-Prem + Cloud	Combine institutional servers + cloud bursts.	Sensitive data, cost control.

What to use at TU Eindhoven

https://openpar.pages.tue.nl/solution-searcher/index.html#categories

Determining the right resources

CPU (vCPUs)

- Start small (1–2 vCPUs), scale when needed
- Parallel? Use multi-core or multi-process
- Non-parallel? Extra cores won't help

RAM (Memory)

- Dataset + working copies must fit in RAM
- If swapping occurs, add RAM or chunk data
- Typical: 4–8 GB (light), 16–32 GB (data wrangling), 64 GB+ (ML / big joins)

OS & Packages

- Pick OS your toolchain supports (Ubuntu LTS is a safe default)
- Prefer containers to pin dependencies
- Record environment (Dockerfile / environment.yml)

Storage (Capacity & I/O)

- Capacity: dataset + intermediates + 30% headroom
- I/O: SSD/local NVMe for temp scratch; network storage for sharing
- Watch egress time for large transfers

Key Lesson

You pay for running resources — not for finished jobs.

Always shut down or stop your workspace when done.

SURF Research Cloud Demo

✅ **Topic Suggestion:** Working in the shell.

Logging Onto the Cloud

Concepts:

- A *remote instance* = virtual computer in the cloud.
- You access it via **SSH** (Secure Shell).
- Uses key-pair authentication or passwords.

Hello

✅ **Topic Suggestion:** Your workspace was created successfully.