Alphacruncher for Education

Zurich Initiative on Computational Economics (ZCCFE)

alexandru popescu

University of Lugano and Alphacruncher AG
popescu@alphacruncher.com

January 16, 2017

Outline

Alphacruncher: A New Approach to Data-Driven Research

Computation Services

Key Concepts
- Parallel Computing
- High Performance Computing
- Version Control

Live session

Summary

- John W. Tukey (1962, The Future of Data Analysis):

  "... my central interest is in data analysis, which I take to include, among other things: procedures for analyzing data, techniques for interpreting the results of such procedures, ways of planning the gathering of data to make its analysis easier, more precise or more accurate, and all the machinery and results of (mathematical) statistics which apply to analyzing data."

- Alphacruncher is the cloud platform for scientific research:

  - Endorsed by the Swiss Finance Institute.
  - Supporting research at the following Universities: Geneva, Zurich, Lugano, St. Gallen, HEC Paris and Oxford.

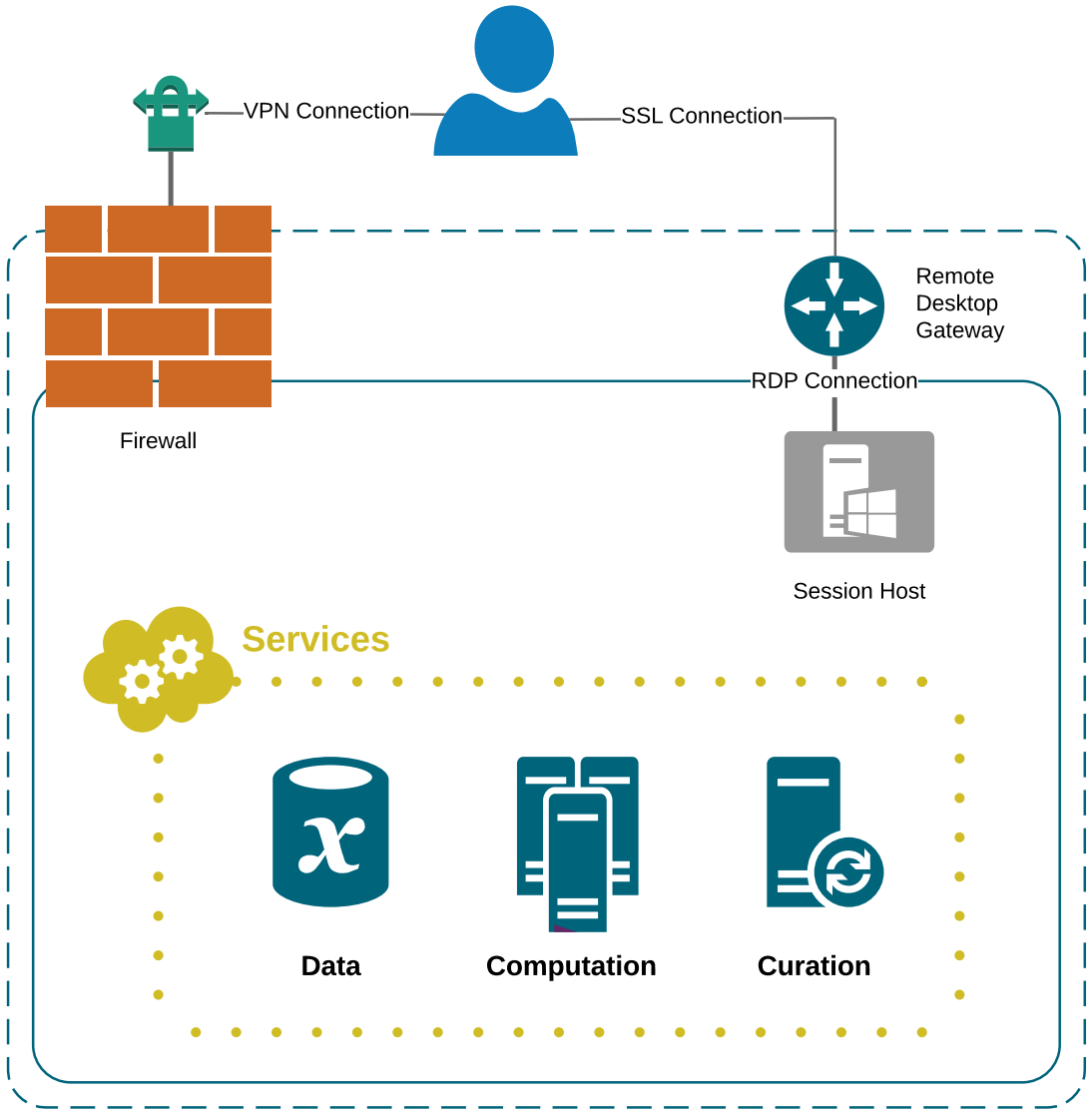
Alphacruncher: A New Approach to Data-Driven Research

- Infrastructure for data-analytic research, integrating services for:

  1. Data.
  2. Computation.
  3. Curation.

- Easily accessible and user-friendly platform on a secure cloud.
- Neutrality with respect to data properties and application-type.

Alphacruncher: Architecture



Computation Services

- Computational ressources are costly and not easily accessible.
- Alphacruncher boosts research, by harnessing the power of the cloud.

1. Parallel computing in a scalable high-performance cluster.
2. Trusted environment with smooth learning curve for beginners.
3. Full integration with data services.
4. Exponential productivity gains.

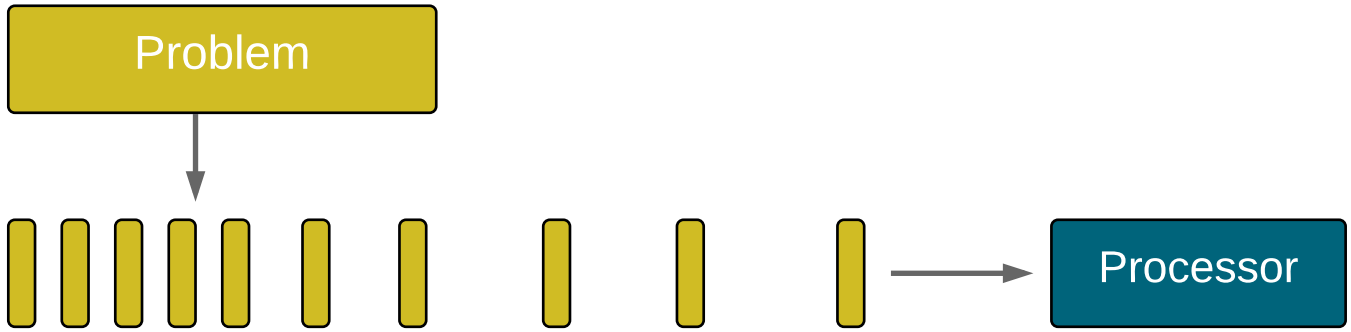
Key Concept: Serial vs. Parallel Computing

- Serial Computing:
  - A problem is broken into a discrete series of instructions
  - Instructions are executed sequentially one after another
  - Executed on a single processor
  - Only one instruction may execute at any moment in time

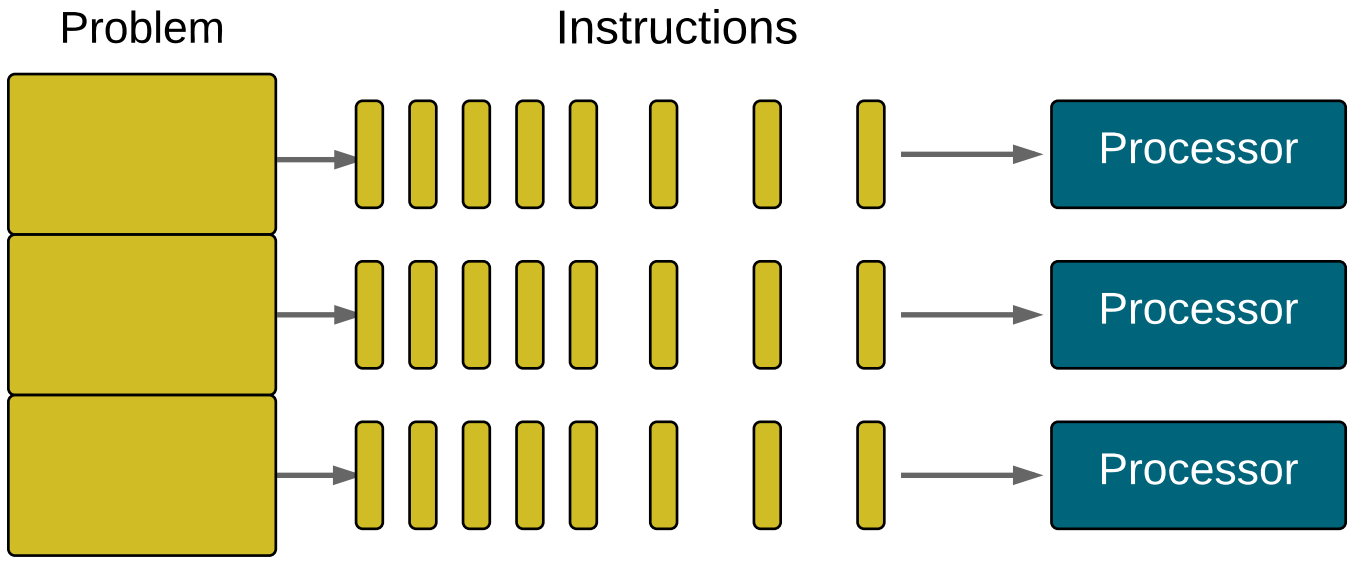
- Parallel Computing:
  - A problem is broken into discrete parts that can be solved concurrently
  - Each part is further broken down to a series of instructions
  - Instructions from each part execute simultaneously on different processors
  - An overall control/coordination mechanism is employed

Key Concept: Serial vs. Parallel Computing

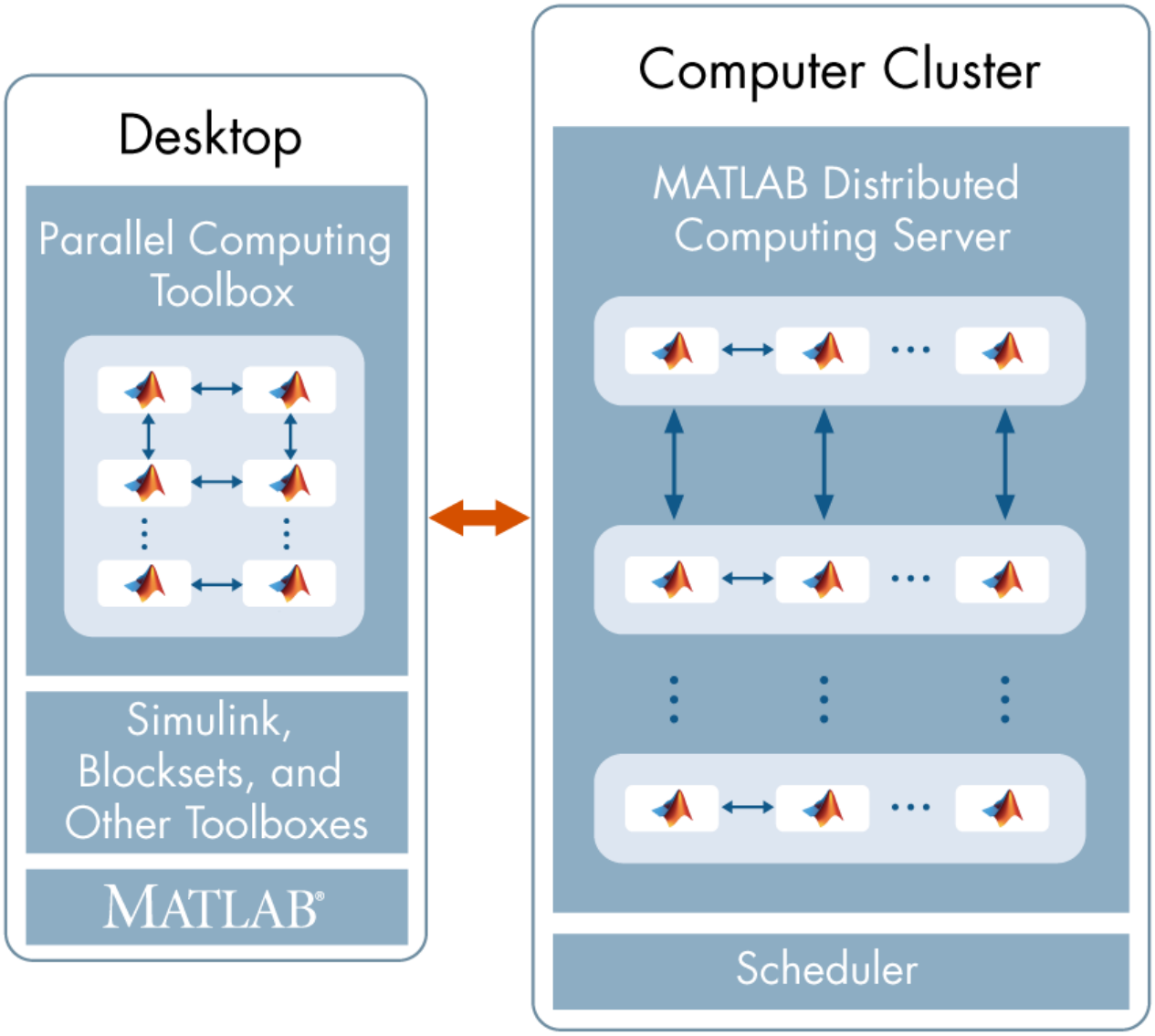
- Serial



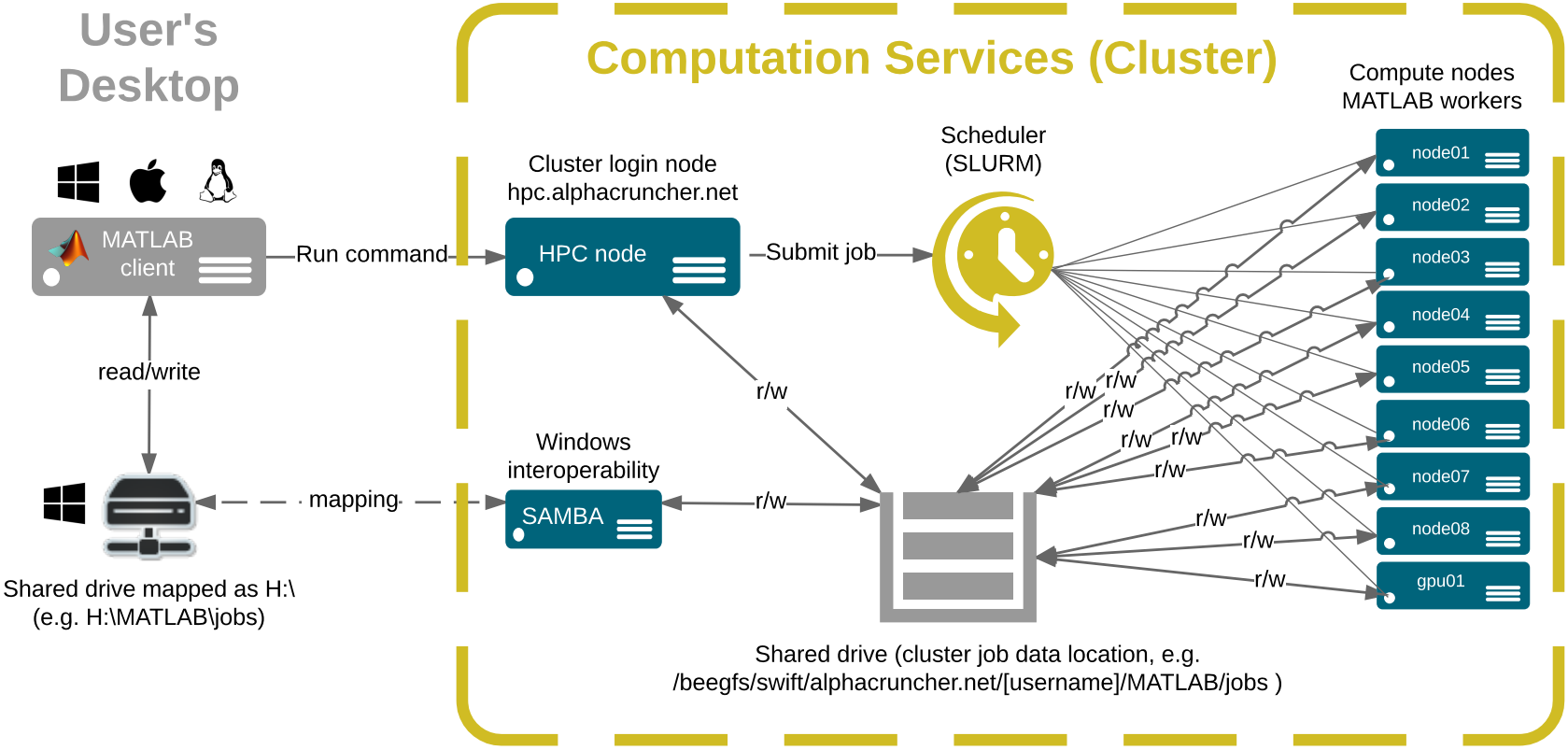
- Parallel



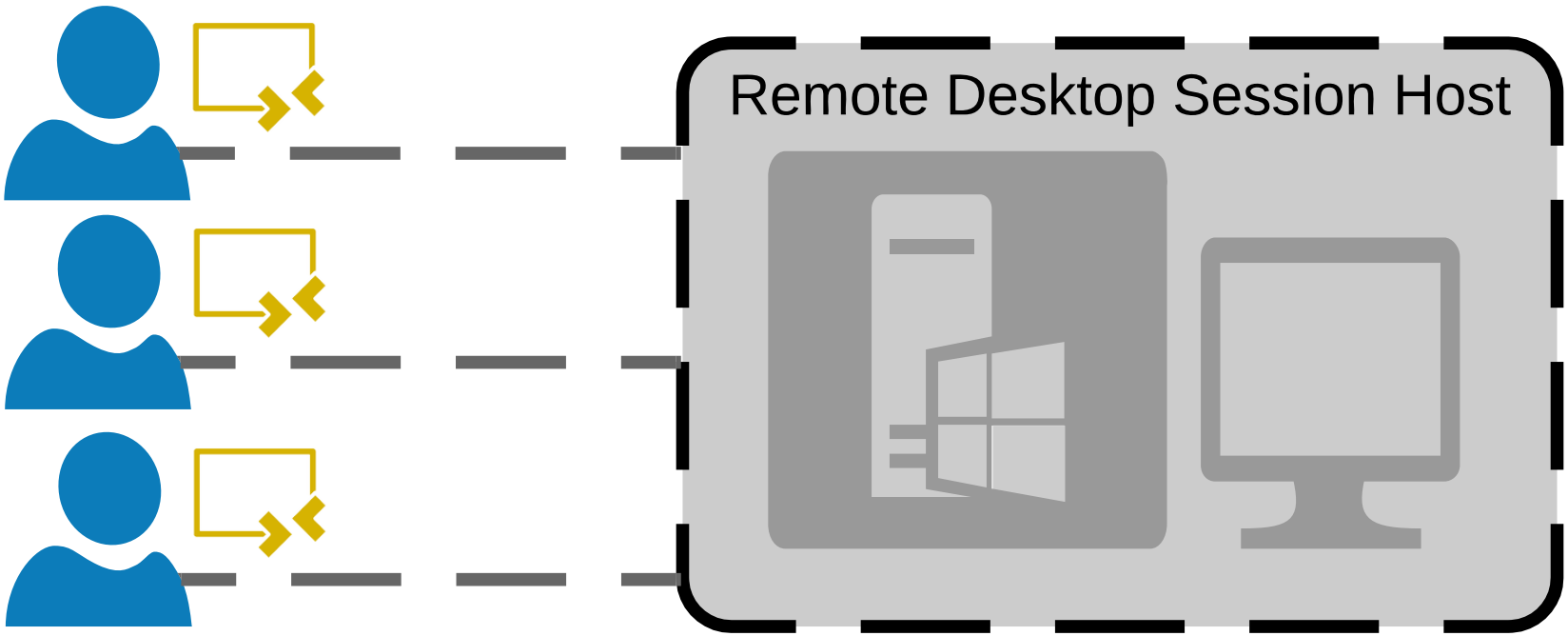
Key Concept: Parallel Computing in MATLAB



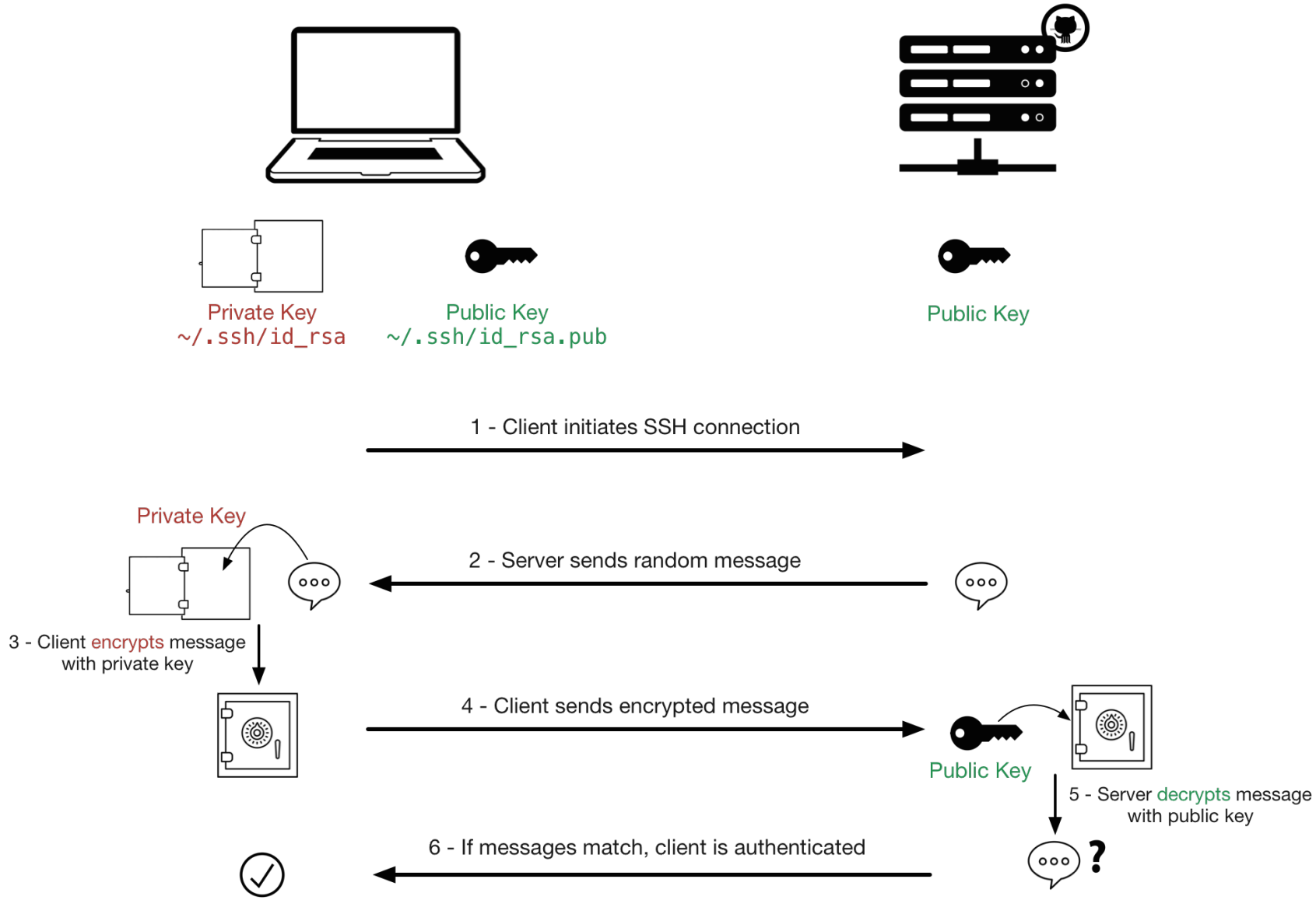
Key Concept: High Performance Computing Cluster



Key Concept: Session-based desktop deployment



Key Concept: SSH Key Connection



Key Concept: Version Control

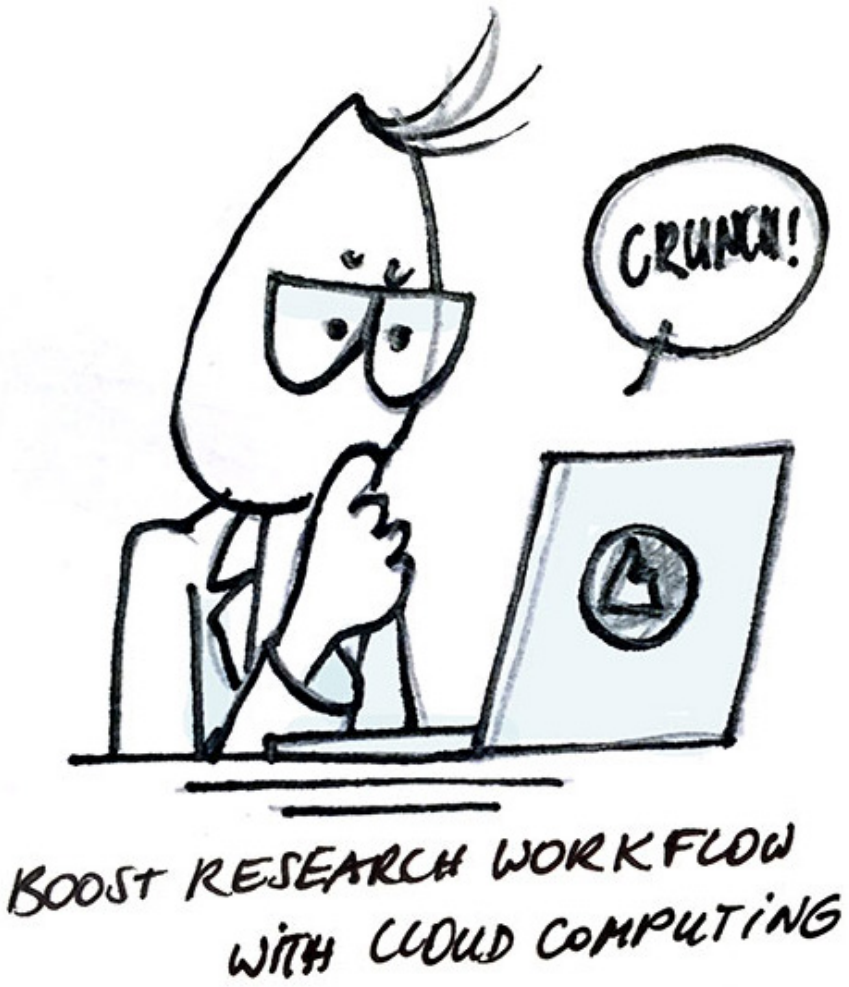
- Why keep track of the various versions of (your) work?

    1. Data is augmented and curated.
    2. Code is adapted and improved.
    3. Writing is revised and expanded.

- Learn the Command Line
  `https://www.codecademy.com/learn/learn-the-command-line`

- Learn Git
  `https://www.codecademy.com/learn/learn-git`
  `https://backlogtool.com/git-guide/en/`

Live Session on Alphacruncher platform



alphacruncher

MASTERING DATA

APPLYING SCIENCE

BOOSTING RESEARCH

www.alphacruncher.com