

# Portable HPC workflows & software re-use

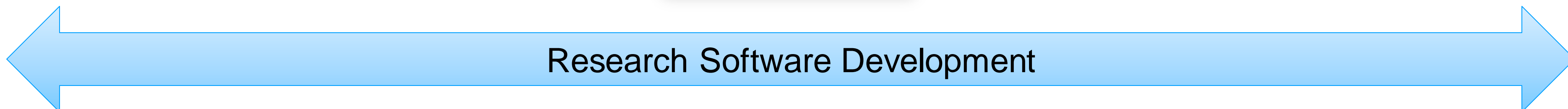
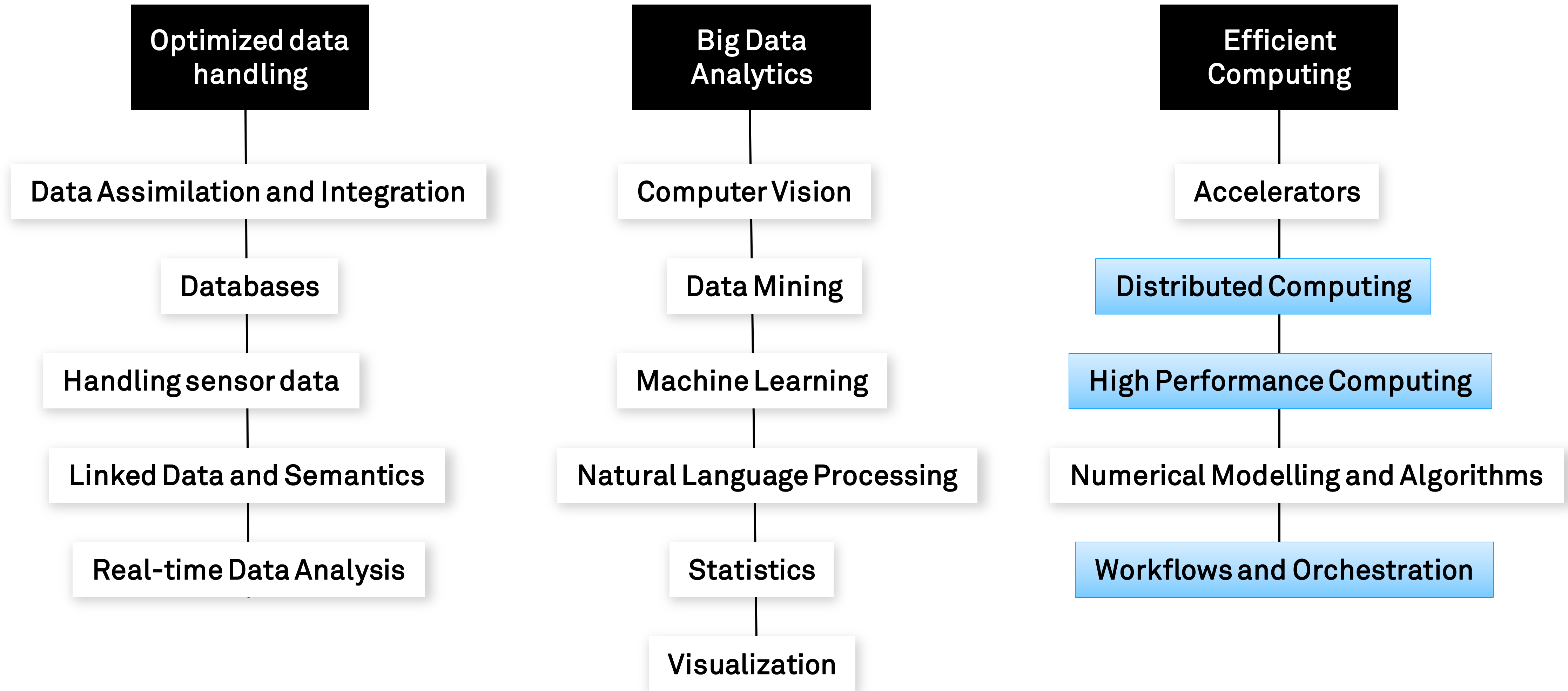
Arnold Kuzniar  
eScience research engineer

✉ [a.kuzniar@esciencecenter.nl](mailto:a.kuzniar@esciencecenter.nl)

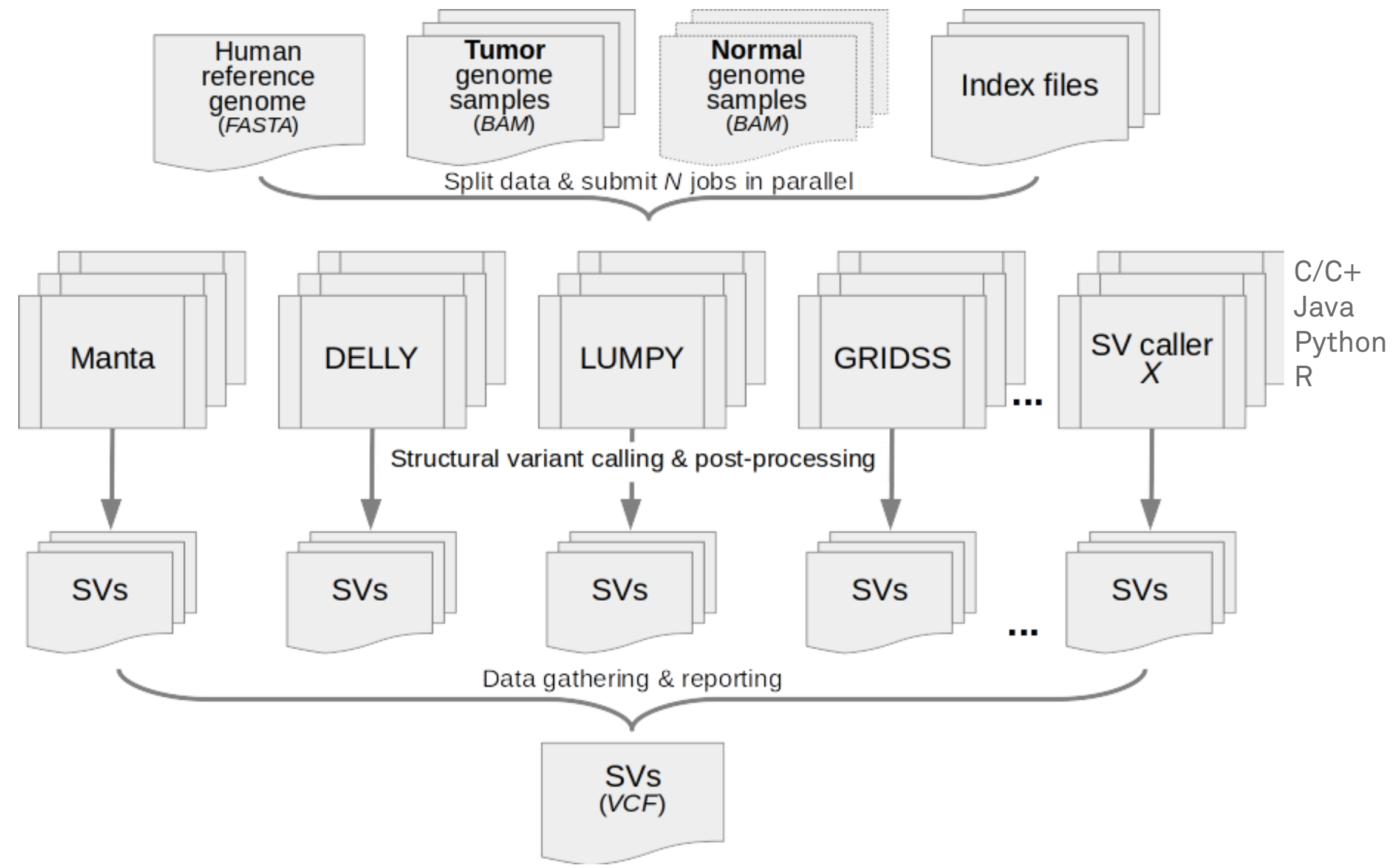
🌐 [www.esciencecenter.nl](http://www.esciencecenter.nl)

# Our expertise in collaborative research

---



# Example from Life Sciences | Highly portable genomics workflow

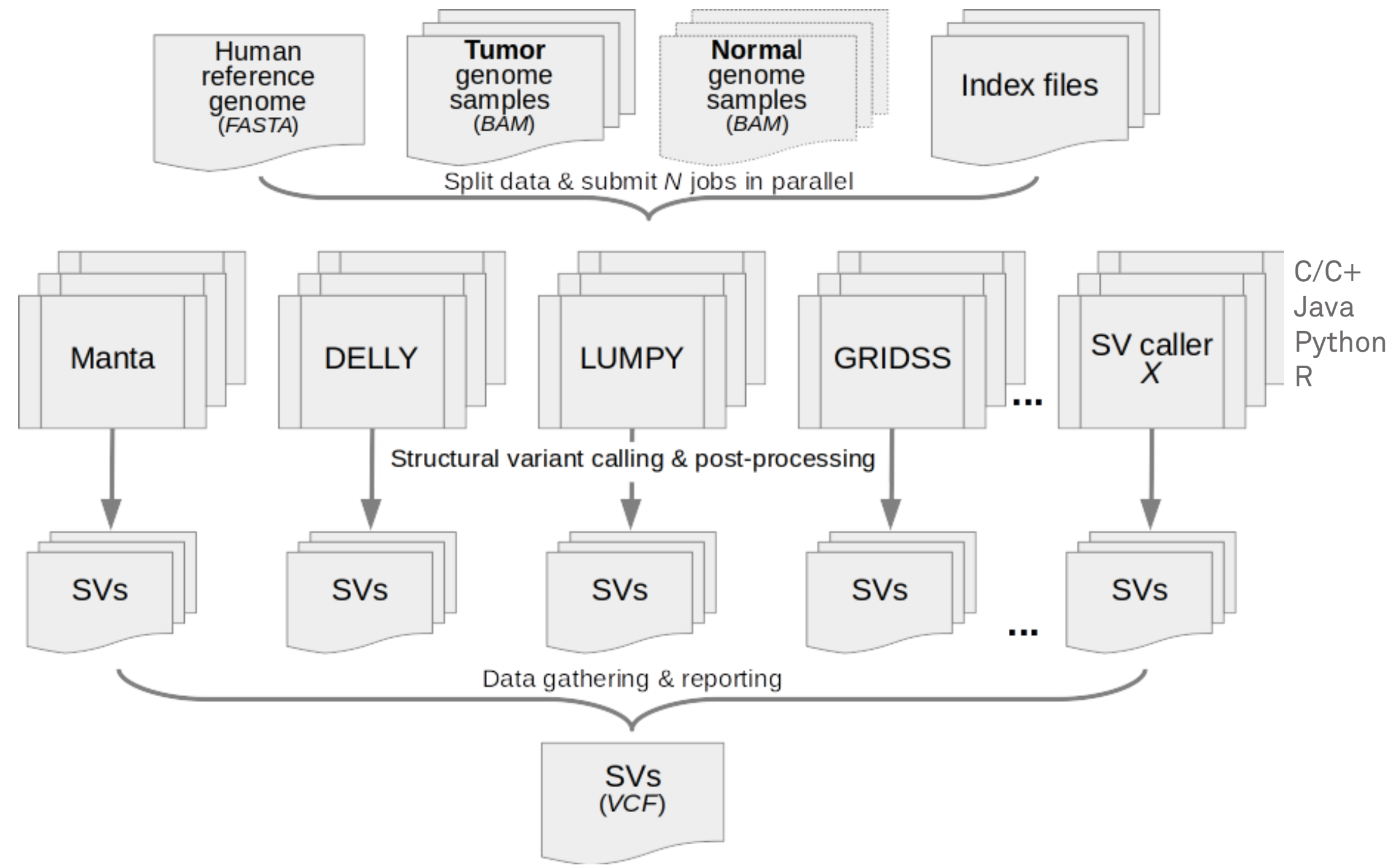


Big challenge!

supercomputer / HPC cluster / cloud



# Example from Life Sciences | Highly portable genomics workflow



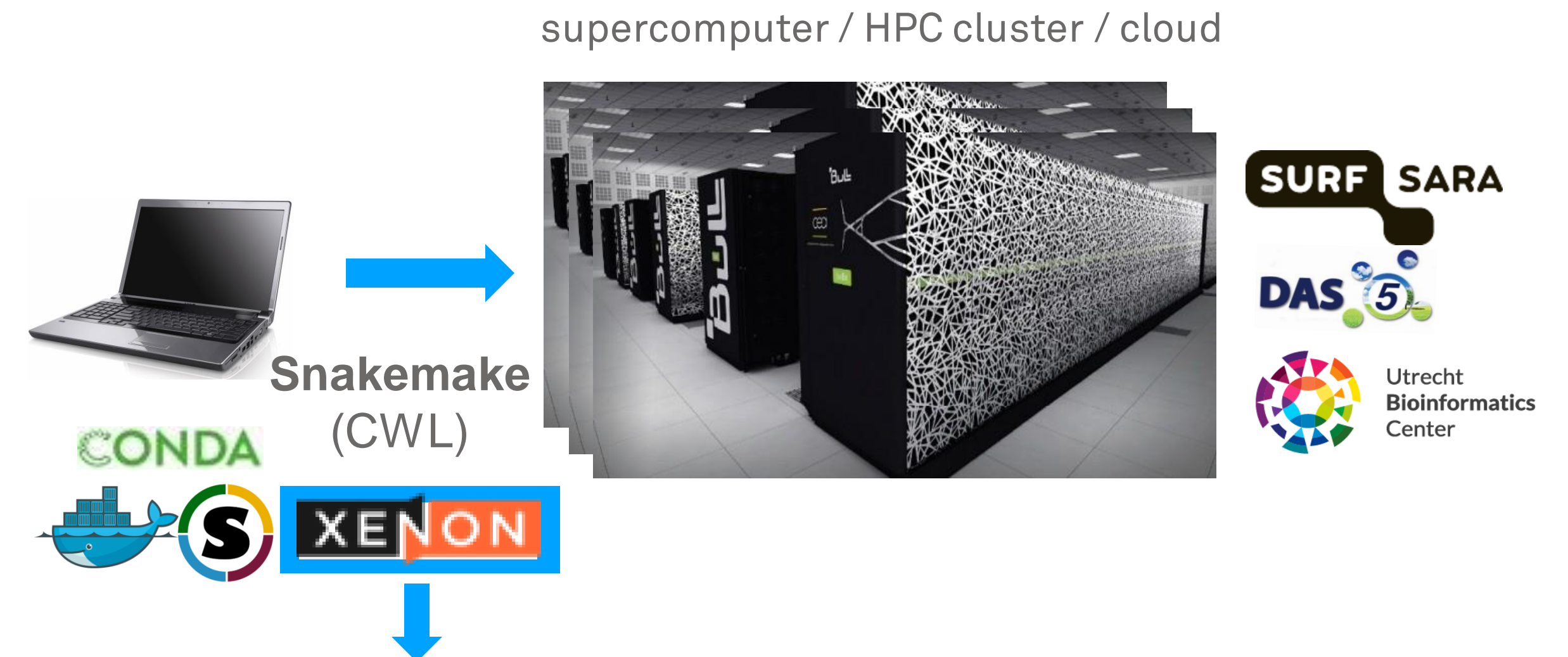
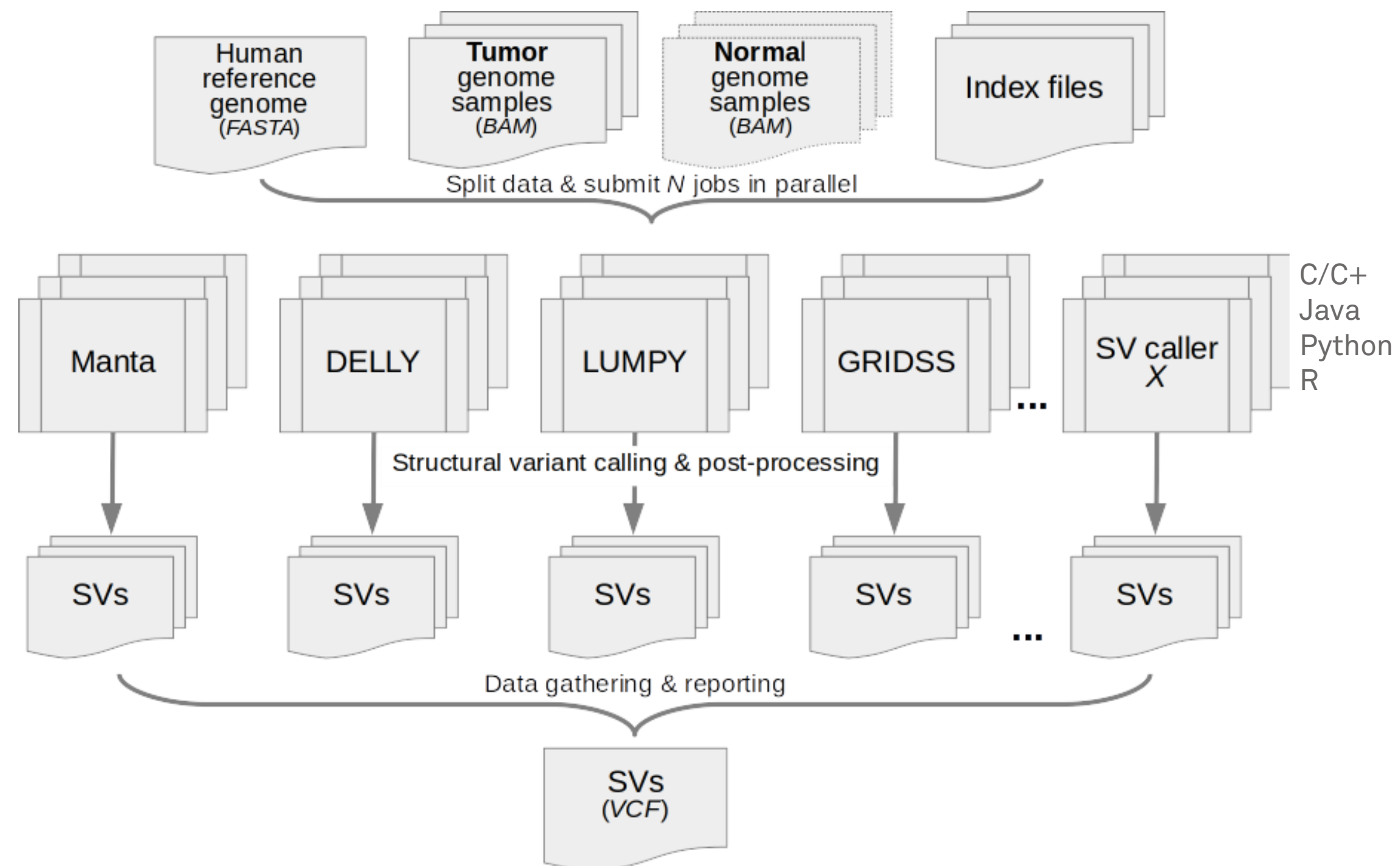
Snakemake (CWL)




supercomputer / HPC cluster / cloud





# Example from Life Sciences | Highly portable genomics workflow





- Enabling portable & reproducible analyses
- Making tools easy to use / extend / deploy
- Following SWE best practices

Googling the cancer genome  UMC Utrecht

Chemo-informatics for metabolite identification  WAGENINGEN UR  
For quality of life

The evolution of start clusters  Universiteit Leiden

Predicting sea level changes  Utrecht University

Mapping the Via Appia in 3D  VU UNIVERSITY AMSTERDAM

...

Software re-use  
@Science center

# It's Open Science...

GitHub repository: **sv-callers** (10 repositories, 8 people, 0 teams, 0 projects). Pinned repositories include **elixir bio.tools** (Search tool and data services registry).

Zenodo page: **sv-callers workflow** (May 6, 2019). Software, Open Access. Author: Kuzniar, Arnold. Description: "It combines several state-of-the-art tools for detecting structural variants... workflow is easy to use and to deploy on any Linux-based machine. In software deployment, easy configuration and addition of new analysis tools..."

Research Software Directory: Encouraging the re-use of research software. URL: <https://research-software.nl/software/sv-callers>

sv-callers: Available versions: 1.0.0. Genomics, Emerging, Command-line. A user-friendly... Tags: Big data (0), GPU (0), High performance computing (1).

**Xenon** (6 mentions, 9 contributors). Description: "If you are using remote machines to do your computations, and don't feel like learning and implementing many different APIs, Xenon is the tool for you."

Get started [link]. 2366 commits | Last update: June 13, 2019.

Cite this software: DOI: 10.5281/zenodo.3244998. Copy to clipboard. Choose a version: 3.0.0. Choose a citation style: BibTeX. Download file.

<https://www.research-software.nl/software/xenon>

2018 IEEE 14th International Conference on e-Science. DOI: [10.1109/eScience.2018.00064](https://doi.org/10.1109/eScience.2018.00064). eScience IEEE INTERNATIONAL CONFERENCE.

## A portable and scalable workflow for detecting structural variants in whole-genome sequencing data

Arnold Kuzniar (Netherlands eScience Center, Amsterdam, The Netherlands, a.kuzniar@esciencecenter.nl), Jason Maassen (Netherlands eScience Center, Amsterdam, The Netherlands, j.maassen@esciencecenter.nl), Stefan Verhoeven (Netherlands eScience Center, Amsterdam, The Netherlands, s.verhoeven@esciencecenter.nl), Luca Santuari (Center for Molecular Medicine, University Medical Center Utrecht, The Netherlands, l.santuari@umcutrecht.nl), Carl Shneider (Center for Molecular Medicine, University Medical Center Utrecht, The Netherlands, c.shneider@umcutrecht.nl), Wigard Kloosterman (Center for Molecular Medicine, University Medical Center Utrecht, The Netherlands, w.kloosterman@umcutrecht.nl), Jeroen de Ridder (Center for Molecular Medicine, University Medical Center Utrecht, The Netherlands, j.deridder-4@umcutrecht.nl).

**Keywords**—whole genome sequencing, cancer genomics, structural variants, workflow, reproducible research, HPC

I. INTRODUCTION  
Cancer affects millions of people worldwide. With the advent of novel DNA sequencing technologies, whole genome sequencing (WGS) has become a standard tool for genomic research. However, WGS data is massive and analyzing it requires specialized software, easy configuration and addition of new analysis tools. Moreover, the workflow's parallel execution environment enables to scale from a single computer to high-performance compute clusters with minimal effort. For this, we used the actively maintained *Snakemake* workflow system [8], *Conda* package manager and the newly developed *Xenon* software suite.

eScience center (Writer). Title: **Portable HPC workflows with Snakemake, Conda and Xenon**. Author: Juriaan H. Spaaks (Follow). Date: Dec 20, 2018 · 4 min read.

**Migrate your scientific workflows to different remote machines in minutes instead of days**

By *Jurriaan H. Spaaks, Jason Maassen, and Arnold Kuzniar*

<https://blog.esciencecenter.nl/>