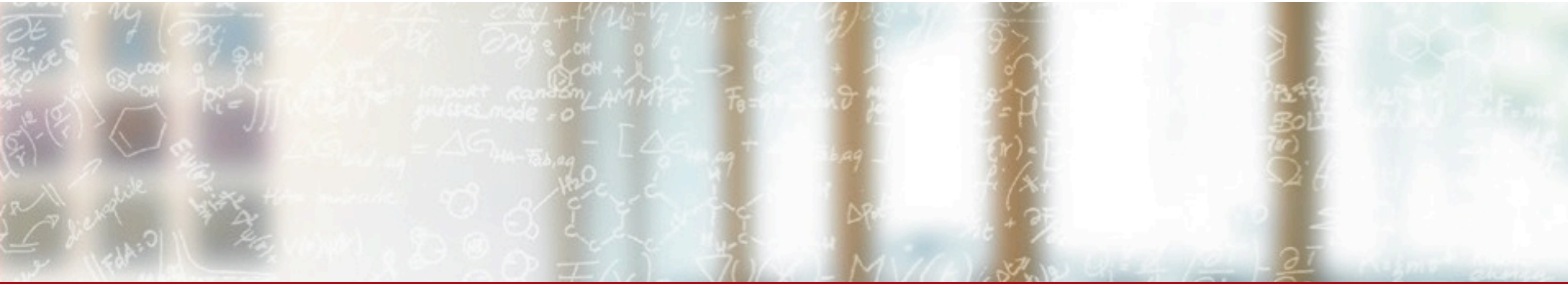




**CSCS**

Centro Svizzero di Calcolo Scientifico  
Swiss National Supercomputing Centre

**ETH**zürich



# Tools and Utilities for Data Science Workloads and Workflows

CUG 2018

Maxime Martinasso, CSCS

May 21, 2018

# Data science at CSCS

## ■ Tools

- Urika-XC
- CSCS modules for: TF, Spark, Theano and Shifter (container)
- ABCpy (Approximate Bayesian Computation): active development

## ■ Users

- Data analytics users:
  - SDSC (Swiss Data Science Centre)
  - USI IDIS (InterDisciplinary Institute of Data Science)
- User Lab interested in data science: simulation + analytics

## ■ Machines

- Large GPU-based system (Piz Daint)
- OpenStack systems

# User perspective – one example



- SDSC RENKU platform
  - RENKU allows scientists to securely *Share/Reproduce/Reuse* Data, Metadata and Code
- RENKU and Jupyter Notebook
  - Spawns notebooks to compute nodes
- Jupyter notebooks with Urika-XC
  - Complex setup to deploy a notebook (ssh, copy url, change ports)
  - Using JupyterHub: requires to write a JupyterHub Spawner

# Looking forward

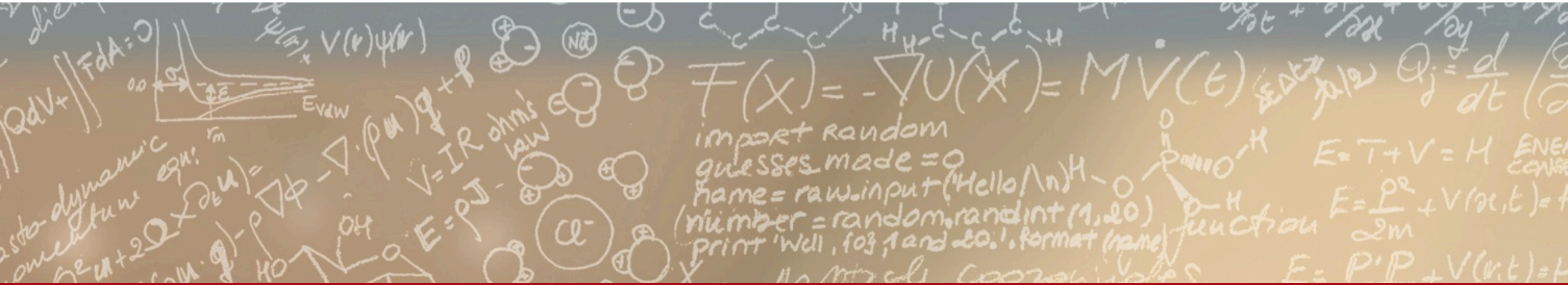
- Provide a workflow service (simulation + analytics)
  - Common Workflow Language?
  - Eclipse?
  - Data management
  
- Promote the usage of containers
  - Decoupled Urika-XC from Shifter?
  - Provide an OCI compliant container solution



CSCS

Centro Svizzero di Calcolo Scientifico  
Swiss National Supercomputing Centre

ETH zürich



Thank you for your attention.