Trusted Research Environment on AWS

Piotr Kasprzak (GWDG)
GÉANT Cloud Framework Workshop, 05.06.2023

Agenda

- Introduction
- It´s Science, not servers – Value of AWS Cloud to research
- EOSC Future Tenders – A joint success in European Research
- Trusted Research Environment (TRE) on AWS
- Questions / Discussion

Introduction

Gesellschaft für Wissenschaftliche Datenverarbeitung Göttingen (GWDG)

An IT service provider to:

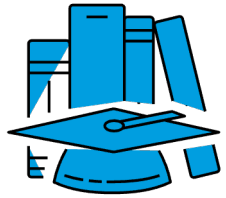
- Max Planck Society (#3 research institution globally, 86 institutes)
- University of Göttingen
- Various academic institutions in Europe

Gesellschaft für Wissenschaftliche Datenverarbeitung Göttingen (GWDG)

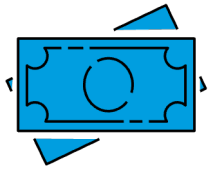
- Takes part and carries out its own research in applied computer science and HPC
- One of eight national Centers for High Performance Computing (NHR Centers) and one of two HPC sites of the German Aerospace Centre (DLR)
- Core player in German Research Data Initiative (NFDI) taking part in different domain specific projects and contributing to core services
- One of four national AI service centers targeting AI applications in the fields of medicine and energy

It´s Science, not servers – Value of AWS Cloud to research

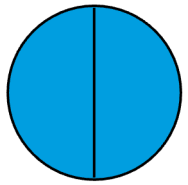
Researcher challenges and pain points



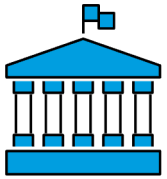
Pressure to publish



Competition for funding

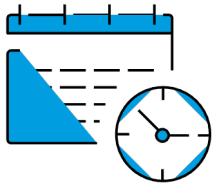


Cross-institution collaboration



Complicated compliance

Difficulty leveraging large datasets



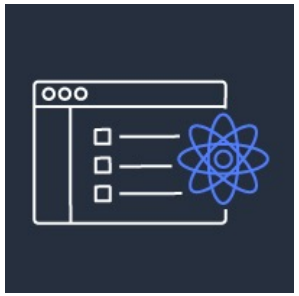
Limited technical knowledge

Cost

What is a researcher looking for?

Science, not servers
Use compute when you need
It to do large-scale analysis

Collaborate and share
Access data sets that span institutions



Share effort
High compatibility with existing workflows

Reproduce research
A common platform for reproducing scientific analyses

State-of-the-art analytics
Ability to run the latest analysis tools

Freedom
No constraints on choice of tools and methods

Why AWS Cloud – A GWDG perspective

- More and more demand from researchers
  - Established workflows that utilize AWS services
  - Collaborations with researchers from other countries (e.g. USA) where use of cloud services is widespread
  - Easy scalability, ease-of-use
- Increasing number of services cannot be provided anymore on-prem
  - DL models
  - Highly sophisticated or very specific services (e.g. AWS Mechanical Turk)
- The future will be hybrid: use cloud and on-prem services at the same time based on scalability, cost and functional requirements

EOSC Future Tenders –
A joint success in European Research

EOSC Future Tenders – A joint success in European research

- EU funding to drive the adoption of commercial cloud services in European research
- Tender addressed European research aggregators
- GWDG proposal supported by AWS & OCRE partner Rackspace:
  - Develop proposal idea
  - Define scope of project
  - Technical guidance and support (e.g. AWS architecture design)
  - Cloud calculation
  - Value for money and more

EOSC Future Tenders – A joint success in European research

- GWDG was awarded with EOSC funding early 2023
- GWDG will implement a Trusted Research Environment (TRE) to secure and analyze sensitive data in the AWS cloud
- Award includes funding to onbaord first research projects on TRE
- TRE will be accessible to European researchers through the EOSC marketplace

Trusted Research Environment (TRE) on AWS

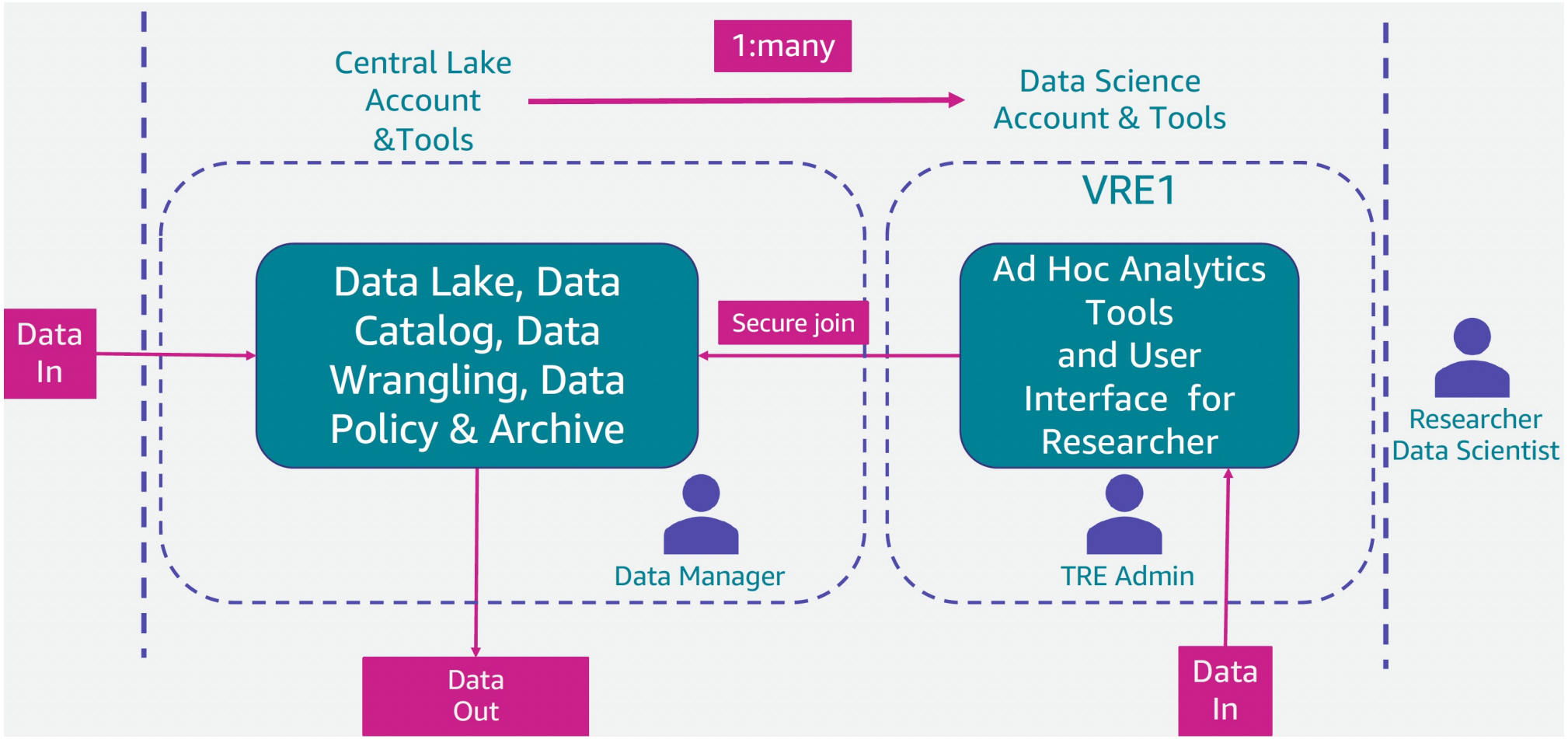
What is Trusted Research Environment on AWS?

- An open source self-service research solution to secure and analyse sensitive data
- Two main components
  - Data lake (AWS Lake Formation)
  - Virtual Research Environment (VRE) based on Service Workbench on AWS
- Leverages AWS security model & broad range of AWS tools
- Based on customer requirements & recognized frameworks
- Repeatable

Secure data analysis environments

The Five Safes Framework	
Safe People	Trained and accredited researchers trusted to use data appropriately
Safe Projects	Only used for valuable, ethical research that delivers clear public benefits
Safe Data	Researchers can only use data that have been de-identified
Safe Setting	Access to data is only possible using our secure technology systems
Safe Outputs	Outputs are checked to ensure they cannot identify data subjects

Architecture



Thank you!

Fragen / Diskussion