



Data Cage

for EstBB data processing

BY PRIIT KLEEMANN

ESTONIAN BIOBANK

About the presentation

- ▶ Quick overview of the constraints – The „problem“
- ▶ About the solution – The Data Cage and business use case
- ▶ Challenges

Any problems?

- ▶ Legal context
 - ▶ The General Data Protection Regulation (GDPR)
 - ▶ Estonian Human Genetic Research Act
- ▶ Data context - special categories of personal data
 - ▶ Genetic data
 - ▶ Medical data
 - ▶ Lifestyle data
- ▶ The fact of being gene donor must be **secret**
- ▶ Data breach could be fatal for EstBB

Business requirements

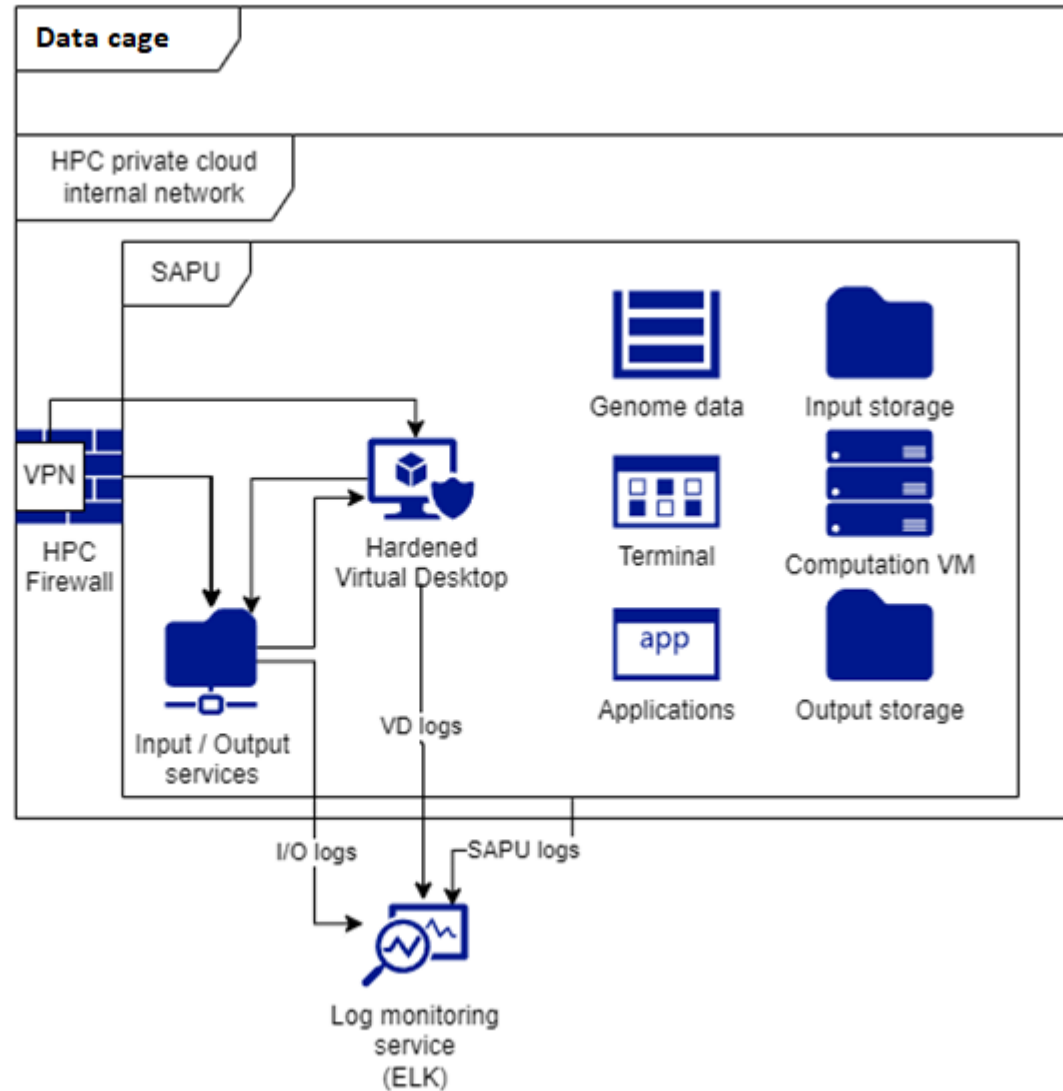
- ▶ Find computational answers to research questions
- ▶ Run custom algorithms to find answers to the research question
- ▶ Be able to use new and „legacy“ software versions
- ▶ Receive research results from secure environment

Business requirements

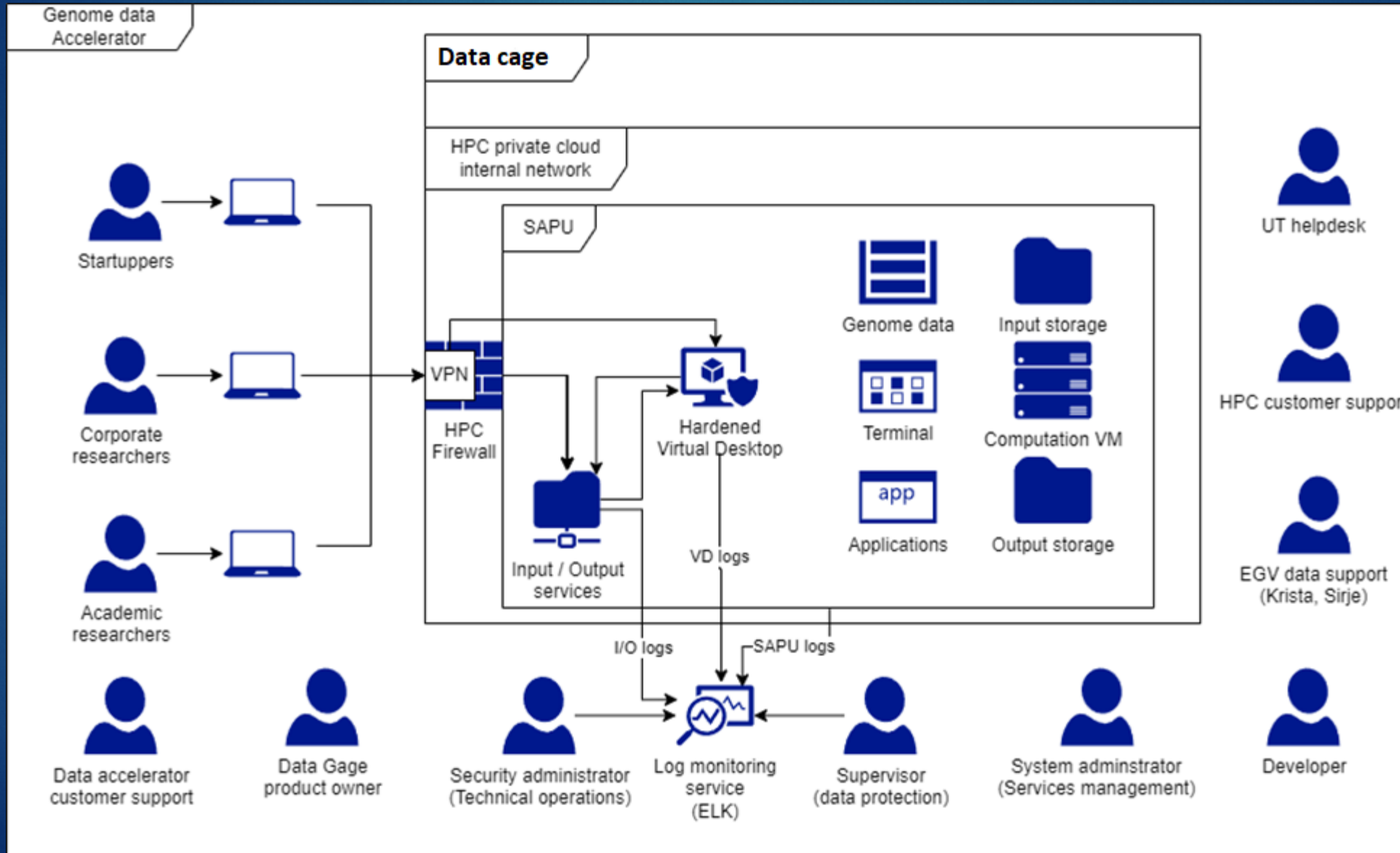
- ▶ Find computational answers to research questions
- ▶ Run custom algorithms to find answers to the research question
- ▶ Be able to use new and „legacy“ software versions
- ▶ Receive research results from secure environment

- ▶ EstBB data **shall not** be exfiltrated

The scope of Data Cage



How Data Cage is provided



Tech stack

- ▶ CentOS 7 virtual machines
- ▶ XRDP remote server
- ▶ Guacamole XRDP as HTML5 proxy
- ▶ ELK stack for log management
- ▶ FreeIPA identity management and Keycloak for OpenID Connect
- ▶ Object storage for data-in and data-out
- ▶ Firezone VPN

Challenges

- ▶ What identity provider to use?
- ▶ What are the **real** risks?
- ▶ Do we spend over or under to control the risks?
- ▶ Reaction time to patch new security flaws
- ▶ Environment usability for researchers (end user experience)
- ▶ Discover data misuse or breaches from the logs
- ▶ Customer willingness to pay for using the data gage



Thank you for your valuable time!

Questions and discussion