

Accelerated big data systems: from application to computation

Zaid Al-Ars & Peter Hofstee

z.al-ars@tudelft.nl

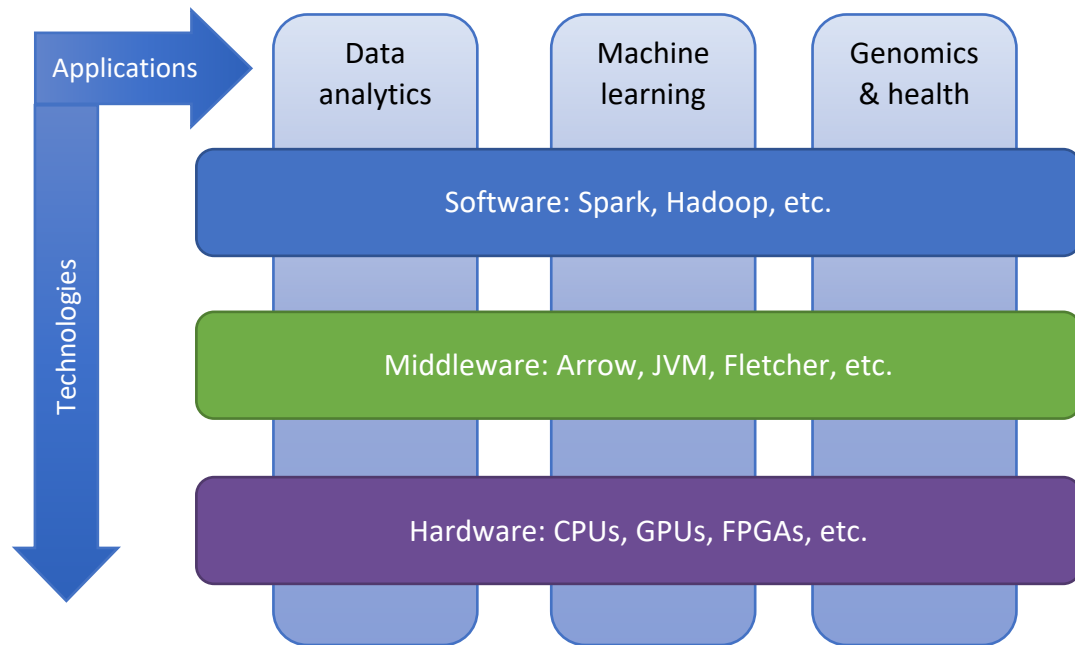
Accelerated Big Data Systems

Delft University of Technology

Acceleration enablement at various levels

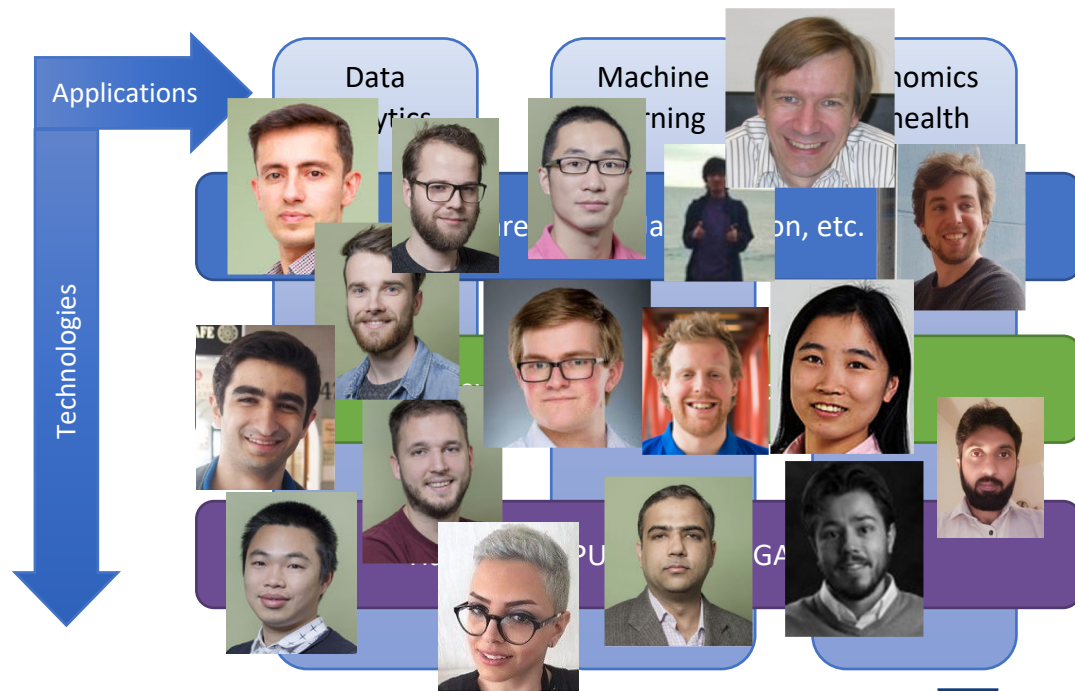
- Objective is to demo technology to accelerate data center applications in critical domains

- Data analytics
- Machine learning
- Genomics



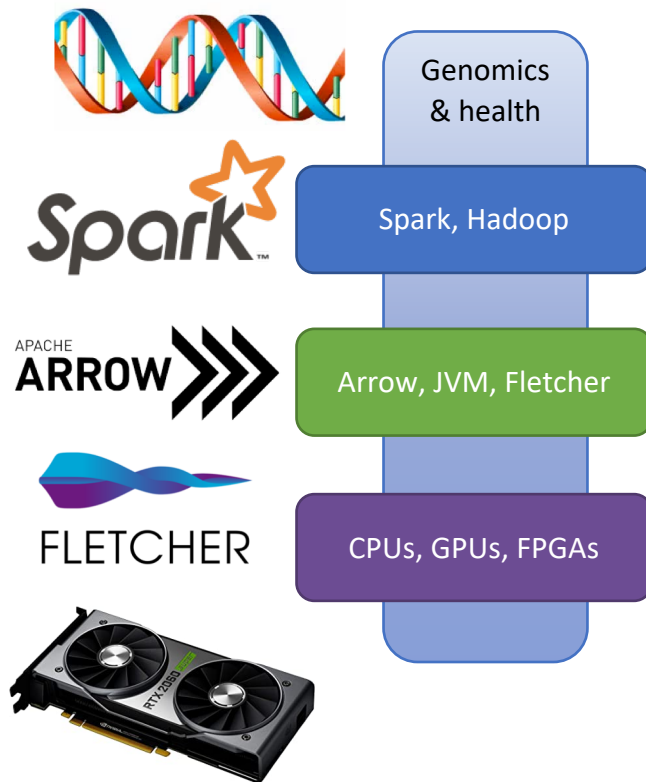
Large academic-industry effort

- Large effort of 15 employees and 20 students
- Together with industry
 - IBM, Xilinx, Nvidia
- & medical centers
 - LUMC, EMC, DKFZ



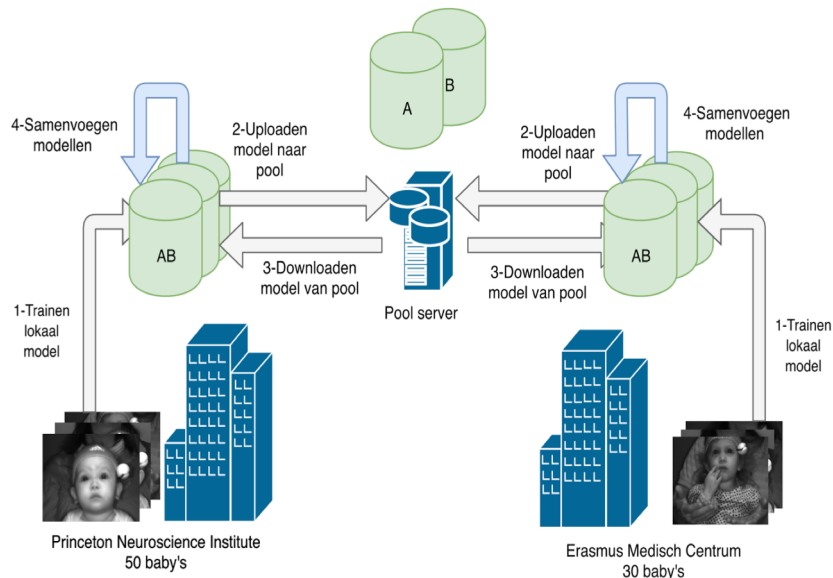
Acceleration MSc topics in genomics

- Genomics is largest source of big data
- Applications in cancer hereditary disease diagnosis
- Genomics acceleration on GPUs and FPGAs
- Big data scalability using Apache Spark
- In-memory computing using Arrow and Fletcher



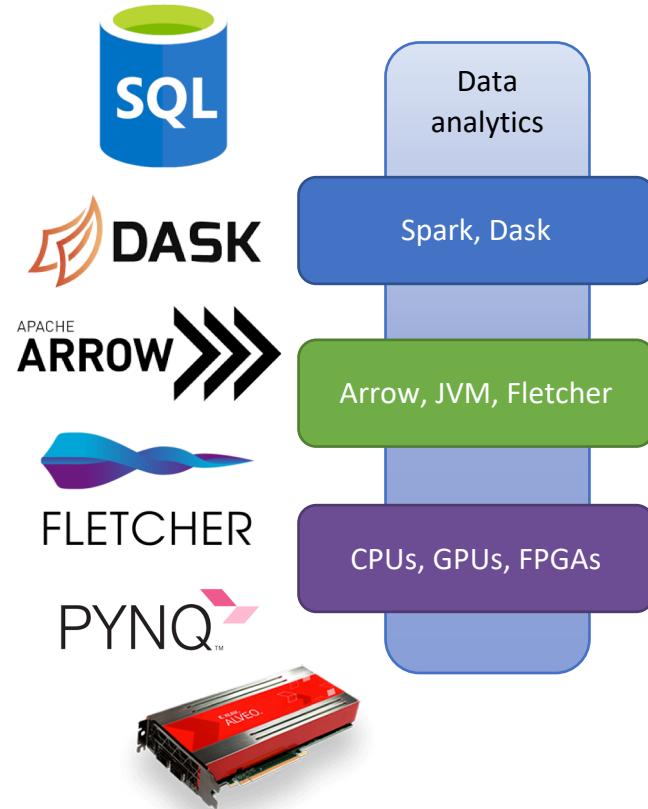
Machine learning MSc thesis topics

- Federated ML to protect privacy in medical applications
 - Local training on the edge, centralized aggregation
 - No individual data transfer
 - Stroke identification & MRI image cleaning



Acceleration MSc topics in data analytics

- General purpose processor architecture for big data analytics
- High throughput map-reduce with 100s of cores
- Deployment on FPGA for low-latency streaming



Questions:

Johan Peltenburg, Zaid Al-Ars, Peter Hofstee

j.w.peltenburg@tudelft.nl, z.al-ars@tudelft.nl

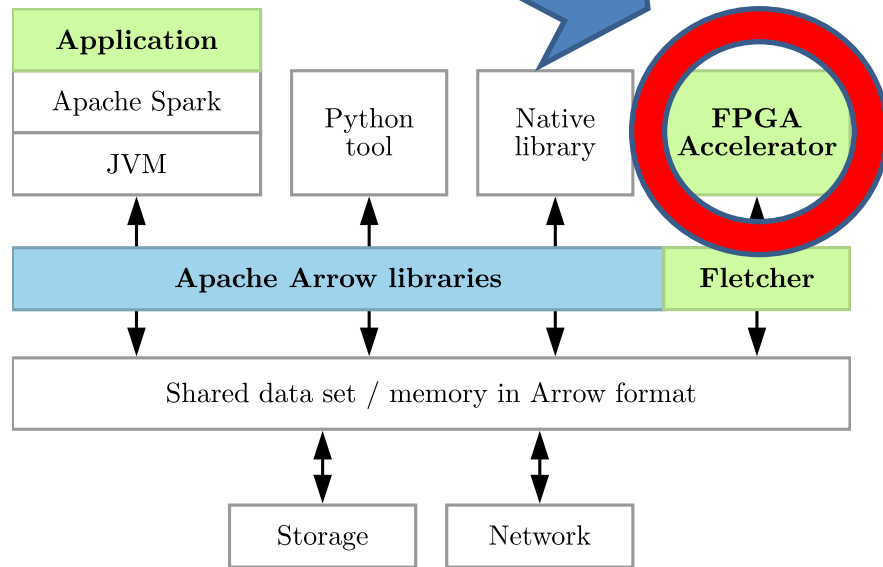
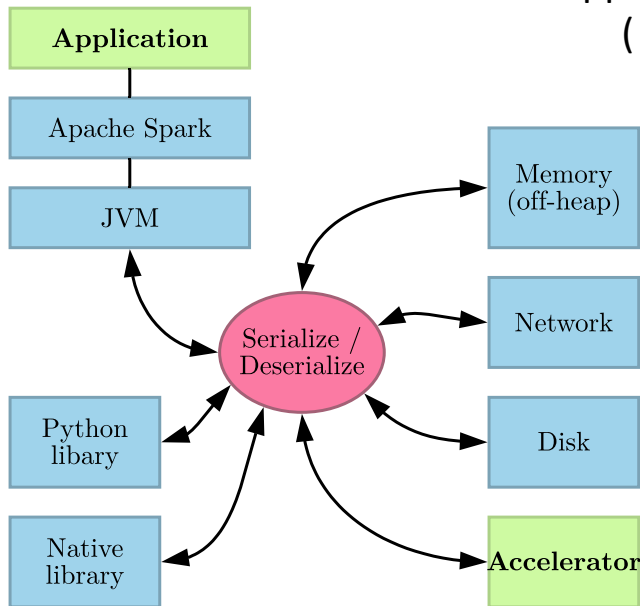
Accelerated Big Data Systems

Delft University of Technology

Apache Arrow & Fletcher

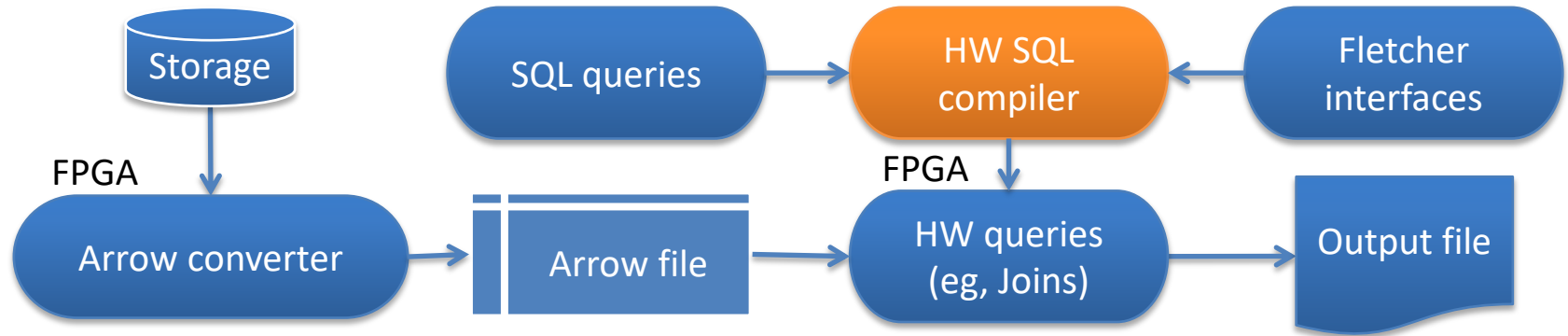


Opportunity to do new and cool things!
(Without a lot of overhead.)



Current implementation of Fletcher on POWER9/CAPI 2.0 supports 10GB/s processing!
(5x Amazon EC2 F1)

Database acceleration use case



- Automated high level SQL script acceleration
- HW compiler to accelerate
 - SQL queries execution kernels
 - Integrated Fletcher interfaces
- Fully automated data path design on FPGA