

HELIX TO HEALTH – BIG DATA CHALLENGES IN GENOMICS

contact: Anil Kurella, Aruna Kumar, Monica Martinez-Canales, Abhi Basu, Stefano Guerra

Problem statement: As genome sequencing becomes affordable computational analysis rather than sequencing will be the rate limiting factor in personalized medicine.

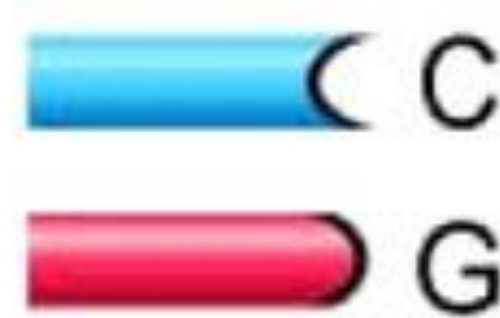
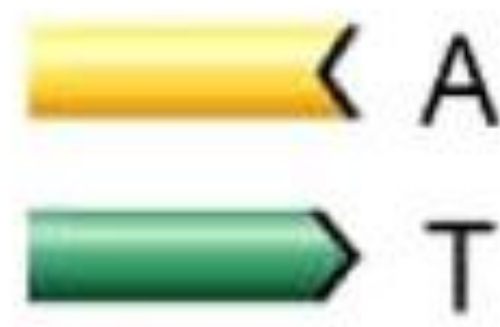
Why sequencing is a
Big data
problem?

Humans have
 3.2×10^9 Base Pairs

Canopy Plant has
 150×10^9 Base Pairs



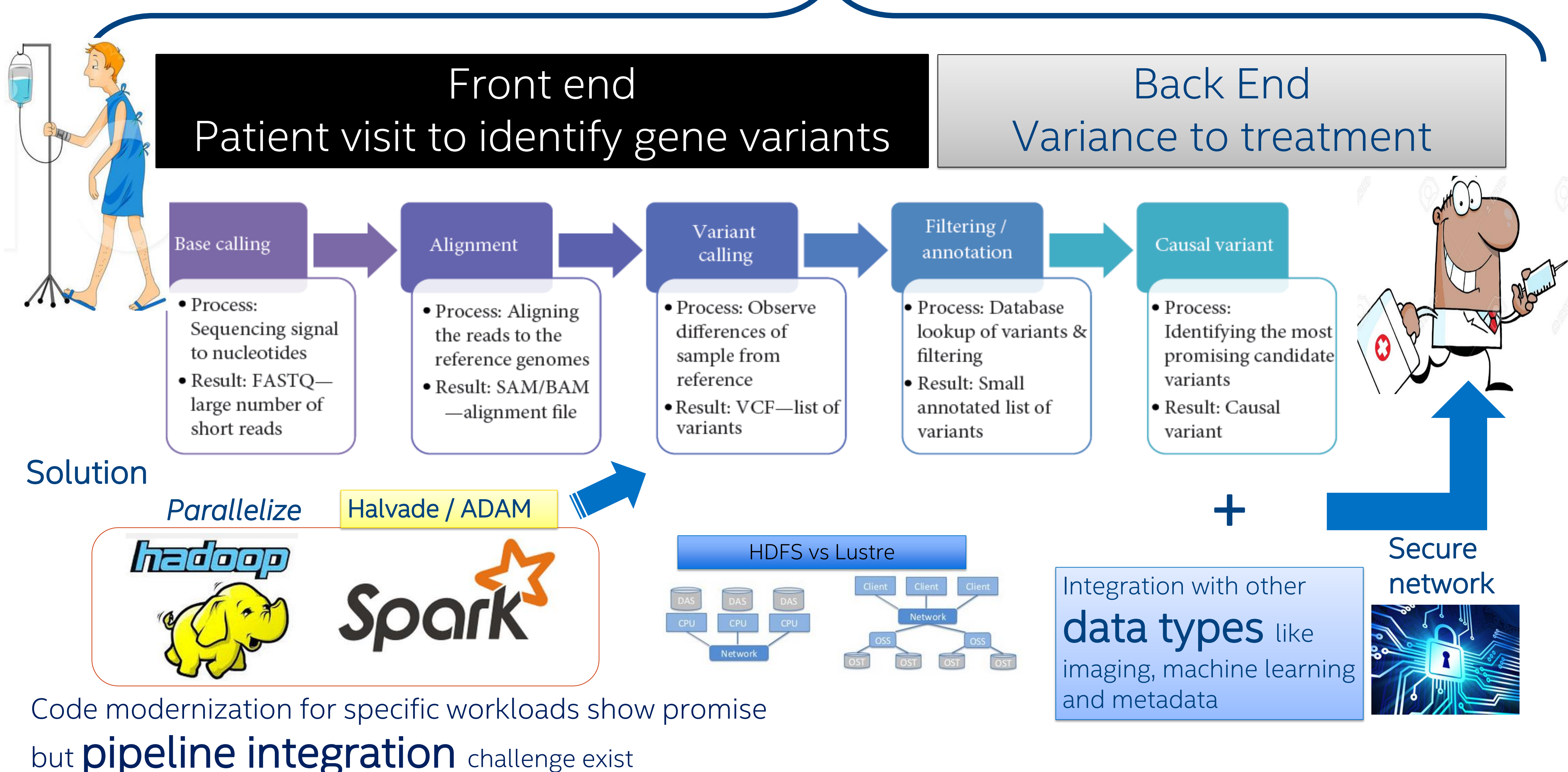
Sequencing refers to the exact order of base pairs in a segment of DNA



Storage by 2025 genomics would generate **40×10^{18}** bytes more than YouTube or Twitter

Big data **traffic jam-**
20 days to transfer 50 genomes between an analyst & customer

Current end to end **Computation** takes **days** with single nodes



Low cost end to end solution stack is the need of the hour- lets collaborate