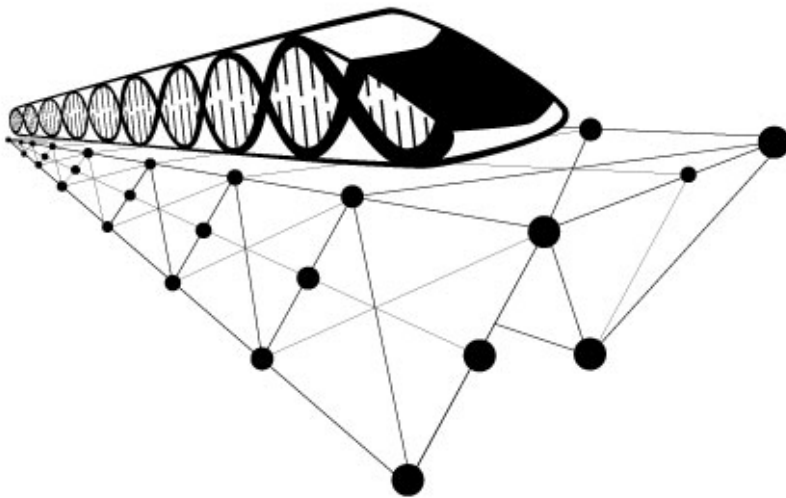


KNOWeng

BIG DATA TO KNOWLEDGE
CENTER OF EXCELLENCE

(PRONOUNCED “KNOWING”)



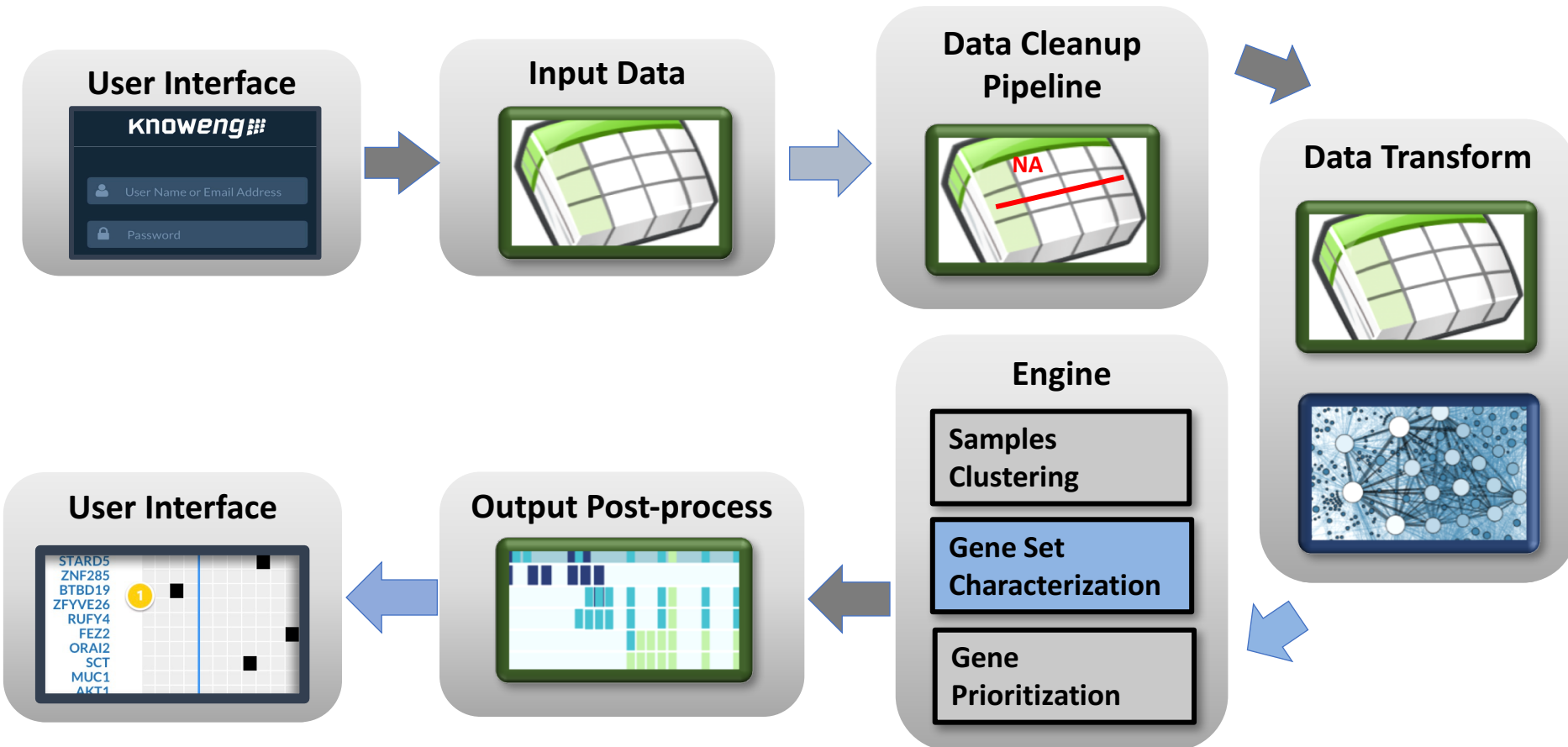
Knowledge Engine for Genomics

Jing Ge

Research Programmer

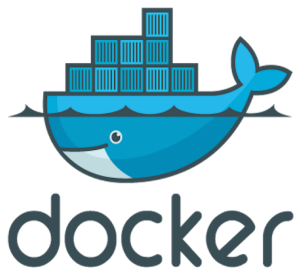
National Center for Supercomputing
Applications

System Infrastructure



System Infrastructure

- We encapsulated each analysis task in **Docker containers** for easy execution
- Whose execution is orchestrated by the **Chronos** job scheduling framework
- On a compute cluster managed by **Apache Mesos**
- That syncs user and community data through **AWS S3**



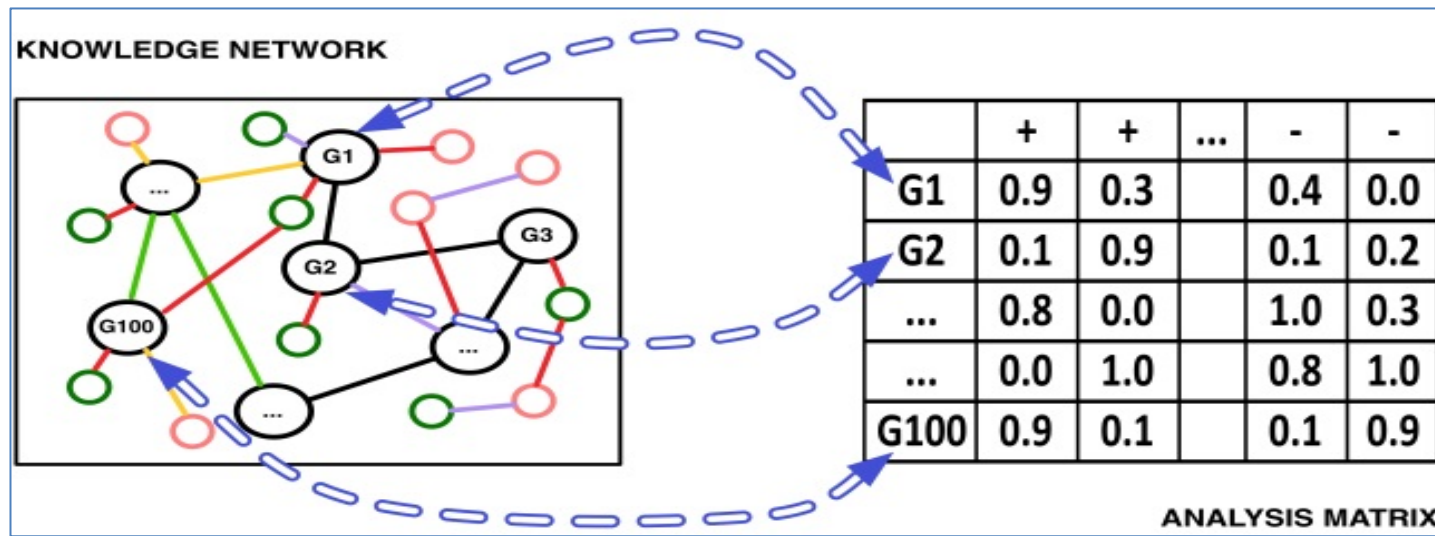
Apache
MESOS



GitHub

Knowledge Network

- Knowledge network: a heterogeneous graph whose nodes and edges represent genes/proteins, their properties and relationships.
- We want knowledge network to guide the pipeline



Thank you !

