Optimizing serverless analytics for molecule identification

Germán T. Eizaguirre PhD Student Universitat Rovira i Virgili

Thesis supervisor Marc Sánchez-Artigas Universitat Rovira i Virgili Empresarial tutor Bernard Metzler IBM Research Zurich

Special thanks Animesh Trivedi, Radu Stoica IBM Research Zurich









ADAPTIVE VIRTUALIZATION FOR AI-ENABLED CLOUD-EDGE CONTINUUM

The METASPACE metabolomics annotation pipeline

- European Molecular Biology Laboratory (EMBL).
- Production-ready and publicly available. Hosted in AWS.
- Built on Cloud functions.

www.metaspace2020.eu





Goals

- 1. Overall performance (execution time) optimization.
- 2. Smart resource provisioning.

Is the number of functions relevant?

Example: METASPACE preprocessing phase.

• All-to-all communication.





Is the number of functions relevant?



- Parent != child computation times.
- Cloud object storage systems provide asymmetric read/write performance.



- We take provisioning decisions online, at each dependency, immediately before the exchange operation.
- We use an analytic model of IO, CPU and invocation time.

Preliminary results

Thank you!

Germán T. Eizaguirre PhD Student Universitat Rovira i Virgili

Thesis supervisor Marc Sánchez-Artigas Universitat Rovira i Virgili Empresarial tutor Bernard Metzler IBM Research Zurich

Special thanks Animesh Trivedi, Radu Stoica IBM Research Zurich

ADAPTIVE VIRTUALIZATION FOR AI-ENABLED CLOUD-EDGE CONTINUUM