Software containers for scientific applications on clouds

Mikołaj Baranowski, Marian Bubak, Adam Belloum

Informatics Institute, SNE group

Email: baranowski@uva.nl

November 3, 2015
Past work

Building data workflows from application sources¹

- Optimization based on application syntax for application model based on general-purpose programming language
- Transformations to workflow applications (workflow engines)
- Visualization
- Provenance data

Abstraction of MapReduce jobs²

- Is it possible to abstract logic – map and reduce routines – from implementation – data storage and MapReduce implementation
- Abstraction of data types and data transformations


Research questions

Scientific applications issues:

▶ **Reproducibility** – scientific software is maintained only during the experiment and usually not supported later, often used only in one environment
  
  ▶ if it is impossible to maintain it – we have to find a way to preserve scientific applications for further usage and experimentation
  
▶ Difficulties accessing data – diversity of libraries, standards

▶ Modules are closely coupled, big number of dependencies, hard to share applications

▶ **Hide complexity** behind an abstraction

Cloud computing issues (from a science perspective):

Many **data sources** Diversity of **libraries** **Authentication** methods

We addressed some of these questions³

Cookery

Cookery is a framework for building scientific applications

- High abstraction – Cookery language, Jupyter notebook
- Malleable – 3 layer architecture, each assigned to a user role
- Designed for a distributed, cloud environment

Data abstraction – subjects
Process abstraction – actions

github.com:
- Cookery: mikolajb/cookery
- Jupyter kernel: mikolajb/icookery
language-test.cookery

Lang = google-prediction Text.
send-email {to: baranowski@uva.nl, topic: "Language test"}. 
Software containers

Technologies (e.g., Docker and its growing ecosystem) and methodologies of software development

- fast deployment
- micro-services
- People agreed to some standards
- Popularity

Only Linux but maybe not for long – Open Container Project targets standardization and portability among platforms

It gives new capabilities for reproducibility in science

---

Introduction to architecture

Service provider: supplies Cookery actions or subjects and software container: **Cookery layer 2 + Docker container**

User: executes Cookery application in **Jupyter notebook** (locally or remotely), Cookery asks for permissions to access:

- Existing services e.g. Google prediction API – user authorizes Cookery
- Complex services distributed with Docker files – Cookery asks for a permission to deploy them on user’s infrastructure
Architecture

Repository
- Cookery layer 2
- Cookery layer 2
- Cookery layer 2
- Cookery layer 2

Jupyter notebook
Cookery

User

Deploying

Container engine

Cloud services
Future work

**Separation of application logic from its implementation** and its theoretical foundation

Data in reproducible research

Questions?