IFC workshop: “Data Science in Central Banking: Applications and tools”
14-17 February 2022, BIS Basel, Switzerland

Containerization for Research Collaboration

Platform Independent Economics

Venkat Balasubramanian, Danielle Handel, Anson Ho, Kim Huynh, David Jacho-Chávez, Carson Rea

Previous work

IFC Workshop Part 1: Data Science in Central Banking: Machine learning applications

Cloud Computing Research Collaboration: An Application to Access to Cash and Financial Services

Danielle V. Handel, Anson T. Y. Ho, Kim P. Huynh, David T. Jacho-Chavez, Carson H. Rea

Abstract

We illustrate the utility of cloud computing tools for big data management and analysis serving the functions of the Bank of Canada. These tools provide the opportunity to easily leverage increasingly complex and large-scale data in an interactive coding environment without worrying about backend infrastructure. As an empirical use case to demonstrate these advantages, we use a cloud computing platform to expedite a computationally intensive spatial analysis mapping access to financial services in Canada.

Keywords: High-Performance Computing; Big data; Spark; Jupyter.

Introducing: the "**but it works on my machine**" problem

See "Use of Docker for Reproducibility in Economics," from AEA data editor ***Lars Vilhuber***

Containers Offer a Solution

A *container* is an executable package of software that includes everything needed to run an application: code, runtime, system tools, system libraries and settings

Solves the "**works on my machine**" problem*

REPRODUCIBILITY

PORTABILITY

COLLABORATION

Thank you to Bruno Tissot for highlighting this pathway

Use Case

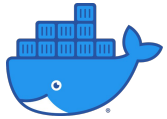
Access to Cash and Financial Services

From our previous work: mapping consumers and their nearest bank branch

- Challenge: **24,000+** postal codes
- Computed straight line distances from population centroids to financial institutions
- Run with PySpark docker image
    - Port to Microsoft Azure and AWS EC2 to manage computational needs

Repro repo on GitHub with all necessary input files

What is Docker?



Industry standard tool for the creation and deployment of containers

FILE

Dockerfile: instructions for image assembly

BUILD

Construct an image based on instructions in a Dockerfile

SHIP

Container image: package with all dependencies needed to run an application

RUN

Container: active instance of a Docker image

SHARE

Docker Hub repositories

Constructing our Docker Container

Mirror executable(s) from GitHub

Load a prebuilt PySpark image

Execute relevant code and exit with success

Execution Across Cloud Platforms


Containerization for Research Collaboration
Platform Independent Economics
Azure
aws

Singularity as an Alternative

- Singularity provides the same functionality as Docker
- Load images from Docker
- Better suited for HPCs
- Only compatible with Mac OS and Linux

Added Value

AEA Data Editor

Repro repo on GitHub

COLLABORATION

EFFICIENCY

REPRODUCIBILITY

PORTABILITY

Thanks/Merci

Venkat Balasubramanian
balv@bank-banque-canada.ca

Danielle Handel
dvhandel@stanford.edu

Anson Ho
atyho@ryerson.ca

Kim Huynh
khuynh@bank-banque-canada.ca

David Jacho-Chávez
djachocha@emory.edu

Carson Rea
chrea@emory.edu