Persistent Reproducible Reporting

Nan Xiao
Genomic Data Scientist, Seven Bridges
Reproducible Research
R Markdown + knitr to the rescue
Reproducibility

... has always been a concern in both academia & industry.
Hundreds of automated analysis workflows for petabyte-scale data from The Cancer Genome Atlas.

Cancer Genomics Cloud
www.cancergenomicscloud.org
Product & Tech Innovations in CGC

1. User interacts with Seven Bridges.
2. Seven Bridges platform is launched.
3. Cloud provider infrastructure is utilized.
4. Worker instances are created.
5. Data sharing between instances is enabled.
6. Encrypted storage buckets are used.

Rabix

Common Workflow Language
liftr

OS-level reproducibility & persistency for reports.
liftr extends the R Markdown metadata format, introducing additional options for containerizing and rendering reports.

By running `lift()` on the RMD file, liftr parses the metadata fields appeared in the R Markdown document; then generates the Dockerfile.

By running `render_docker()`(), liftr will build the Docker image, run the container, and render the R Markdown document.
library("liftr")
input = "demo.Rmd"

lift(input) # Generate Dockerfile
render_docker(input) # Render report with Docker

purge_image(input) # Clean up Docker image
push_image(input) # Push image to registry (devel)
Demo: RNA-Seq Data Analysis

Example workflow from Bioconductor

- RNA-Seq: biotechnology for measuring the expression of genes. It can help identify key genes in cancer.

- TBs of RNA-Seq data are generated. Computational tools and workflows are developed to analyze such data.

- How to ensure such reports are reproducible through time, when datasets, analysis tools are both evolving?

Step 1

Add liftr metadata to the R Markdown document: base image, system dependencies, package dependencies, etc.
Step 2

Use `liftr::lift` to generate Dockerfile
Step 3

`liftr::render_docker` will build the image, run the container, and render into PDF/HTML/Docx.

Re-compilation: cached image layers are used to improve speed.

Remove the used image, or push to registry.
Future works

- Cloud-based rendering and containerization services for dynamic documents
- Democratize reproducible report creation/sharing
Thank You!

liftr.me

@road2stat
#dockercon  #liftr