Enterprise R Platform

eRum 2016

Motivation

The “Joel Test” for Data Science

1. Can new hires get set up in the environment to run analyses on their first day?
2. Can data scientists utilize the latest tools/packages without help from IT?
3. Can data scientists use on-demand and scalable compute resources without help from IT/dev ops?
4. Can data scientists find and reproduce past experiments and results, using the original code, data, parameters, and software versions?

Motivation

The “Joel Test” for Data Science

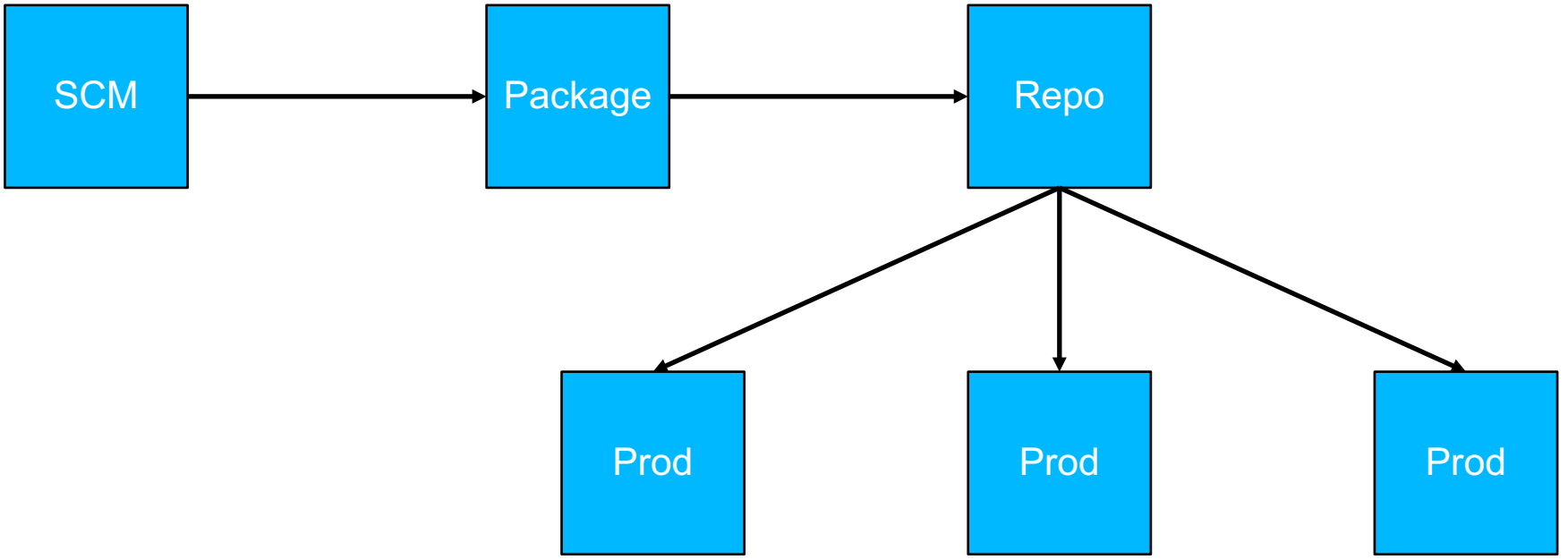
5. Does collaboration happen through a system other than email?
6. Can predictive models be deployed to production without custom engineering or infrastructure work?
7. Is there a single place to search for past research and reusable data sets, code, etc.?
8. Do your data scientists use the best tools money can buy?

Source: https://blog.dominodatalab.com/joel-test-data-science/

+1 Can you access your statistics without using R?

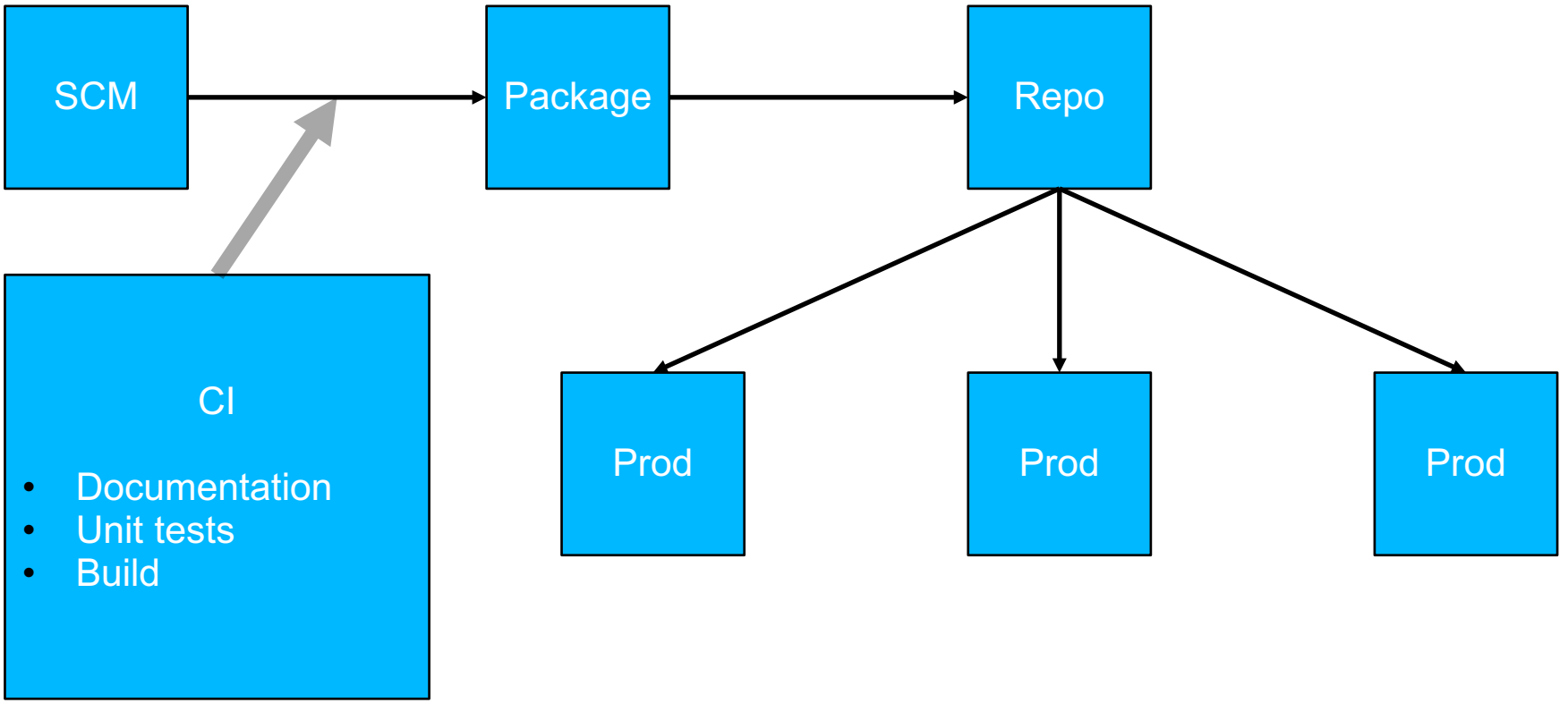
Ideal Process

From code to production



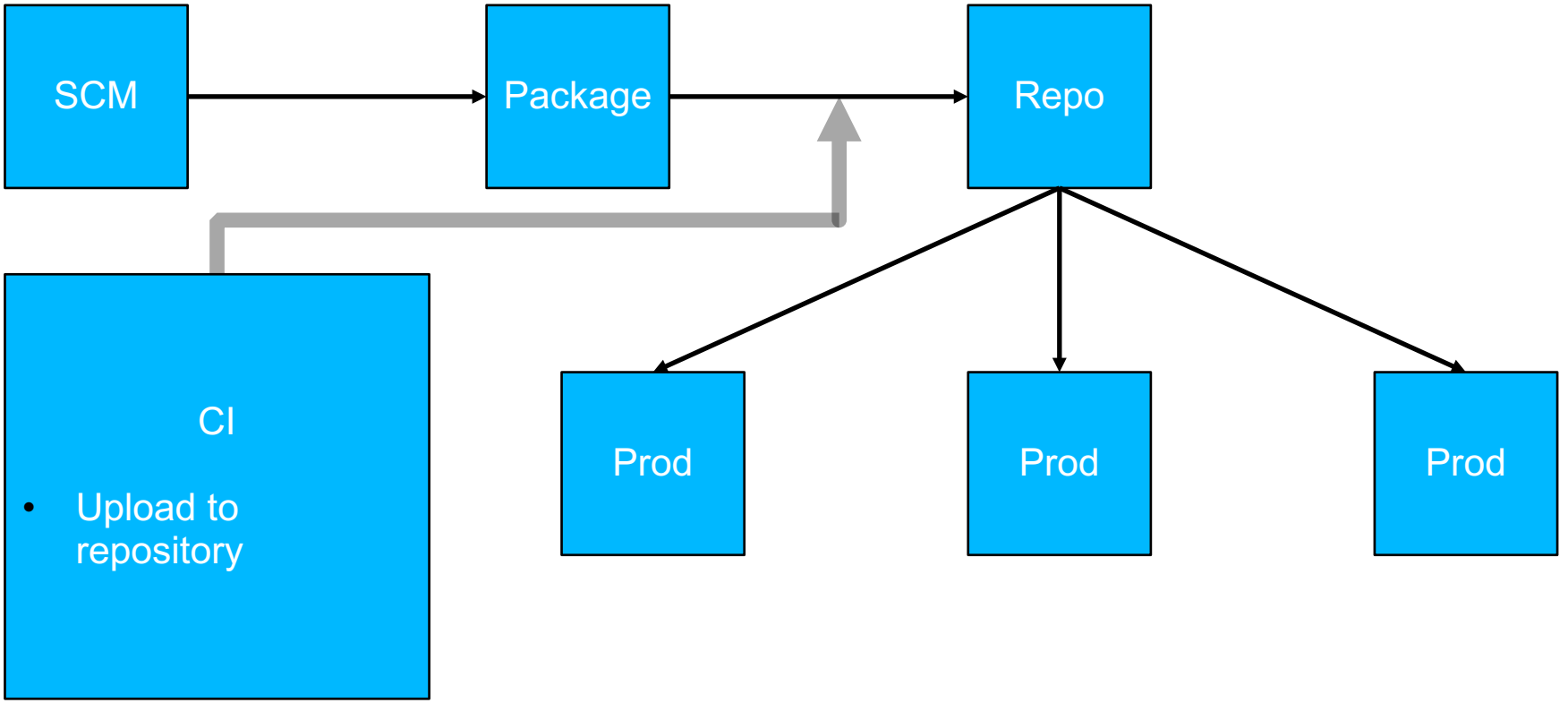
Ideal Process

From code to production



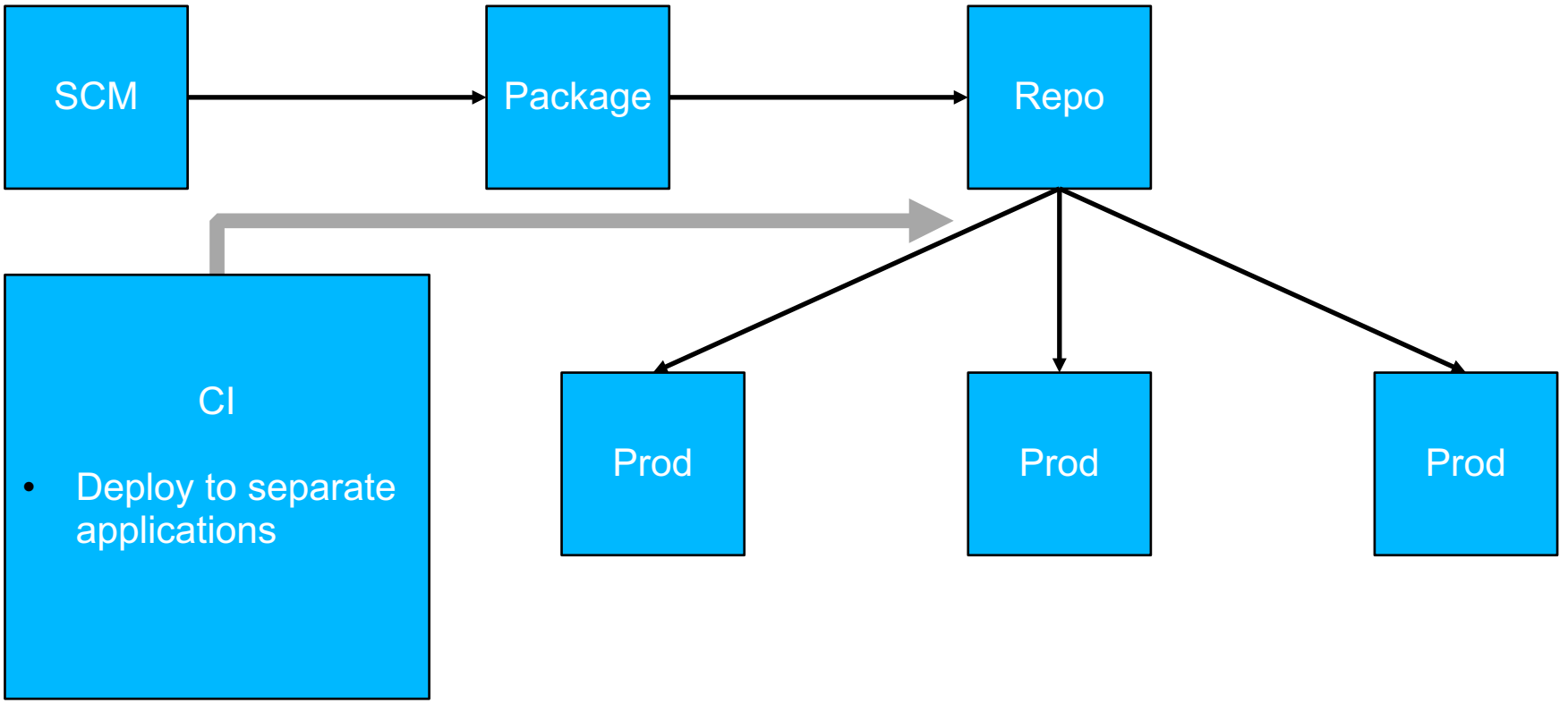
Ideal Process

From code to production



Ideal Process

From code to production



laiR

The R package repository

1. Multiple repositories
    - Shared repositories without authorization
    - Private repositories
    - CRAN & Bioconductor
2. Upload any tar.gz with a DESCRIPTION
3. Search packages
4. Dependency management across all repositories
5. Zero-config installation
6. Maintain multiple versions of the same package

laiR

The “Joel Test” for Data Science

1. Can new hires get set up in the environment to run analyses on their first day?
2. Can data scientists utilize the latest tools/packages without help from IT? ✅
3. Can data scientists use on-demand and scalable compute resources without help from IT/dev ops?
4. Can data scientists find and reproduce past experiments and results, using the original code, data, parameters, and software versions?

laiR

The “Joel Test” for Data Science

5. Does collaboration happen through a system other than email? ✅
6. Can predictive models be deployed to production without custom engineering or infrastructure work?
7. Is there a single place to search for past research and reusable data sets, code, etc.? ✅
8. Do your data scientists use the best tools money can buy?

Source: https://blog.dominodatalab.com/joel-test-data-science/

+1 Can you access your statistics without using R?

roveR

R container management

1. Preconfigured, separated R environments
2. Open-source R package
3. Linking R environments to laiR installations
4. Release packages into laiR
5. Install dependencies with specific versions using laiR API
6. Install rover using the following command:

```
install.packages("rover",
                 repos = c("https://lair.functionalfinances.com/repos/cran/",
                           "https://lair.functionalfinances.com/repos/shared/"),
                 method = "curl")
```

roveR + laiR

The “Joel Test” for Data Science

1. Can new hires get set up in the environment to run analyses on their first day? ✅
2. Can data scientists utilize the latest tools/packages without help from IT? ✅
3. Can data scientists use on-demand and scalable compute resources without help from IT/dev ops?
4. Can data scientists find and reproduce past experiments and results, using the original code, data, parameters, and software versions? ✅

roveR + laiR

The “Joel Test” for Data Science

5. Does collaboration happen through a system other than email? ✅
6. Can predictive models be deployed to production without custom engineering or infrastructure work? ✅
7. Is there a single place to search for past research and reusable data sets, code, etc.? ✅
8. Do your data scientists use the best tools money can buy?

Source: https://blog.dominodatalab.com/joel-test-data-science/

+1 Can you access your statistics without using R?

exposeR

REST API for R containers

1. Access selected R calculations via REST
2. roveR container management - CRUD
3. Select functions to expose from packages in container
4. Allows other application to integrate R calculation
5. R in a dedicated environment
6. Zero-config installation

exposeR + roveR + laiR

The “Joel Test” for Data Science

1. Can new hires get set up in the environment to run analyses on their first day? ✅
2. Can data scientists utilize the latest tools/packages without help from IT? ✅
3. Can data scientists use on-demand and scalable compute resources without help from IT/dev ops?
4. Can data scientists find and reproduce past experiments and results, using the original code, data, parameters, and software versions? ✅

exposeR + roveR + laiR

The “Joel Test” for Data Science

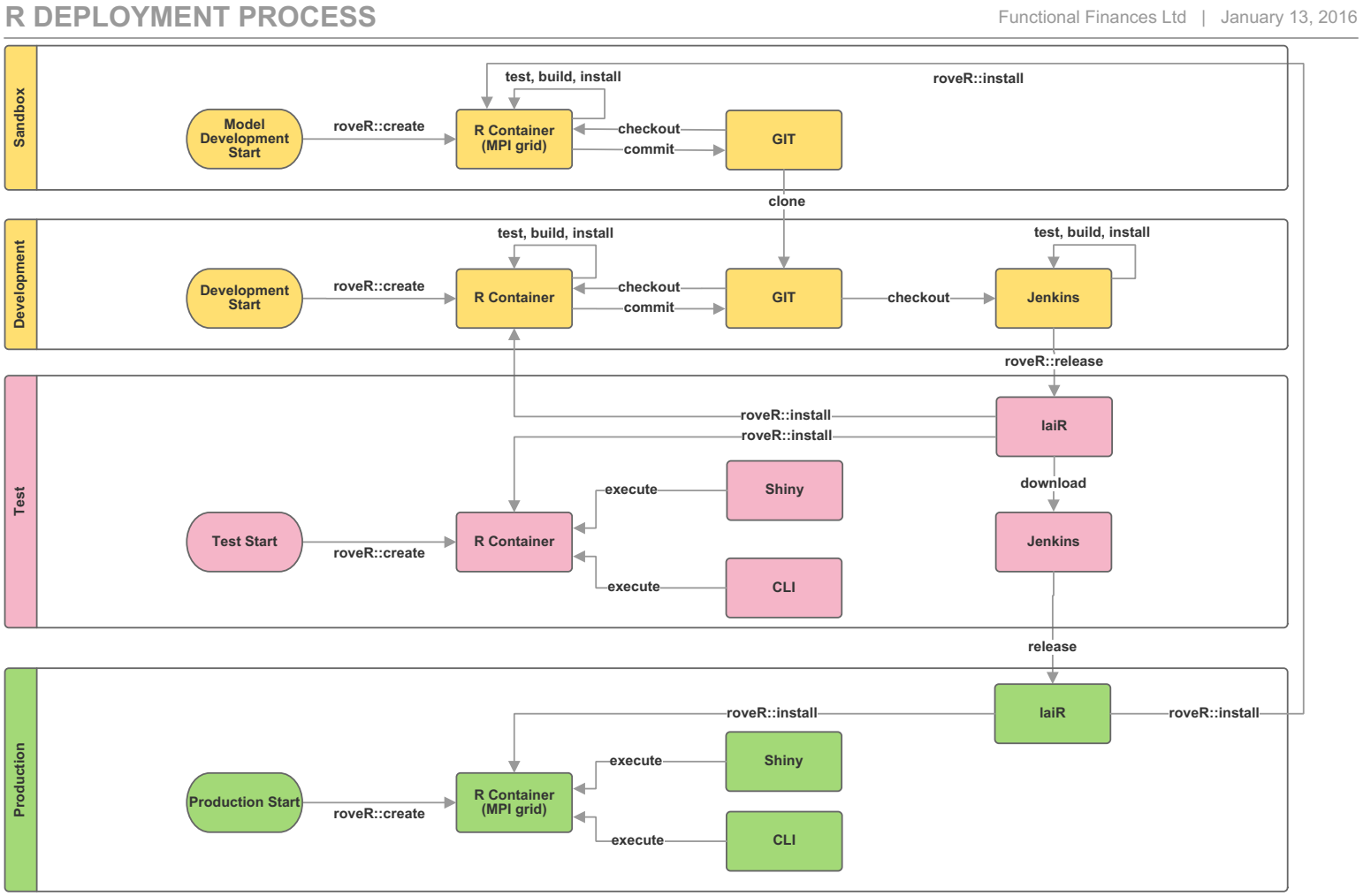
5. Does collaboration happen through a system other than email? ✅
6. Can predictive models be deployed to production without custom engineering or infrastructure work? ✅
7. Is there a single place to search for past research and reusable data sets, code, etc.? ✅
8. Do your data scientists use the best tools money can buy?

Source: https://blog.dominodatalab.com/joel-test-data-science/

+1 Can you access your statistics without using R? ✅

Reference Implementation

The Delta-Lloyd case study



Reference Implementation

The “Joel Test” for Data Science

1. Can new hires get set up in the environment to run analyses on their first day? ✅
2. Can data scientists utilize the latest tools/packages without help from IT? ✅
3. Can data scientists use on-demand and scalable compute resources without help from IT/dev ops? ✅
4. Can data scientists find and reproduce past experiments and results, using the original code, data, parameters, and software versions? ✅

Reference Implementation

The “Joel Test” for Data Science

5. Does collaboration happen through a system other than email? ✅
6. Can predictive models be deployed to production without custom engineering or infrastructure work? ✅
7. Is there a single place to search for past research and reusable data sets, code, etc.? ✅
8. Do your data scientists use the best tools money can buy? ✅

Source: https://blog.dominodatalab.com/joel-test-data-science/

+1 Can you access your statistics without using R? ✅

Questions?