

Easy and Stable Deployment of R-based Software with GitHub, devtools, and Docker

Daeyoung Kim, Jinsung Kim, Seonjeong Lee, Hyoseop Lee at Encored Technologies, Inc.

An unexpected upgrade in dependent packages can **break** your production

Our failure: the default operation of `dplyr::distinct` changed

a	b	c
1	3	4
2	6	8
1	3	4

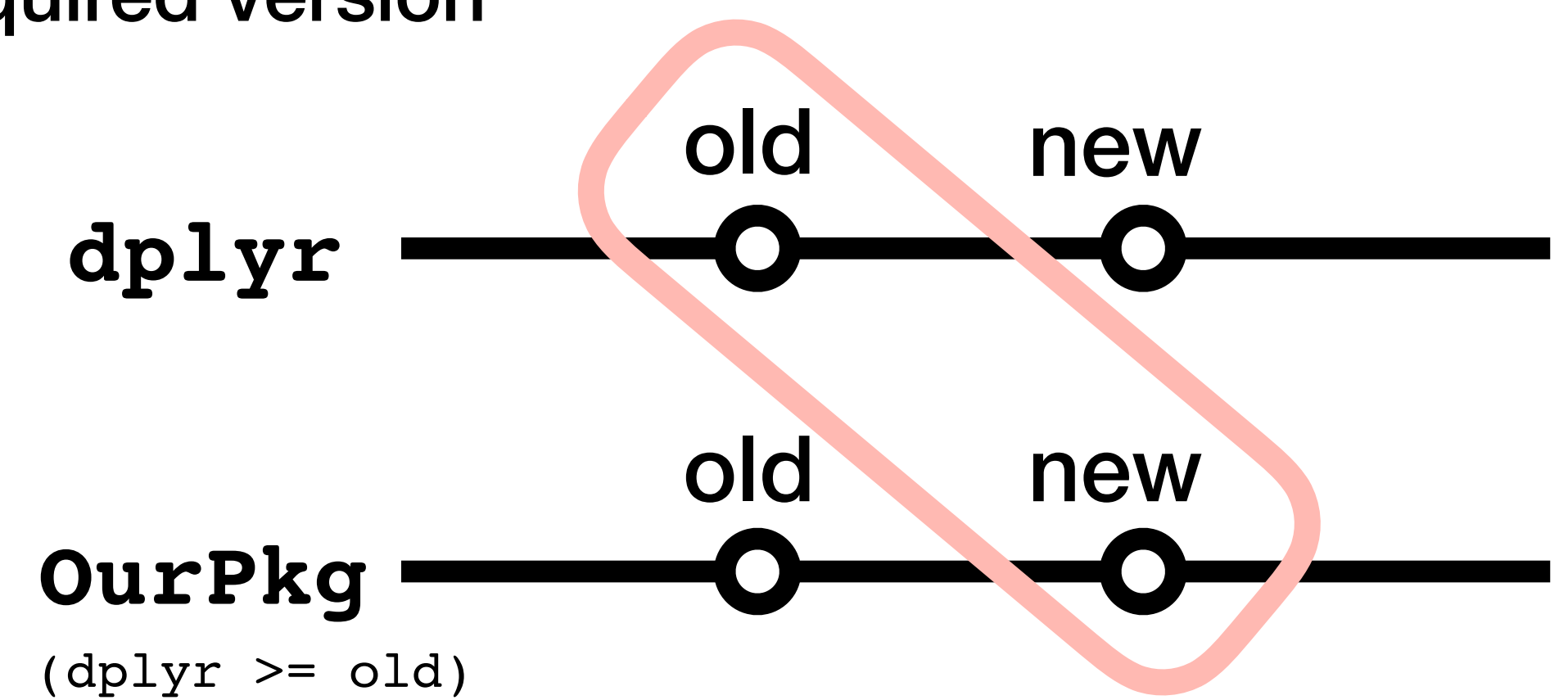
`%>% distinct(a, b)`

expected			after update	
a	b	c	a	b
1	3	4	1	3
2	6	8	2	6

new
`distinct(.data, ..., .dots, .keep_all = FALSE)`

Requirements on **new** `devtools::install_github`

- Do **NOT** upgrade dependent packages unless it violates version requirements
- Dependent packages are installed as a minimum required version



Reform `devtools::install_github` to be *conservative* on installing dependencies

Benchmark `pip install`



- Installs the *oldest* package *if possible*

```
$ pip install SomePackage # latest version
$ pip install SomePackage==1.0.4 # specific version
$ pip install 'SomePackage>=1.0.4' # minimum version
```

https://pip.pypa.io/en/stable/user_guide/#installing-packages

- Planning** “Only if needed” recursive upgrade

Currently, it performs an “eager” recursive upgrade, i.e. it upgrades all dependencies regardless of whether they still satisfy the requirements.

https://pip.pypa.io/en/stable/user_guide/#only-if-needed-recursive-upgrade

Reformed `devtools::install_github` works as following:

At 2016-06-30

DESCRIPTION of `HanSolo` v0.1.0

```
Package: HanSolo
Version: 0.1.0
Imports:
  dplyr (>= 0.4.3),
  MillenniumFalcon (>= 0.2.3)
```

Latest version of dependent packages

- `dplyr` : 0.5.0
- `MillenniumFalcon` : 0.2.3

At 2017-06-20

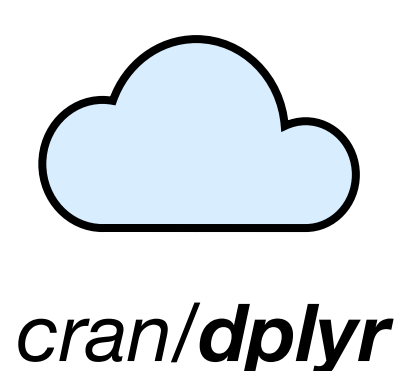
DESCRIPTION of `HanSolo` v0.2.0

```
Package: HanSolo
Version: 0.2.0
Imports:
  dplyr (>= 0.4.3),
  MillenniumFalcon (>= 0.3.0)
```

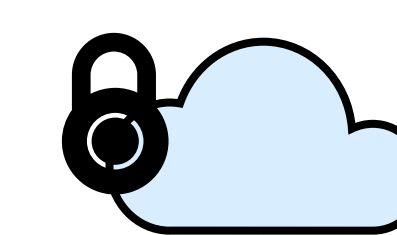
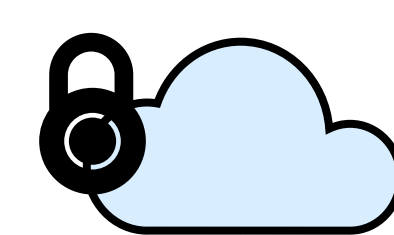
Latest version of dependent packages

- `dplyr` : 0.7.0
- `MillenniumFalcon` : 0.3.0

GitHub **public** repo



GitHub **private** repo



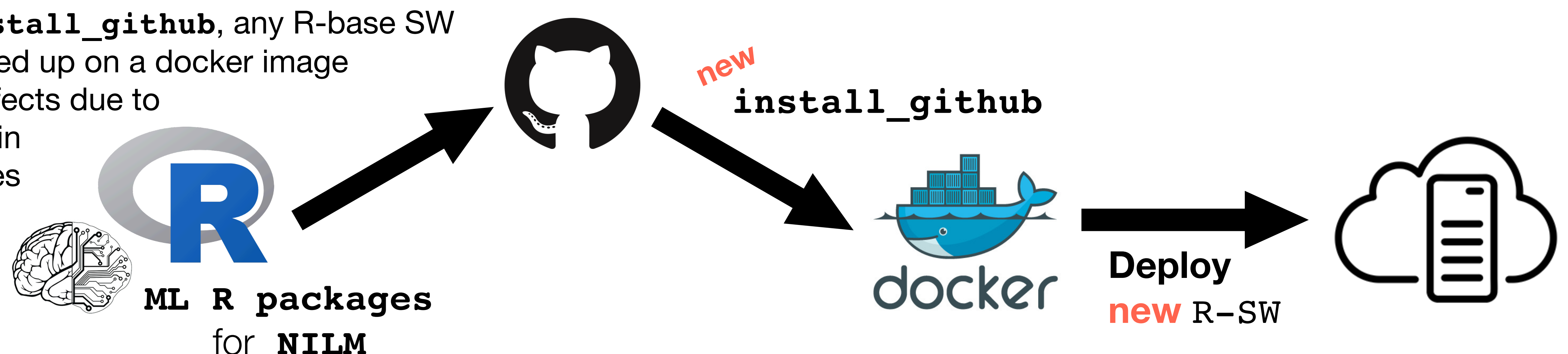
- Case 1-1 : initial installation on totally empty environment at 2016-06-30**
 - Any packages related to `HanSolo` are **not** installed yet
 - Running `install_github("Encored/HanSolo")` would install `dplyr(0.4.3)`, `MillenniumFalcon(0.2.3)` and `HanSolo(0.1.0)`
 - Latest version `dplyr(0.5.0)` would not be installed since it's not the minimum required version

- Case 2 : initial installation on preoccupied environment at 2016-06-30**
 - `dplyr(0.5.0)` is currently installed
 - Running `install_github("Encored/HanSolo")` would install `MillenniumFalcon(0.2.3)` only
 - `dplyr` would not be installed since its requirement already being satisfied
 - It would not awkwardly install the minimum required version when requirements are satisfied with margin

- Case 1-2 : Re-installation on initially deployed environment at 2017-06-20**
 - The only unsatisfied requirement is `MillenniumFalcon`
 - Running `install_github("Encored/HanSolo")` would install `MillenniumFalcon(0.3.0)` and `HanSolo(0.2.0)`
 - `dplyr` would not even being considered to install since its requirement already being satisfied

Iteratively build Docker images using *reformed* `devtools::install_github`

Using *reformed* `install_github`, any R-base SW can be safely stacked up on a docker image without any side-effects due to unwilling upgrades in dependent packages



ACKNOWLEDGMENTS

This work was supported by the Korea Institute of Energy Technology Evaluation and Planning (KETEP) and the Ministry of Trade, Industry & Energy (MOTIE) of the Republic of Korea (No. 20161210200410).