



Docker & Containers

Slides: <https://bit.ly/bioc-docker-2020>

Nitesh Turaga



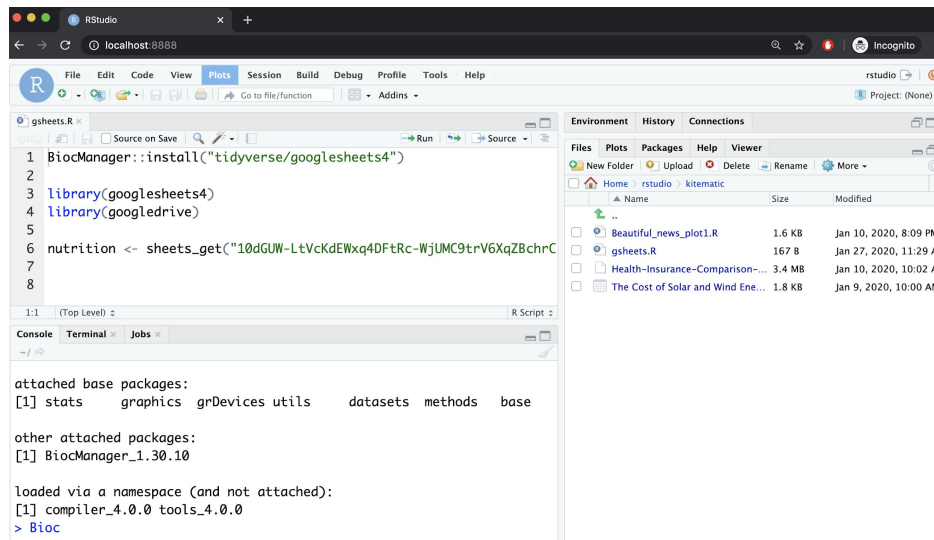
@niteshturaga

Demo 1

Introduction

Key Takeaways

- On demand **RStudio** instance
- **Different versions** of R and Bioconductor



Run R-devel, Bioconductor - devel 3.11

```
docker run \  
  -p 8888:8787 \  
  -e PASSWORD=bioc \  
    bioconductor/bioconductor_docker:devel
```

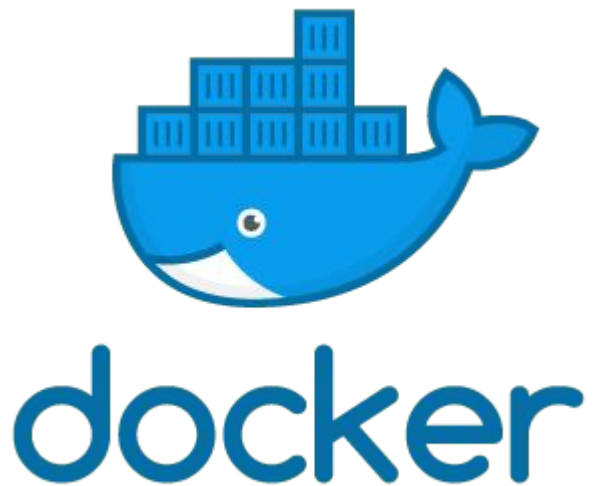
Run R-3.6.2 and Bioconductor 3.10

```
docker run \  
  -p 8787:8787 \  
  -e PASSWORD=bioc \  
    bioconductor/bioconductor_docker:latest
```

RStudio login details

Username: **rstudio**

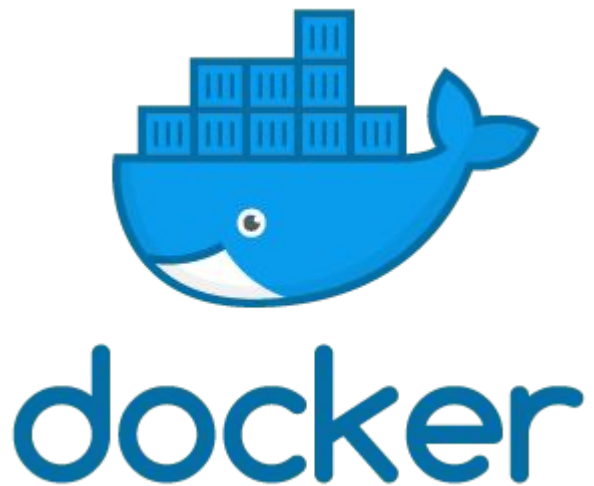
Password: **bioc**



WHAT?

WHY?

HOW?

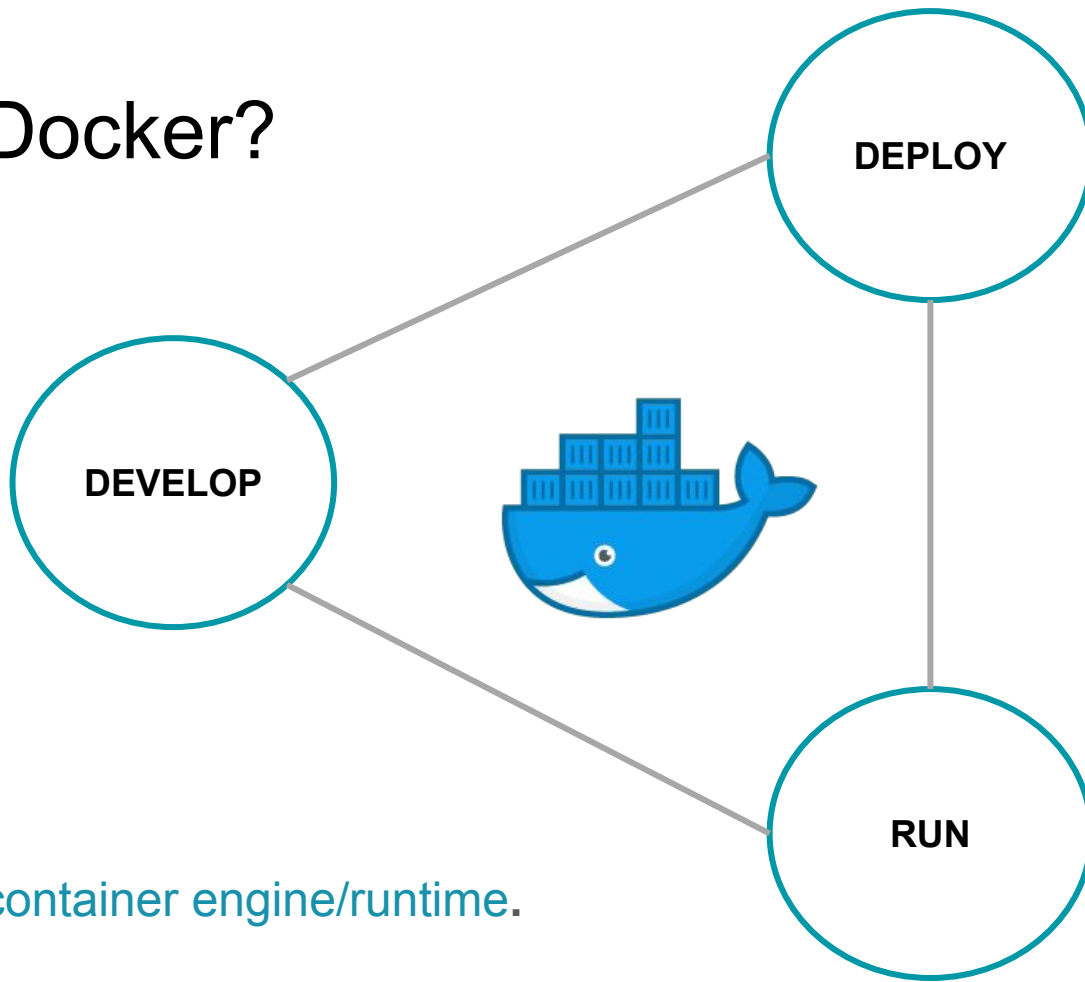


WHAT?

WHY?

HOW?

What is Docker?

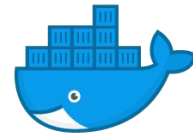
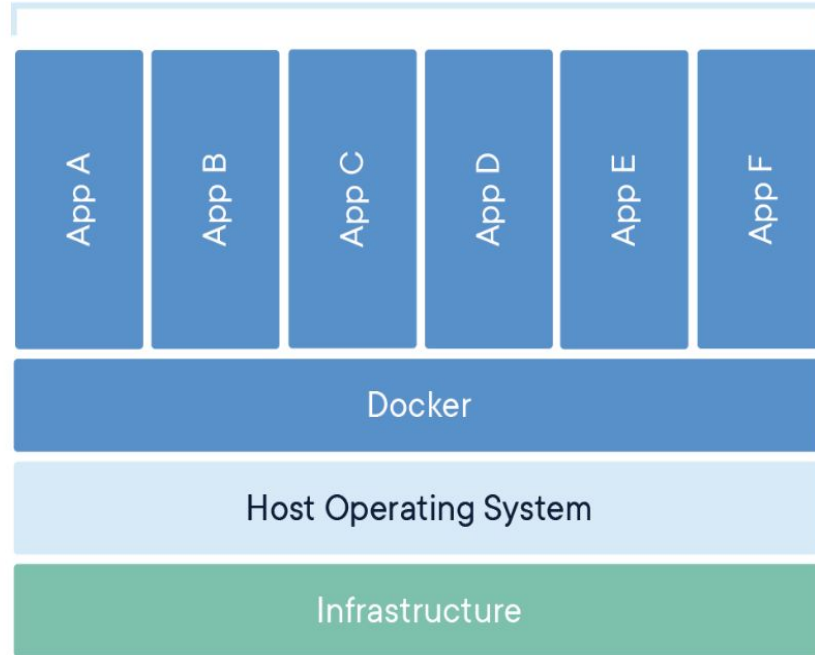


Most popular [container engine/runtime](#).

What are containers?

- Completely isolated environments
- Do not bundle a full operating system
 - Lightweight
 - Efficient
- Cross-platform

Containerized Applications



macOS



Bioconductor

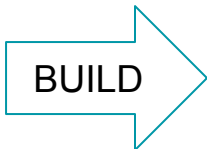
OPEN SOURCE SOFTWARE FOR BIOINFORMATICS



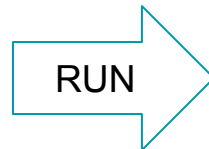
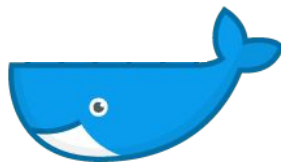
docker

Dockerfile

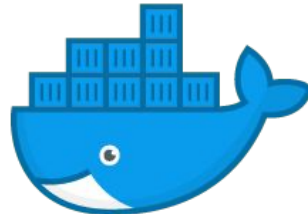
```
1 # The suggested name for this image is: bioconductor/bioconductor_docker:devel
2 FROM rocker/xtls:devel
3
4 # Version is 3.11.0 because this is the first iteration of the bioconductor Dockerfile
5 LABEL name="bioconductor/bioconductor_docker" \
6       version="3.11.0" \
7       url="https://github.com/bioconductor/bioconductor_docker" \
8       vendor="Bioconductor Project" \
9       maintainer="maintainer@bioconductor.org" \
10      description="Bioconductor docker image with system dependencies to install most packages." \
11      license="Artistic-2.0"
12
13 # make cache dirs before installing php; tip from Dirk E. Fries broken img
14 RUN rm -f /usr/lib/php/available && rm -rf /var/cache/apt/*
15
16 # layout with /usr/lib/php/available not found
17 # this will recreate
18 RUN dpkg --clear-avail
19
20 # This is to avoid the error
21 # 'debconf: unable to initialize frontend: Dialog'
22 ENV DEBIAN_FRONTEND noninteractive
23
24 # Update apt-get
25 RUN apt-get update \
26     && apt-get install -y --no-install-recommends apt-utils \
27     && apt-get install -y --no-install-recommends \
28     glib \
29     libxml2-dev \
30     python3-pip \
31     libz-dev \
32     liblzma-dev \
33     libbz2-dev \
34     libjpeg-dev \
35     libmer1-dev-compat \
36     && rm -rf /var/lib/apt/lists/* \
37     && apt-get install -y \
38     php-conf \
39     fortran77-compiler \
40     byacc \
41     autoconf \
42     curl \
43     && rm -rf /var/lib/apt/lists/*
```



docker
image



docker
container



What is the Bioconductor image called?

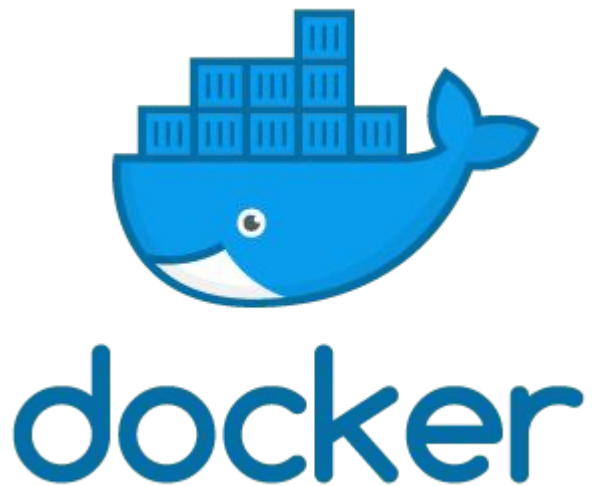
bioconductor / bioconductor_docker : latest

Username
or
Organization

Image Name

Tag

REPOSITORY



WHAT?

WHY?

HOW?

Why should you use Bioconductor with Docker?

- Avoid installing system dependencies
- Reproducible and consistent environment.
- Cross-platform, run anywhere
- Portability / Sharing
- Isolation



Reproducible environment - Tags

bioconductor/bioconductor_docker: **devel**

bioconductor/bioconductor_docker: **RELEASE_3_10**

latest == RELEASE_3_10

Reproducible environment - Dockerfile

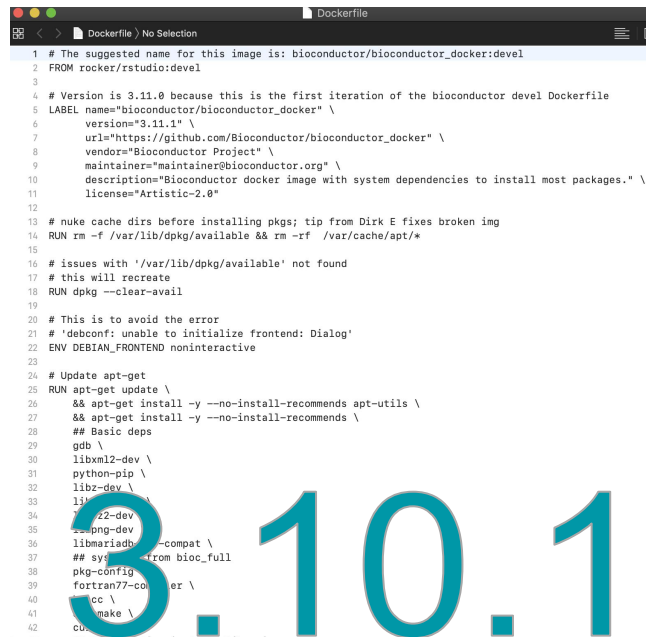
devel



```
1 # The suggested name for this image is: bioconductor/bioconductor_docker:devel
2 FROM rocker/rstudio:devel
3
4 # Version is 3.11.0 because this is the first iteration of the bioconductor devel Dockerfile
5 LABEL name="bioconductor/bioconductor_docker" \
6     version="3.11.1" \
7     url="https://github.com/Bioconductor/bioconductor_docker" \
8     vendor="Bioconductor Project" \
9     maintainer="maintainer@bioconductor.org" \
10    description="Bioconductor docker image with system dependencies to install most packages." \
11    license="Artistic-2.0"
12
13 # nuke cache dirs before installing pkgs; tip from Dirk E fixes broken img
14 RUN rm -f /var/lib/dpkg/available && rm -rf /var/cache/apt/*
15
16 # issues with '/var/lib/dpkg/available' not found
17 # this will recreate
18 RUN dpkg --clear-avail
19
20 # This is to avoid the error
21 # 'debconf: unable to initialize frontend: Dialog'
22 ENV DEBIAN_FRONTEND noninteractive
23
24 # Update apt-get
25 RUN apt-get update \
26     && apt-get install -y --no-install-recommends apt-utils \
27     && apt-get install -y --no-install-recommends \
28     ## Basic deps
29     gdb \
30     libxml2-dev \
31     python-pip \
32     libz-dev \
33     libbz2-dev \
34     libpng-dev \
35     libbz2-dev \
36     libbz2-dev \
37     ## sys deps from bioc_full
38     pkg-config \
39     fortan77-compiler \
40     gcc \
41     autotools \
42     curl \
```

3.11.1

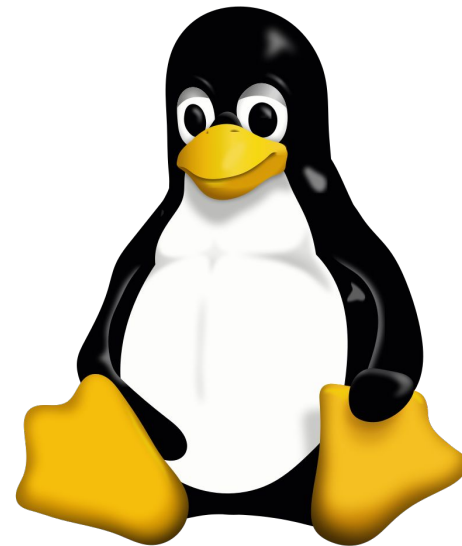
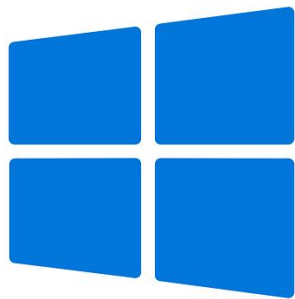
RELEASE_3_10



```
1 # The suggested name for this image is: bioconductor/bioconductor_docker:devel
2 FROM rocker/rstudio:devel
3
4 # Version is 3.11.0 because this is the first iteration of the bioconductor devel Dockerfile
5 LABEL name="bioconductor/bioconductor_docker" \
6     version="3.11.1" \
7     url="https://github.com/Bioconductor/bioconductor_docker" \
8     vendor="Bioconductor Project" \
9     maintainer="maintainer@bioconductor.org" \
10    description="Bioconductor docker image with system dependencies to install most packages." \
11    license="Artistic-2.0"
12
13 # nuke cache dirs before installing pkgs; tip from Dirk E fixes broken img
14 RUN rm -f /var/lib/dpkg/available && rm -rf /var/cache/apt/*
15
16 # issues with '/var/lib/dpkg/available' not found
17 # this will recreate
18 RUN dpkg --clear-avail
19
20 # This is to avoid the error
21 # 'debconf: unable to initialize frontend: Dialog'
22 ENV DEBIAN_FRONTEND noninteractive
23
24 # Update apt-get
25 RUN apt-get update \
26     && apt-get install -y --no-install-recommends apt-utils \
27     && apt-get install -y --no-install-recommends \
28     ## Basic deps
29     gdb \
30     libxml2-dev \
31     python-pip \
32     libz-dev \
33     libbz2-dev \
34     libpng-dev \
35     libbz2-dev \
36     libbz2-dev \
37     ## sys deps from bioc_full
38     pkg-config \
39     fortan77-compiler \
40     gcc \
41     autotools \
42     curl \
```

3.10.1

Cross platform



Bioconductor images are built on Debian

Share your work for free

GitHub



Amazon ECR



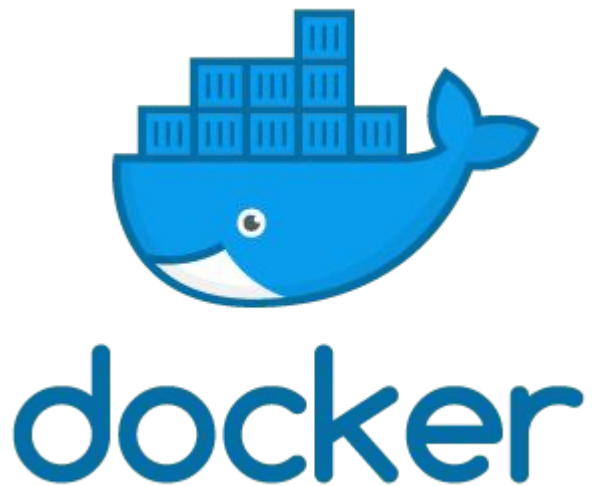
Google Container Registry

Features of the Bioconductor docker image

- Image has the **system dependencies to all*** Bioconductor packages
- Test packages as **devel** image **resembles the linux nightly builders.**

* devel image: 23 packages don't install

* latest (release) image: 10 packages don't install



WHAT?

WHY?

HOW?

How to use Bioconductor on Docker?

A. Pull an image

B. Use image

C. Extend image



Most people will fall in category A and B.

Demo 2

Attaching volumes

Docker demo

```
## get a list of docker images  
$ docker images
```

```
## download/pull an image  
$ docker pull bioconductor/bioconductor_docker:latest
```

```
## run docker image
```

```
$ docker run \  
    -p 8787:8787 \  
    -e PASSWORD=bioc \  
        bioconductor/bioconductor_docker:latest
```

-p maps the port from container to host.

-e set environment variables required for the image.

<https://localhost:8787> or <https://127.0.0.1:8787> (localhost == 127.0.0.1)

```
## get a list of containers
```

```
$ docker ps
```

```
## Stop docker image
```

```
$ docker stop <CONTAINER ID>
```

Mount volumes on the container

```
## Make a new directory
```

```
$ mkdir /release
```

```
## Mount the volume /release
```

```
$ docker run \  
  -v /release:/usr/local/lib/R/host-site-library \  
  -e PASSWORD=bioc \  
  -p 8787:8787 \  
  bioconductor/bioconductor_docker:latest
```

```
## Check the library search path  
> .libPaths()
```

```
## Install a package  
> BiocManager::install("Biobase")
```

```
/usr/local/lib/R/host-site-library
```

- Packages in that directory will be directly available to use as it is configured to show up in the search path.
- Prevents your docker image from growing in size due to package installation and data.

Mount libraries and data

```
## Mount 1. libraries 2. data
```

```
$ docker run \  
  -v /release:/usr/local/lib/R/host-site-library \  
  -v /Documents:/home/rstudio/  
  -e PASSWORD=bioc \  
  -p 8787:8787 \  
  bioconductor/bioconductor_docker:latest
```


Run containers interactively

```
## Run interactively and start bash
```

```
$ docker run -it \  
    bioconductor/bioconductor_docker:latest bash
```

```
## Run interactively and start R
```

```
$ docker run -it \  
    bioconductor/bioconductor_docker:latest R
```

Connect to a running container

```
## List containers and connect to one interactively
```

```
$ docker ps
```

```
$ docker exec -it <CONTAINER ID> bash
```

Special case - Disable authentication

```
## Disable authentication
```

```
$ docker run \  
  -e DISABLE_AUTH=true \  
  -p 8787:8787 \  
    bioconductor/bioconductor_docker:latest
```

Special case - Enable root

```
## Enable root and disable authentication
```

```
$ docker run \  
  -e ROOT=true \  
  -e DISABLE_AUTH=true \  
  -p 8787:8787 \  
  bioconductor/bioconductor_docker:latest
```

Demo 3

Extending images

Step 1: Create a new directory

```
## Make a new directory
```

```
$ mkdir my_image
```

Step 2: Create a Dockerfile

```
## Inherit from bioconductor/bioconductor_docker:latest  
  
FROM bioconductor/bioconductor_docker:latest  
  
RUN pip install tensorflow  
  
RUN R -e "BiocManager::install('scAlign')"
```

Step 3: Build image and publish

```
## build the docker image
```

```
$ docker build -t nitesh1989/my_image:v1 .
```

```
## dockerhub account needed, my username - nitesh1989
```

```
$ docker push nitesh1989/my_image:v1
```


Extend images

- Extend images and publish for your software / manuscripts.
- Extend images to contribute to Bioconductor.

Bioconductor now accepts images from the community as contributions.

Summary (TL; DR)

1. Run a bioconductor docker image with latest Bioconductor + RStudio.
2. Mount volumes for compiled libraries and data
3. Extend bioconductor docker images for you own needs.



<https://localhost:8787>



Constant value: first on search path `.libPaths()`

RUN CONTAINER

docker run

MOUNT
VOLUMES

r-libs

`-v /library_X_Y/:/usr/local/lib/R/host-site-library`

data

`-v /data/:/home/rstudio/data`

ENV VARS

`-e PASSWORD=bioc`

MAP PORTS

`-p 8787:8787`

DOCKER IMAGE

`bioconductor/bioconductor_docker : RELEASE_X_Y`

IMAGE NAME

TAG

{ latest
RELEASE_3_10
devel }

Help: <https://bioconductor.org/help/docker>

Coming soon



Google Cloud



kubernetes

Questions ?

**Thanks to the Bioconductor team,
Martin Morgan, and Levi Waldron**

Links to resources

1. [Bioconductor on Dockerhub](#)
2. Github: [Bioconductor/bioconductor_docker](#)
3. [Bioconductor website docker documentation](#)
4. [Available versions \(tags\) for bioconductor_docker image](#)
5. [Best practices and contributions](#)