



Semantix

Hadoop - Básico

Aula 2

Quem sou eu?

Eu sou Rodrigo Augusto Rebouças.

Engenheiro de dados da Semantix
Instrutor do Semantix Mentoring Academy

Você pode me encontrar em:
rodrigo.augusto@semantix.com.br





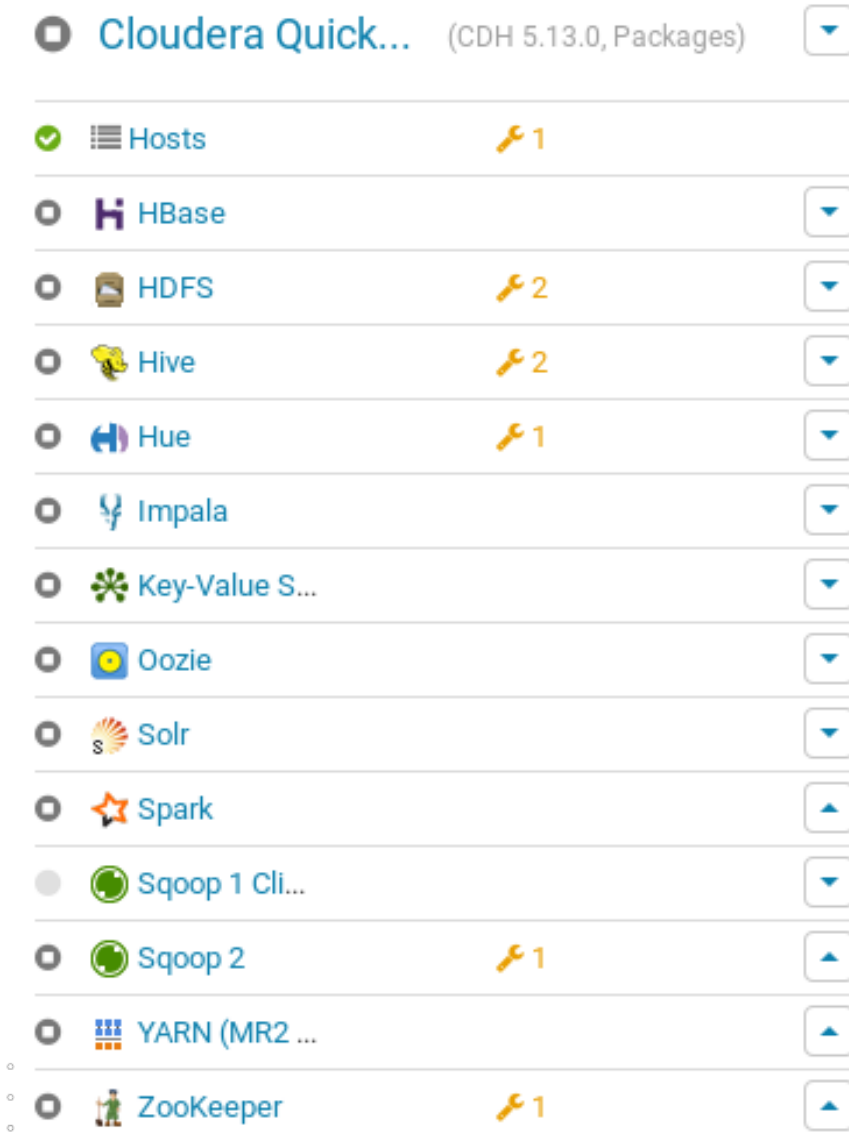
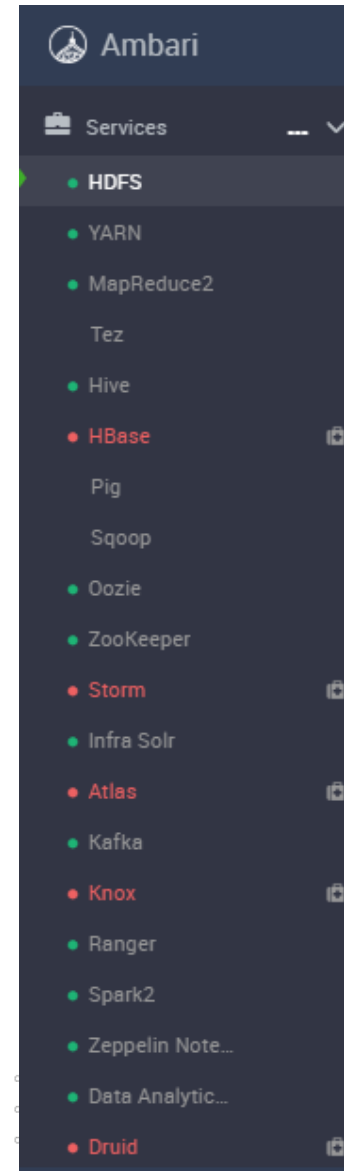
Ambientes de Desenvolvimento



Como Começar?

○ Cloudera

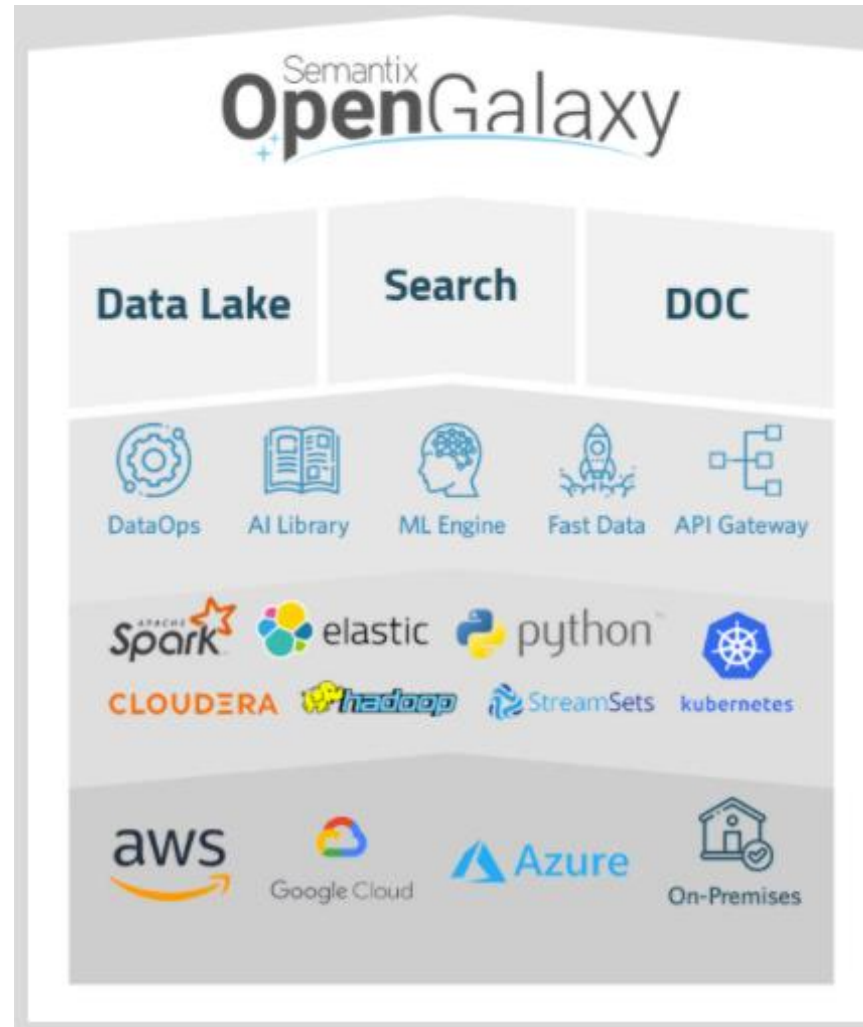
- Hortonworks Sandbox HDP 3.0
 - <https://www.cloudera.com/downloads/hortonworks-sandbox/hdp.html>
- Cloudera Quickstart VM CDH 5.13
 - <https://www.cloudera.com/downloads/>



Como Começar?

- Semantix

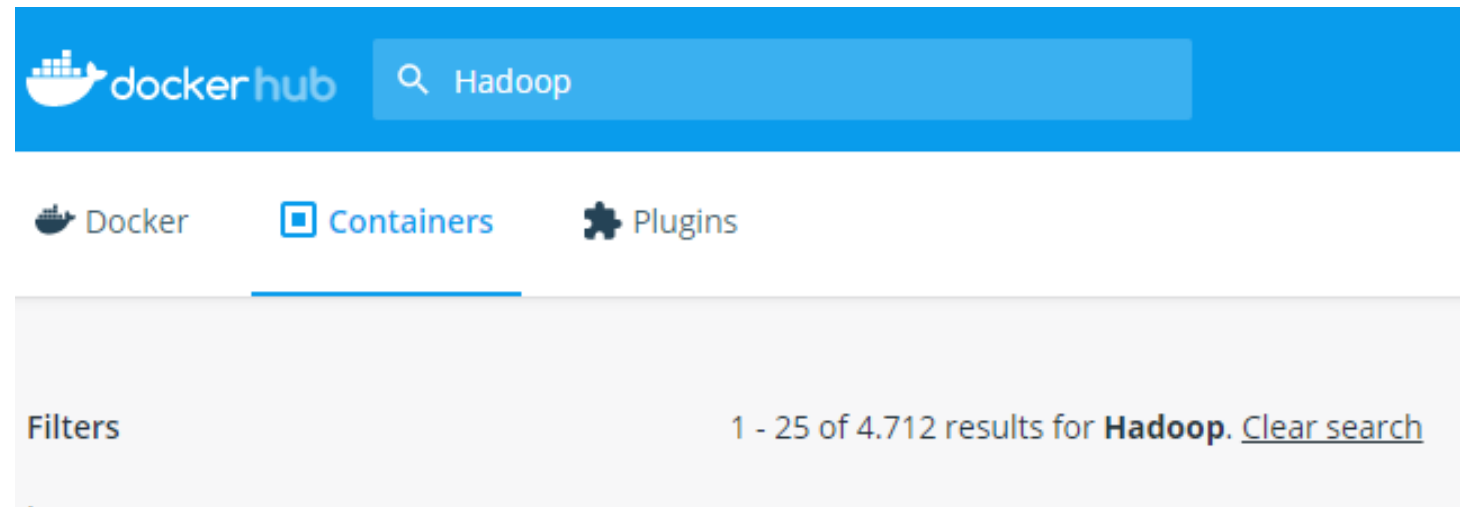
<https://semantix.com.br/opengalaxy/>



Como Começar?

- Containers

<https://hub.docker.com/>



Vamos Começar?

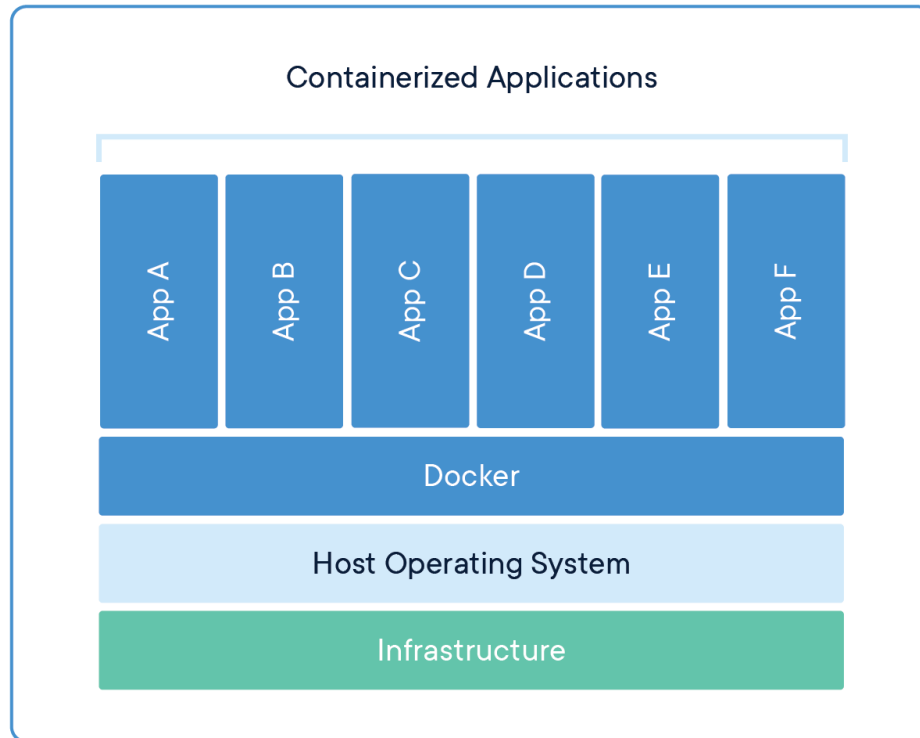
- Containers
 - Software
 - Docker e Docker Compose
 - Sistema Operacional de 64 bits
 - Windows
 - Linux
 - Mac
 - Hardware
 - Memória RAM de 8 GB
 - Acesso a internet
 - HD com no mínimo 20 GB de espaço livre



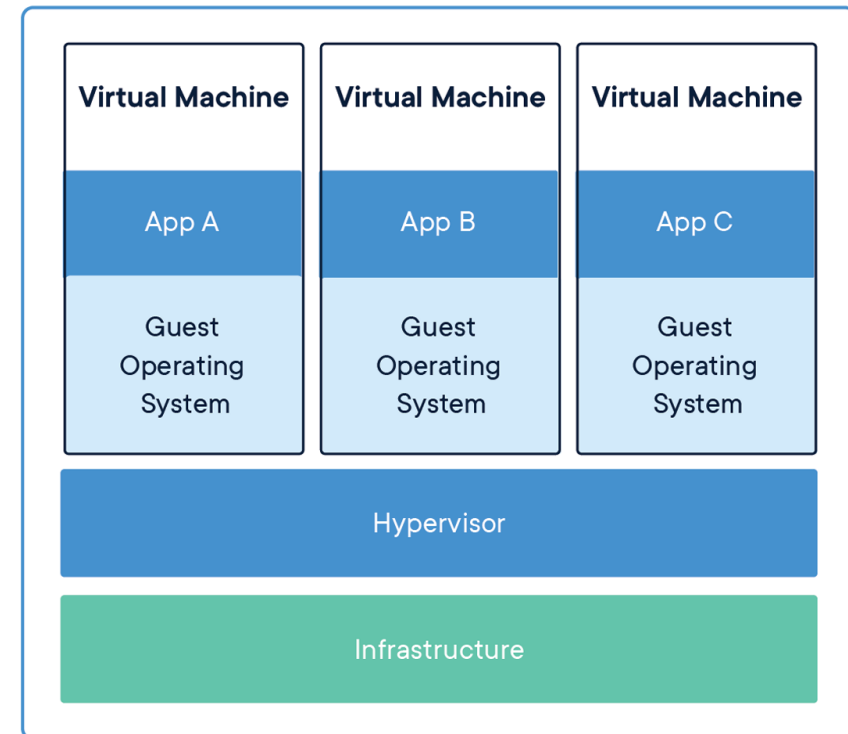
Preparar Ambiente de Desenvolvimento

Cointainers x Maquina Virtual

○ Container



○ Maquina Virtual (VM)



○ Docker

- <https://www.docker.com/>

Preparação Ambiente – Instalação Docker e Docker-compose

○ Instalação

- Docker: <https://docs.docker.com/get-docker/>
- Docker Compose: <https://docs.docker.com/compose/install/>
- SO
 - Windows
 - Docker Desktop (Hyper-V ou WSL2)
 - Docker Toolbox (VirtualBox) - Instalacao_Docker_Windos_10_Home.pdf
 - Linux
 - Docker Engine
 - Docker Compose
 - Mac
 - Docker Desktop



Instalação Docker - Windows

Preparação Ambiente – Instalação Docker Desktop - Windows


○ Instalação

- Docker: <https://docs.docker.com/get-docker/>

Get Docker


Docker is an open platform for developing, shipping, and running applications. Docker enables you to separate your applications from your infrastructure so you can deliver software quickly. With Docker, you can manage your infrastructure in the same ways you manage your applications. By taking advantage of Docker's methodologies for shipping, testing, and deploying code quickly, you can significantly reduce the delay between writing code and running it in production.

You can download and install Docker on multiple platforms. Refer to the following section and choose the best installation path for you.




Docker Desktop for Mac

A native application using the macOS sandbox security model which delivers all Docker tools to your Mac.



Docker Desktop for Windows

A native Windows application which delivers all Docker tools to your Windows computer.



Docker for Linux

Install Docker on a computer which already has a Linux distribution installed.

Preparação Ambiente – Instalação Docker Desktop - Windows

- <https://hub.docker.com/editions/community/docker-ce-desktop-windows/>



Docker Desktop for Windows

By **Docker**

The fastest and easiest way to get started with Docker on Windows

Edition

Windows

x86-64

Get Docker Desktop for Windows

Docker Desktop for Windows is available for free.

Requires Microsoft Windows 10 Professional or Enterprise 64-bit, or Windows 10 Home 64-bit with WSL 2.

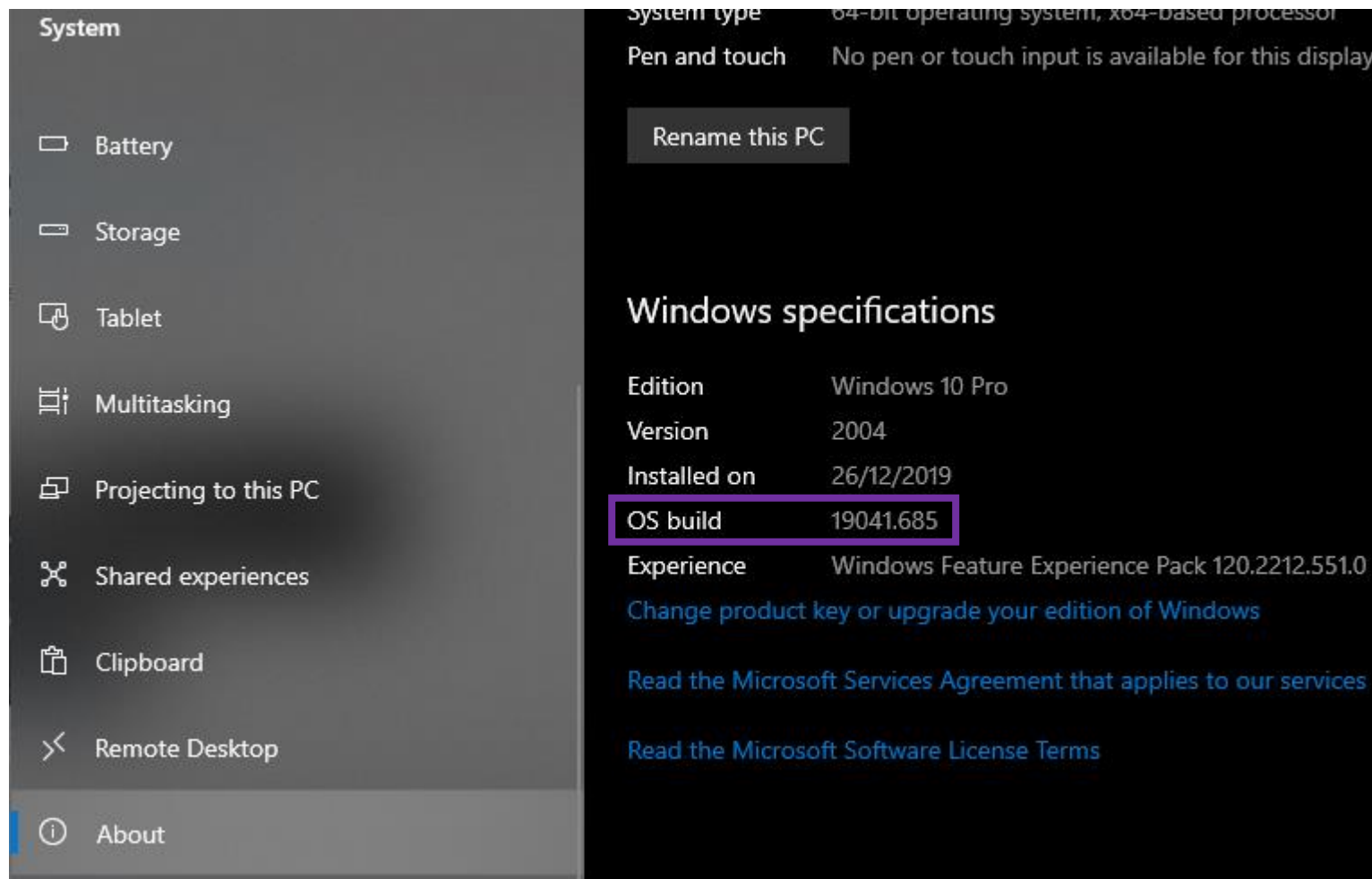
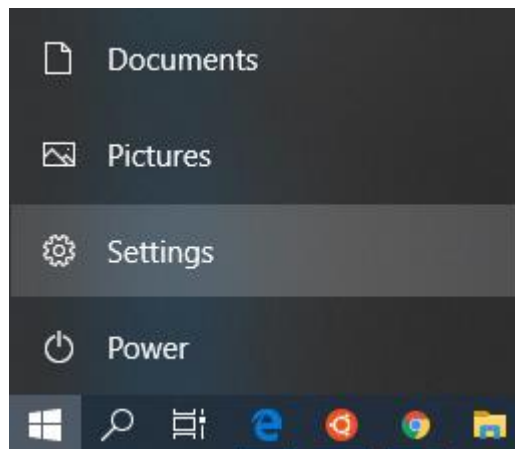
By downloading this, you agree to the terms of the [Docker Software End User License Agreement](#) and the [Docker Data Processing Agreement \(DPA\)](#).



Get Docker

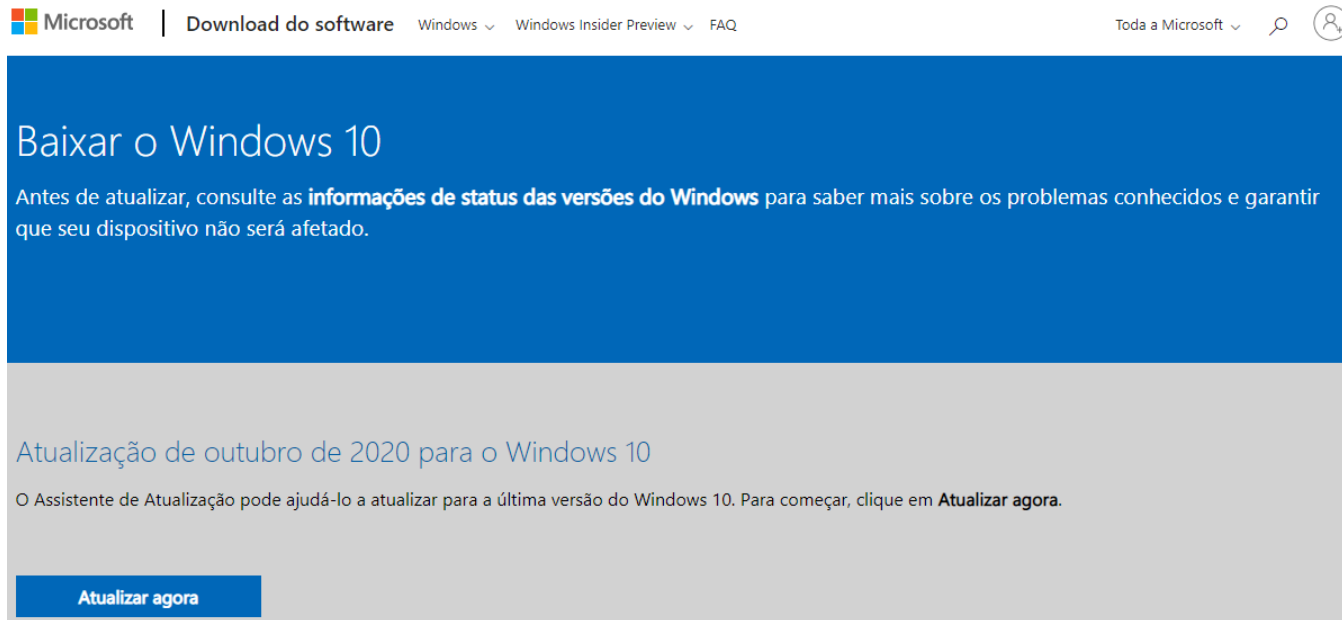
Verificar Atualização Windows

- Windows/System/Settings/About
 - OS Build >18362
 - Sem suporte a WSL 2



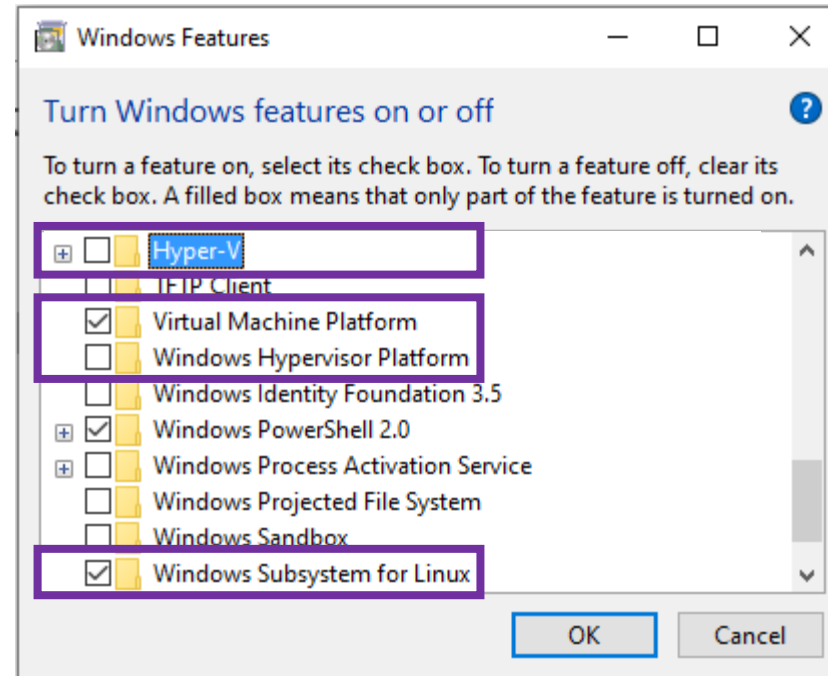
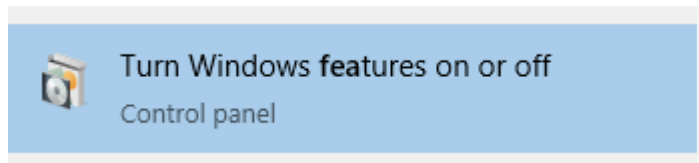
Atualizar Windows

- Tutorial Oficial Microsoft
 - <https://docs.microsoft.com/pt-br/windows/wsl/install-win10>
- Link de Atualização Windows 10 (20.04)
 - <https://www.microsoft.com/pt-br/software-download/windows10>



Habilitar Recursos Windows

- Habilitar o Subsistema do Windows para Linux (WSL) e a Virtual Machine Platform
 - Power Shell (Modo Administrador)
 - `dism.exe /online /enable-feature /featurename:Microsoft-Windows-Subsystem-Linux /all /norestart`
 - `dism.exe /online /enable-feature /featurename:VirtualMachinePlatform /all /norestart`
 - Pesquisar no windows por **recursos**



Atualizar WSL 2

- Download WSL 2
 - https://wslstorestorage.blob.core.windows.net/wslblob/wsl_update_x64.msi
- Execute o comando no PowerShell (Administrador)
 - Definir o WSL 2 como padrão ao instalar uma nova distribuição do Linux
 - `wsl --set-default-version 2`
 - Definir o WSL 2 para uma distribuição já existente
 - `wsl --set-version <distribution name> 2`

```
Administrator: Windows PowerShell
Windows PowerShell
Copyright (C) Microsoft Corporation. All rights reserved.

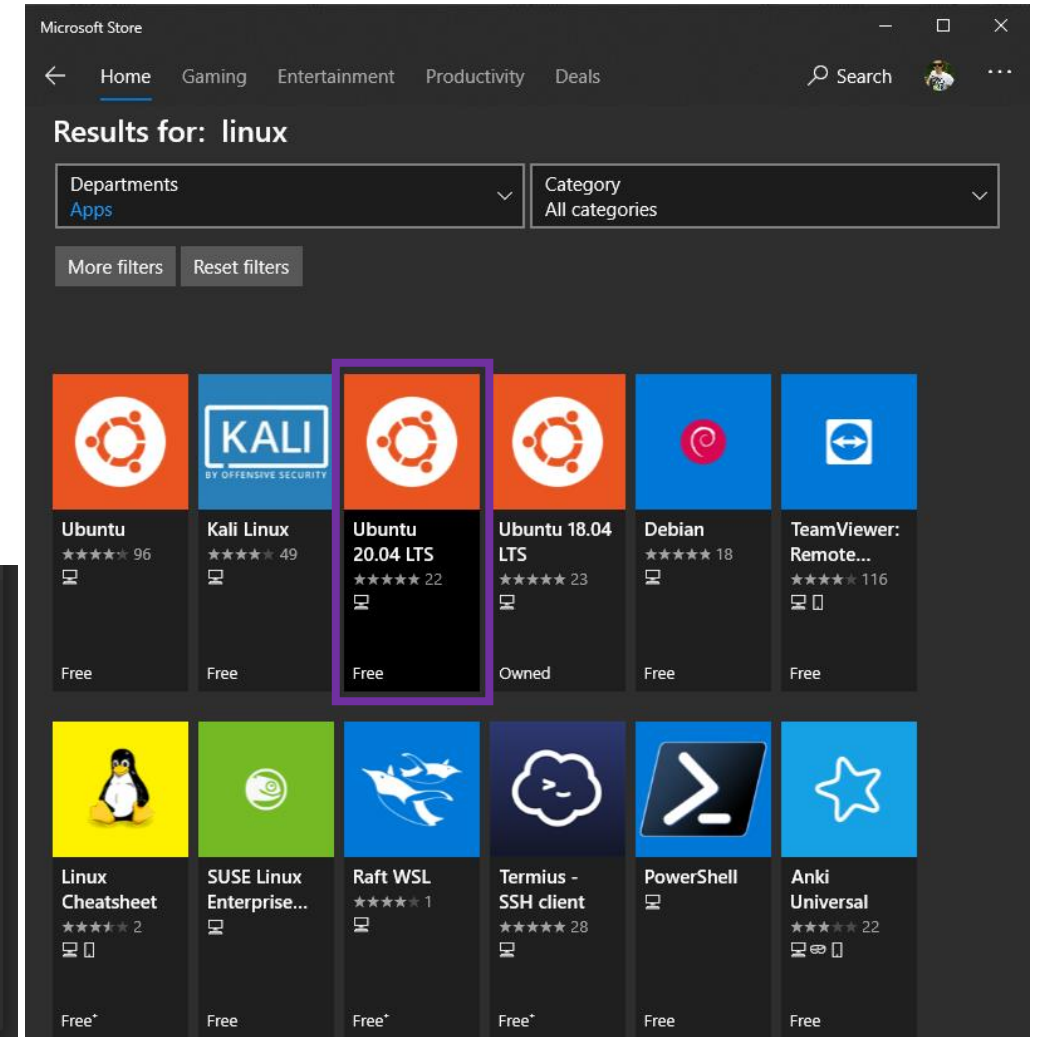
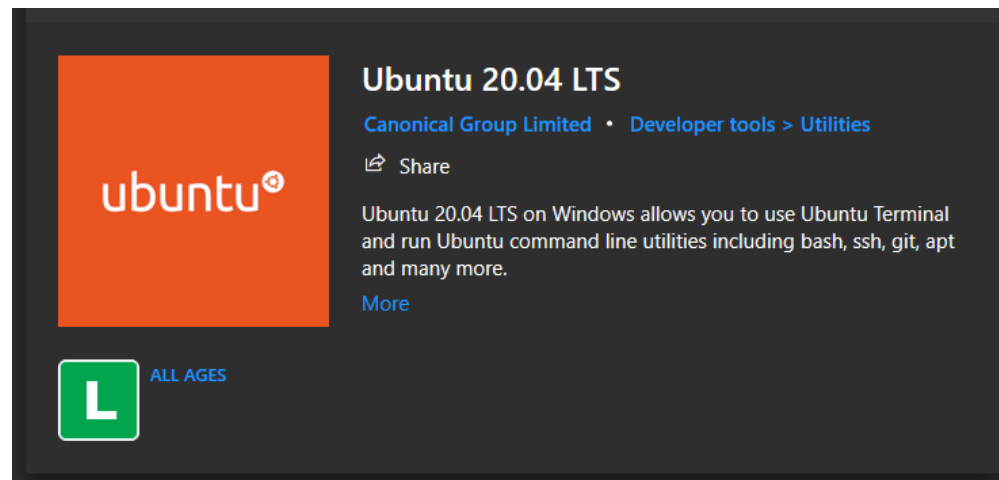
Try the new cross-platform PowerShell https://aka.ms/pscore6

PS C:\WINDOWS\system32> wsl --set-default-version 2
For information on key differences with WSL 2 please visit https://aka.ms/wsl2
PS C:\WINDOWS\system32> █
```

Instalar Ubuntu

- Instalar Distribuição Linux
 - Microsoft Store/Linux
 - Ubuntu 20.04 LTS (500 mb)
- Verificar versão distribuição

```
Administrator: Windows PowerShell
PS C:\WINDOWS\system32> wsl -l -v
NAME                STATE      VERSION
* Ubuntu-20.04      Running    2
PS C:\WINDOWS\system32>
```



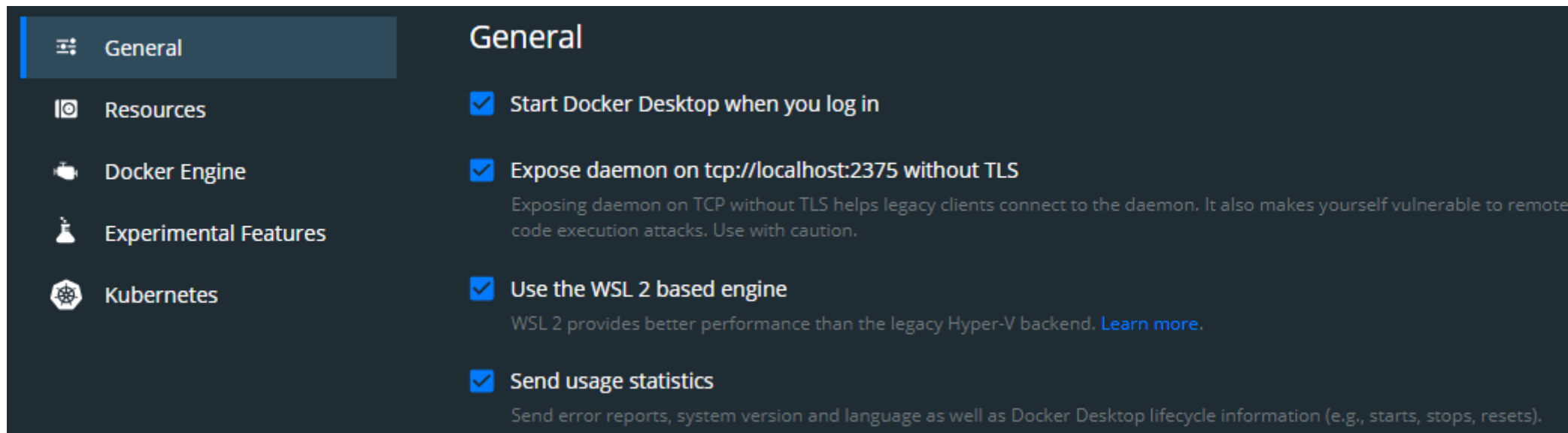
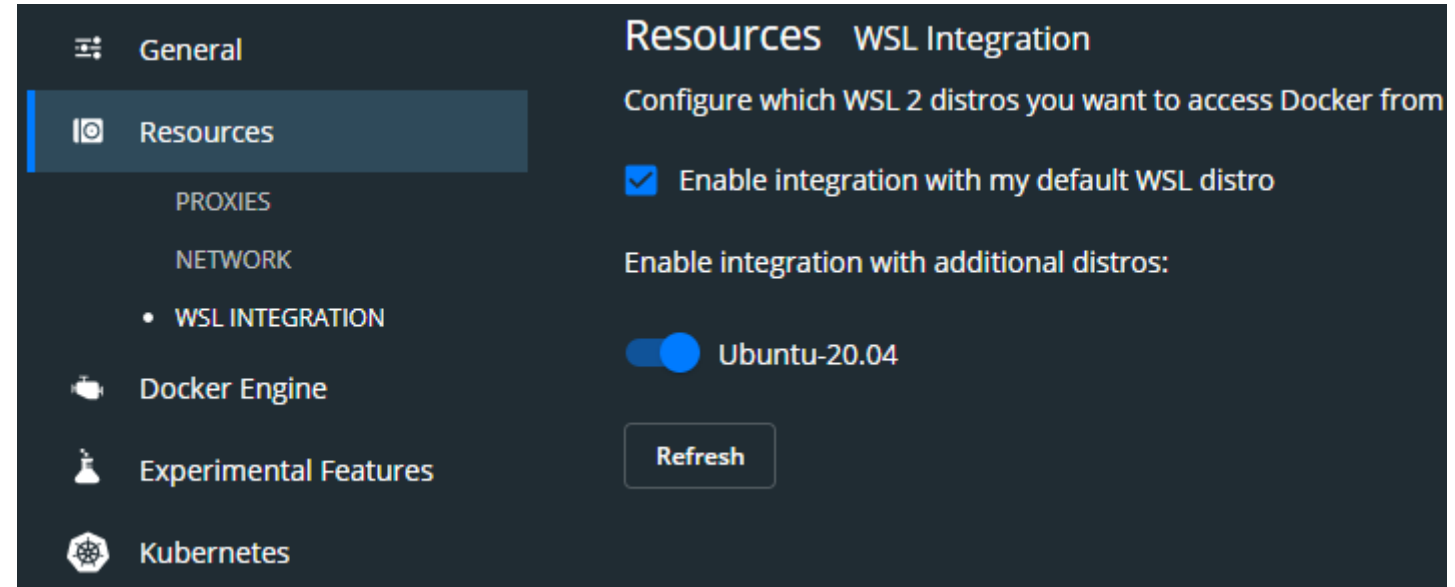
Intalar Docker Desktop

○ Instalação Docker Desktop

🔄 Installing Docker Desktop 3.0.0 (50684)

Configuration

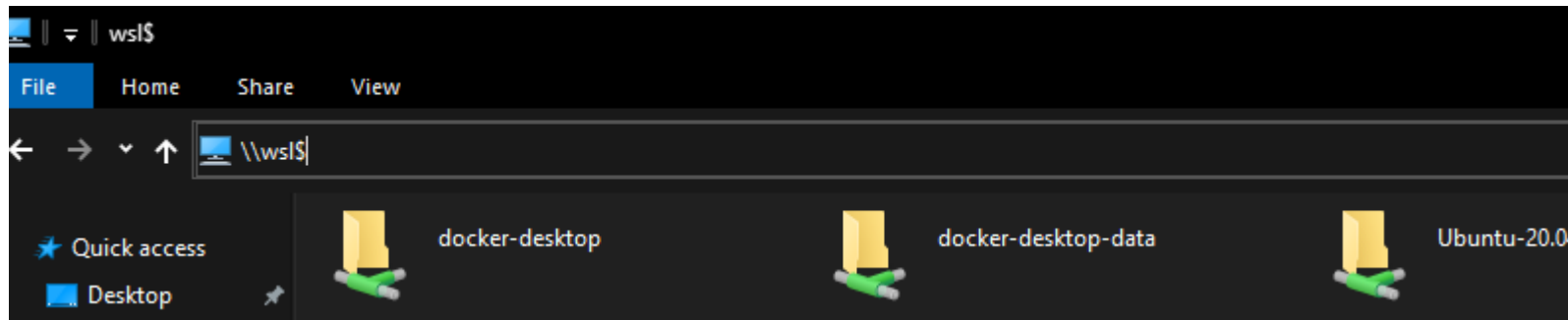
- ☒ Install required Windows components for WSL 2
- ☒ Add shortcut to desktop



WSL2 Visualizar kernel

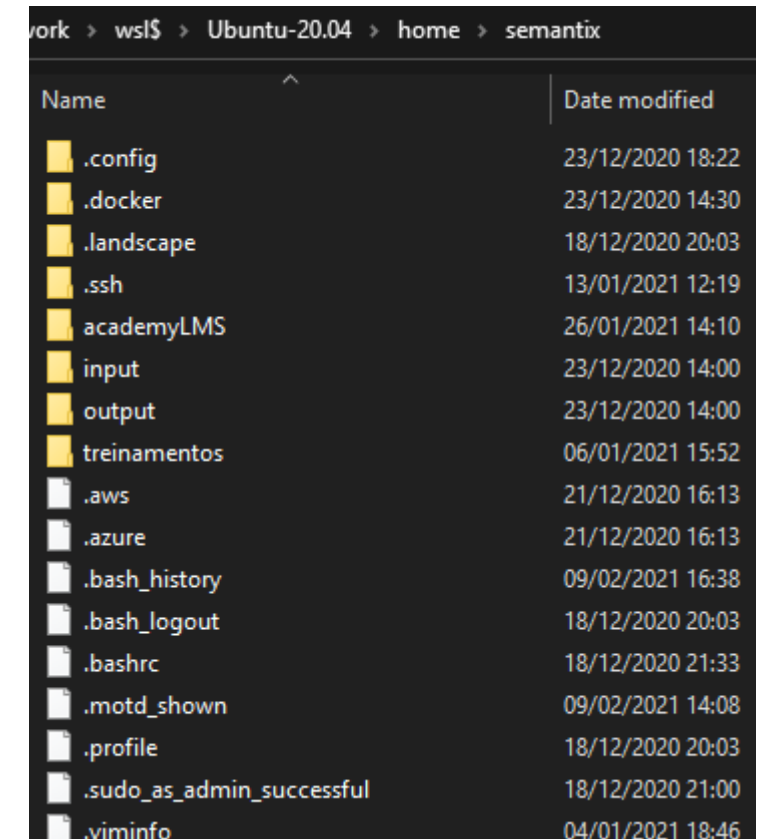
- Visualizar o disco do WSL

- `\\wsl$`



- Abrir o File Explorer(windows E) no terminal

```
semantix@NTBSTX7158: ~  
semantix@NTBSTX7158:~  
$ explorer.exe  
semantix@NTBSTX7158:~  
$ explorer.exe .
```





Instalação Docker - Linux

Preparação Ambiente – Instalação Docker Desktop - Mac


○ Instalação

- Docker: <https://docs.docker.com/get-docker/>

Get Docker


Docker is an open platform for developing, shipping, and running applications. Docker enables you to separate your applications from your infrastructure so you can deliver software quickly. With Docker, you can manage your infrastructure in the same ways you manage your applications. By taking advantage of Docker's methodologies for shipping, testing, and deploying code quickly, you can significantly reduce the delay between writing code and running it in production.

You can download and install Docker on multiple platforms. Refer to the following section and choose the best installation path for you.




Docker Desktop for Mac

A native application using the macOS sandbox security model which delivers all Docker tools to your Mac.



Docker Desktop for Windows

A native Windows application which delivers all Docker tools to your Windows computer.



Docker for Linux

Install Docker on a computer which already has a Linux distribution installed.

Preparação Ambiente – Instalação Docker Engine - Linux

- <https://docs.docker.com/engine/install/>

Server			
Docker provides <code>.deb</code> and <code>.rpm</code> packages from the following Linux distributions and architectures:			
Platform	x86_64 / amd64	ARM	ARM64 / AARCH64
CentOS	✓		✓
Debian	✓	✓	✓
Fedora	✓		✓
Raspbian		✓	✓
Ubuntu	✓	✓	✓

Preparação Ambiente – Instalação Docker Engine - Linux

○ Docker Engine no Ubuntu

- Atualizar e instalar alguns pacotes

```
$ sudo apt-get update
```

```
$ sudo apt-get install apt-transport-https ca-certificates curl  
gnupg-agent software-properties-common
```

- Adicionar Chave GPG do Docker

```
$ curl -fsSL https://download.docker.com/linux/ubuntu/gpg  
| sudo apt-key add -
```

- Configurar repositório

```
$ sudo add-apt-repository "deb [arch=amd64]  
https://download.docker.com/linux/ubuntu $(lsb_release -  
cs) stable"
```

- Instalar o docker

```
$ sudo apt-get update
```

```
$ sudo apt-get install docker-ce docker-ce-cli containerd.io
```

- Alterar permissão

```
$ sudo usermod -aG docker $(whoami)
```

```
$ sudo reboot
```

- Verificar o status

```
sudo systemctl status docker
```


Preparação Ambiente – Instalação Docker Compose - Linux

- <https://docs.docker.com/compose/install/>

Install Compose on Linux systems

On Linux, you can download the Docker Compose binary from the [Compose repository release page on GitHub](#). Follow the instructions from the link, which involve running the `curl` command in your terminal to download the binaries. These step-by-step instructions are also included below.

❗ For `alpine`, the following dependency packages are needed: `py-pip`, `python-dev`, `libffi-dev`, `openssl-dev`, `gcc`, `libc-dev`, `rust`, `cargo` and `make`.

1. Run this command to download the current stable release of Docker Compose:

```
sudo curl -L "https://github.com/docker/compose/releases/download/1.28.2/docker-compose"
```

2. Apply executable permissions to the binary:

```
sudo chmod +x /usr/local/bin/docker-compose
```

Preparação Ambiente – Instalação Docker Compose - Linux

○ Docker Compose no Ubuntu

- Baixar

```
$ sudo curl -L "https://github.com/docker/compose/releases/download/1.28.2/docker-compose-$(uname -s)-$(uname -m)" -o /usr/local/bin/docker-compose
```

- Alterar permissão

```
$ sudo chmod +x /usr/local/bin/docker-compose
```

- Testar instalação

```
$ docker-compose --version
```



Instalação Docker - Mac

Preparação Ambiente – Instalação Docker Desktop - Mac


○ Instalação

- Docker: <https://docs.docker.com/get-docker/>

Get Docker


Docker is an open platform for developing, shipping, and running applications. Docker enables you to separate your applications from your infrastructure so you can deliver software quickly. With Docker, you can manage your infrastructure in the same ways you manage your applications. By taking advantage of Docker's methodologies for shipping, testing, and deploying code quickly, you can significantly reduce the delay between writing code and running it in production.

You can download and install Docker on multiple platforms. Refer to the following section and choose the best installation path for you.




Docker Desktop for Mac

A native application using the macOS sandbox security model which delivers all Docker tools to your Mac.



Docker Desktop for Windows

A native Windows application which delivers all Docker tools to your Windows computer.



Docker for Linux

Install Docker on a computer which already has a Linux distribution installed.

Preparação Ambiente – Instalação Docker Desktop - Mac

- <https://hub.docker.com/editions/community/docker-ce-desktop-mac/>



Docker Desktop for Mac

By **Docker**

The fastest and easiest way to get started with Docker on Mac

Edition

macOS

x86-64

Get Docker Desktop for Mac

Docker Desktop for Mac is available for free.

Docker Desktop - macOS must be version 10.14 or newer: i.e. Mojave (10.14), Catalina (10.15), or Big Sur (11.0). Mac hardware must be a 2010 or a newer model with an Intel processor.

By downloading this, you agree to the terms of the [Docker Software End User License Agreement](#) and the [Docker Data Processing Agreement \(DPA\)](#).



Get Docker



Instalação Cluster Big Data

Baixar Cluster de Big Data

- Baixar conteúdo do Cluster

\$ git clone https://github.com/rodrigo-reboucas/docker-bigdata.git

```
semantix@NTBSTX7158:~/treinamentos
$ git clone https://github.com/rodrigo-reboucas/docker-bigdata.git
Cloning into 'docker-bigdata'...
remote: Enumerating objects: 47, done.
remote: Counting objects: 100% (47/47), done.
remote: Compressing objects: 100% (43/43), done.
remote: Total 1303 (delta 24), reused 7 (delta 4), pack-reused 1256
Receiving objects: 100% (1303/1303), 129.13 MiB | 4.84 MiB/s, done.
Resolving deltas: 100% (844/844), done.
semantix@NTBSTX7158:~/treinamentos
$ cd docker-bigdata/
semantix@NTBSTX7158:~/treinamentos/docker-bigdata
$ ls
README.md  data  docker-compose-completo-windows.yml  docker-compose-completo.yml
docker-compose.yml  ecosystem.jpeg
```

Baixar Imagens do Cluster

- Baixar as imagens
 - docker-compose pull
- Listar as imagens
 - docker image ls

```
semantix@NTBSTX7158:~/treinamentos/docker-bigdata
$ docker-compose pull
Pulling namenode          ... done
Pulling datanode          ... done
Pulling hive-metastore-postgresql ... done
Pulling hive-metastore    ... done
Pulling hive-server       ... done
Pulling database          ... done
Pulling zookeeper         ... done
Pulling hbase-master      ... done
Pulling spark             ... done
```

```
semantix@NTBSTX7158:~/treinamentos/docker-bigdata
$ docker image ls
```

REPOSITORY	TAG	IMAGE ID	CREATED	SIZE
academysemantix/cdh-datanode	latest	636211793711	44 hours ago	943MB

Executar e Listar os Containers do Cluster

- Executar os containers
 - `docker-compose up -d`
- Lista os containers
 - `docker container ls`

```
$ docker-compose up -d
Creating network "docker-bigdata_default" with the default driver
Creating zookeeper      ... done
Creating database       ... done
Creating spark          ... done
Creating namenode      ... done
Creating hbase-master   ... done
Creating datanode       ... done
Creating hive-metastore-postgresql ... done
Creating hive_metastore ... done
Creating hive-server    ... done
```

```
$ docker container ls
```

CONTAINER ID	IMAGE	COMMAND	CREATED	STATUS
0c0871b0ecd4	fjardim/hive	"entrypoint.sh /bin/..."	About a minute ago	Up About a minute
3589dba8a75b	fjardim/hive	"entrypoint.sh /opt/..."	About a minute ago	Up About a minute
d2c0f331694e	fjardim/hive-metastore	"/docker-entrypoint..."	About a minute ago	Up About a minute
2480e1629054	fjardim/hbase-master	"/entrypoint.sh /run..."	About a minute ago	Up About a minute
d9f035edd7ac	fjardim/datanode	"/entrypoint.sh /run..."	About a minute ago	Up About a minute (healthy)
18858c0bd728	fjardim/mysql	"docker-entrypoint.s..."	About a minute ago	Up About a minute
3d6193034ef1	fjardim/jupyter-spark	"/opt/docker/bin/ent..."	About a minute ago	Up About a minute
c82b5d0f5aee	fjardim/namenode_sqoop	"/entrypoint.sh /run..."	About a minute ago	Up About a minute (healthy)
5c9fb535be3d	fjardim/zookeeper	"/bin/sh -c '/usr/sb..."	About a minute ago	Up About a minute

Opções Docker Compose

- Iniciar todos os serviços em background (-d)
\$ docker-compose up -d
- Parar os serviços
\$ docker-compose stop
- Iniciar os serviços
\$ docker-compose start
- Término do treinamento
 - Matar os serviços
\$ docker-compose down
 - Apagar todos os volumes sem uso
\$ docker volume prune
 - Apagar tudo (image, volume, network)
\$ docker system prune --all

Acessos Ambiente docker

- Visualizar os container
 - Ativos
 - `$ docker ps`
 - Todos
 - `$ docker ps -a`
- Executar comandos no container
 - `$ docker exec -it <container> <comando>`
- Visualizar os logs
 - `$ docker logs <container>`
 - `$ docker-compose logs`
- Enviar arquivos
 - `$ docker cp <diretório> <container>:/<diretório>`
- Acesso o container namenode
 - `docker exec -it namenode bash`
- Acesso o container do hive
 - `docker exec -it hive-server bash`

Exercícios Instalação de Ambiente

1. Instalação do docker e docker-compose
 - Docker: <https://docs.docker.com/get-docker/> (Links para um site externo.)
 - Docker-compose: <https://docs.docker.com/compose/install/> (Links para um site externo.)
2. Executar os seguintes comandos, para baixar as imagens do Cluster de Big Data:
 - `git clone https://github.com/rodrigo-reboucas/docker-bigdata.git`
 - `cd docker-bigdata`
 - `docker-compose pull`
3. Iniciar o cluster Hadoop através do docker-compose
 - `docker-compose up -d`
4. Listas as imagens em execução
5. Verificar os logs dos containers do docker-compose em execução
6. Verificar os logs do container namenode
7. Acessar o container namenode
8. Listar os diretórios do container namenode
9. Parar os containers do Cluster de Big Data



Semantix

Obrigado!

Alguma pergunta?



Você pode me encontrar em:
rodrigo.augusto@semantix.com.br

GET SMARTER