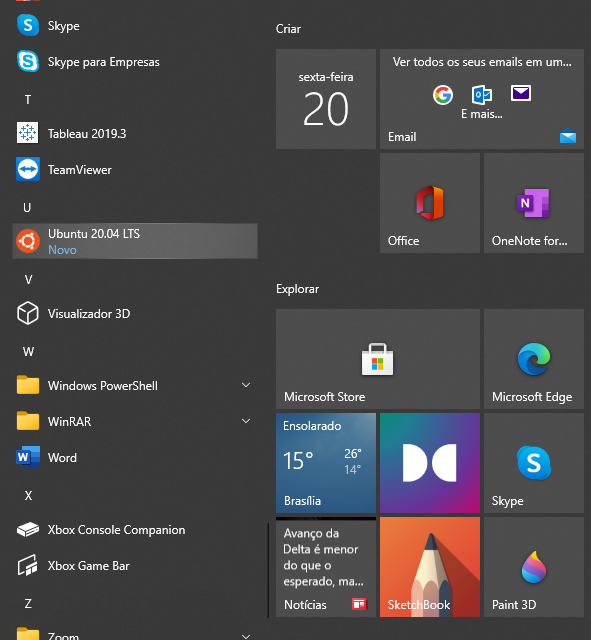
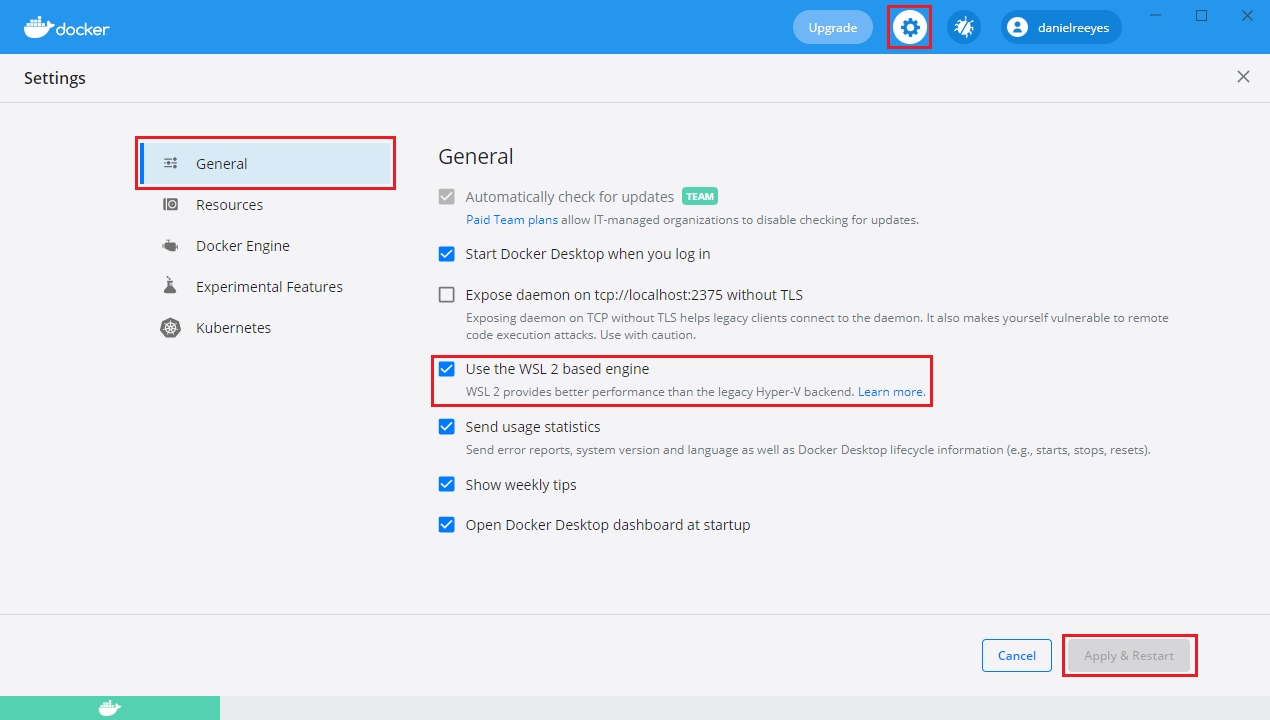
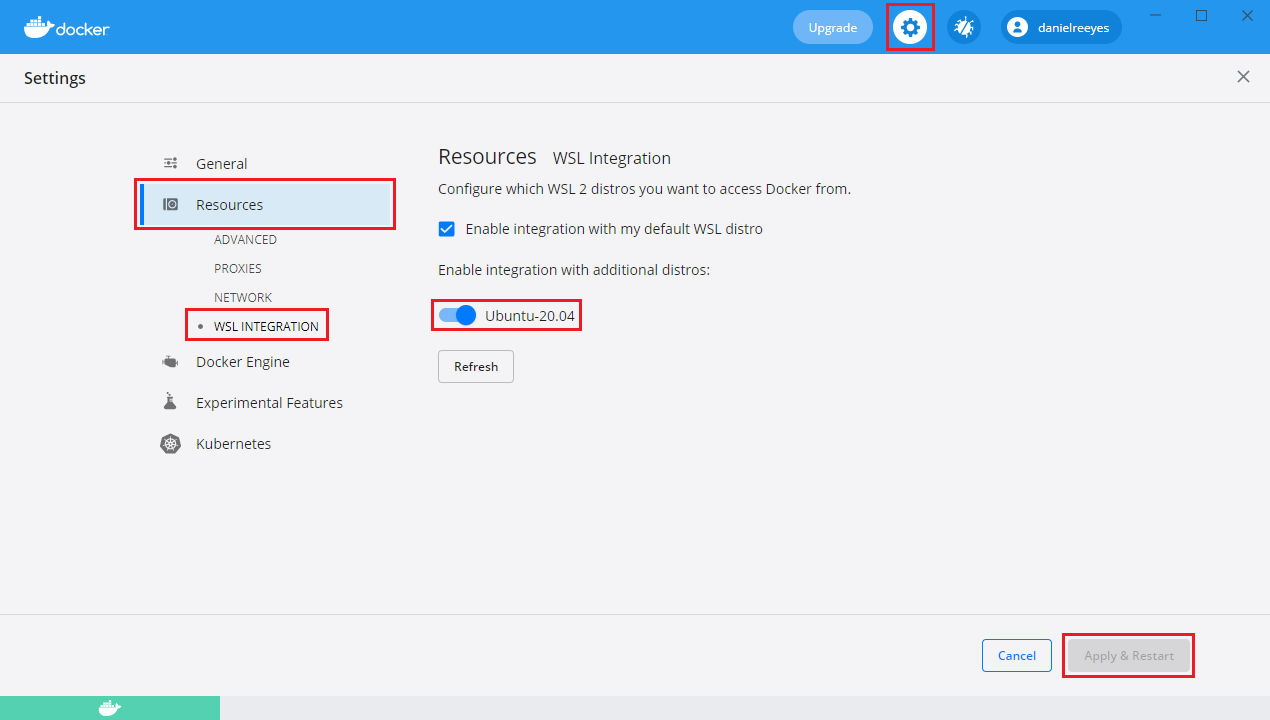
**Requisitos usando Windows**

* Testado em Windows 10
* Habilitar suporte a WSL2: O Hadoop e seus daemons foram desenvolvidos em Linux, dessa forma, precisamos simular o Linux em nossa máquina para que consigamos executar o Hadoop e seus componentes. Quem permite que executemos o Linux dentro do ambiente Windows é o WSL. Assim, siga os 6 passos disponível no link da Microsoft a seguir:
  + <https://docs.microsoft.com/pt-br/windows/wsl/install-win10#manual-installation-steps>
  + Na etapa 6, instalar o Ubuntu 20.04 LTS
    - O Ubuntu 20.04 é a versão que roda sobre o WSL2 no Windows. O Docker utiliza essa versão internamente para disparar os containers.
  + O Linux estará instalado dentro do Windows e executará como se fosse um programa instalado. Veja no menu iniciar:
* Instalação dos softwares: Como dito no passo anterior, o Ubuntu instalado anteriormente vai apenas disparar os containers que serão criados durante o Notebook 1 (Hadoop). Para isso precisamos que seja instalado o gerenciador de Dockers para o Windows (Docker Desktop):
  + Docker Desktop (<https://www.docker.com/>)
    - <https://docs.docker.com/desktop/windows/install/>
  + Habilitar as configurações do Docker Desktop para funcionar com a distribuição do Linux. Abaixo alguns tutoriais de como fazer:
    - Tutorial 1: <https://docs.docker.com/desktop/windows/wsl/>
    - Tutorial 2: <https://docs.microsoft.com/pt-br/windows/wsl/tutorials/wsl-containers>
    - Tutorial 3: <https://andrewlock.net/installing-docker-desktop-for-windows/>
      * Habilitar o suporte a WSL2 (Settings -> General -> Use the WSL 2 based engine -> Apply & Restart). São os passos 3 e 4 do tutorial 1 e passo 3 do tutorial 2.

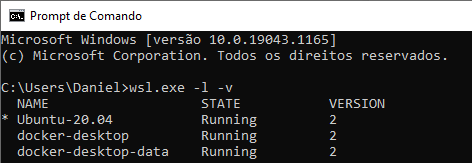


* + - * Habilitar integração com o Ubuntu instalado juntamente com o WSL 2 (Settings -> Resources -> WSL Integration -> Ativar o Ubuntu-20.04 -> Apply & Restart). Passo 4 do tutorial 2.



Antes de finalizar a instalação do ambiente, execute os passos 6 e 7 do tutorial 1. Eles garantem que a instalação do Linux está de acordo com o que precisamos.

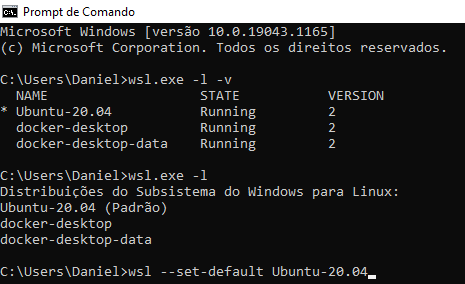
No prompt, executar o comando wsl.exe -l -v. Este comando irá listar as distribuições que temos ativa com o WSL. O resultado, deve ser o abaixo (caso você nunca tenha usado uma distribuição Docker):



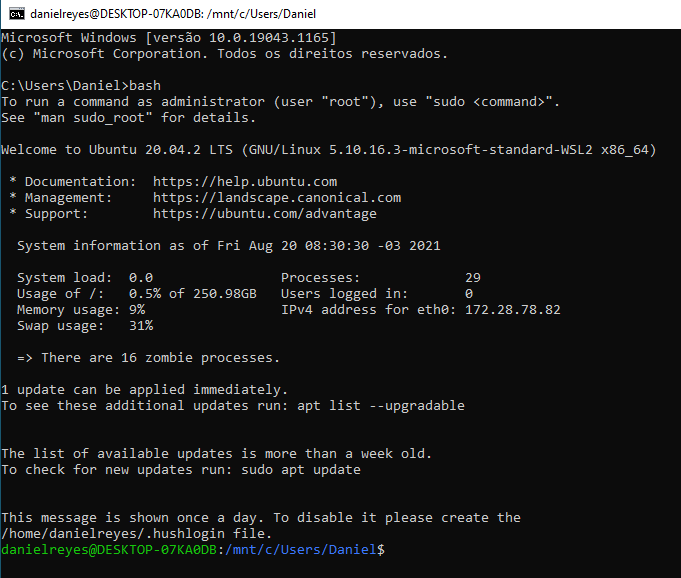
A distribuição também deve estar na versão 2. Para configurar essa opção, siga o passo 6 do tutorial 1 da Microsoft.

O símbolo \* ao lado do Ubuntu-20.04 indica que ela é a integração padrão. Essa configuração padrão é fundamental, pois ela que permitirá que o jupyter notebook execute comandos bash (Linux) direto no prompt de comando, executando o Linux instalado anterior (via prompt de comando).

Para configurar a distribuição padrão, execute o passo 7 do tutorial 1.



Em seguida, reinicie o sistema e faça um teste para verificar se a integração está funcionando. Abra o prompt de comando e execute o comando bash. Esse comando irá abrir o terminal do Linux ubuntu 20.04 instalado na sua máquina anteriormente como na imagem a seguir:



* + Python 3.6 ou superior
    - Sugestão: usar instalador miniconda (<https://docs.conda.io/en/latest/miniconda.html>)
    - Alternativa: <https://python.org.br/instalacao-windows/>
  + Jupyterlab (<https://jupyter.org/>)
    - <https://jupyterlab.readthedocs.io/en/stable/getting_started/installation.html>
    - <https://jupyter.org/install>
* Faça o download do repositório: <https://github.com/tiagoferreto/HadoopJupyter.git>
* Executar Jupyter Lab ou o Jupyter Notebook e abrir notebook 1.HadoopDocker.ipynb

Considerações Finais:

O Ubuntu 20.04 é a versão que roda sobre o WSL2 no Windows. O Docker utiliza essa versão internamente para disparar os containers.

O Notebook 1 (Hadoop) cria uma imagem base de um container Docker com o ubuntu 18.04. Depois no notebook essa imagem base é utilizada para criar o cluster com 4 containers (1 mestre e 3 escravos).

Não existe nenhuma relação direta entre as versões (20.04 e 18.04). Teremos uma "VM" Ubuntu 20.04 rodando sobre o WSL que vai instanciar containers Ubuntu 18.04 para simular o cluster Hadoop.