

# **Banco de Dados 2 – Projeto de Implementação Parte A**

**Davi Kallebe Motta da Silva**

Instituto de Computação – Universidade Federal do Amazonas (UFAM)

CEP 69077-000 – Manaus – AM – Brasil

**dkmds@icomp.ufam.edu.br**

Todos os comandos abaixo foram executados com sucesso, mas com as seguintes modificações:

- O número de works foi diminuído para 3, pois segundo o GCP o disponível era 4 e não 6.
- Criado um bucket com o nome ufam-davi-bgd-2020-bucket-1
- Criado um projeto com o nome ufam-davi-bgd-2020
- Criado um cluster com o nome davi-kallebe-0411

Criação do clusters

```
gcloud dataproc clusters create davi-kallebe-0411 \  
--project ufam-davi-bgd-2020 \  
--region us-central1 \  
--subnet default \  
--zone us-central1-c \  
--master-machine-type n1-standard-1 \  
--master-boot-disk-size 32 \  
--num-workers 3 \  
--worker-machine-type n1-standard-1 \  
--worker-boot-disk-size 32 \  
--image-version 1.4-ubuntu18 \  
--optional-components ANACONDA,JUPYTER \  
--bucket ufam-davi-bgd-2020-bucket-1
```

Wordcount.py foi executado sem maiores problemas

```
gcloud dataproc jobs submit pyspark wordcount.py \  
--cluster=davi-kallebe-0411 -- \  
gs://la-gcp-labs-resources/data-engineer/dataproc/romeoandjuliet.txt \  
gs://ufam-davi-bgd-2020-bucket-1/output/
```

Foi executado mas foi necessário colocar a flag -4 para criar IPV4, pois aparentemente o GCP só criava IPV6, mas mesmo assim não consegui realizar um ssh.

```
gcloud compute ssh davi-kallebe-0411-m --project=ufam-davi-bgd-2020 \  
--zone=us-central1-c -- -D 1080 -N -4
```