**Tarefa 1. Preparar Variáveis ​​de Ambiente**

1. No console, no **menu Navegação** ( 7a91d354499ac9f1.png), clique em **Compute Engine** > **instâncias de VM** .
2. Localize a linha com a instância chamada **training\_vm** .
3. Na extrema direita, em 'conectar', clique em **SSH** para abrir uma janela de terminal.
4. Neste laboratório, você inserirá os comandos da CLI no **training\_vm** .

Por que o Cloud Shell não é usado neste laboratório?

Ao contrário do Compute Engine, o Cloud Shell não possui SLA. Portanto, a disponibilidade da VM do Cloud Shell não pode ser garantida durante o laboratório.

<https://cloud.google.com/shell/sla>

<https://cloud.google.com/compute/sla>

As atividades que você realizará neste laboratório precisarão de vários valores, como o nome de um intervalo do Cloud Storage para arquivos temporários, de hospedagem e de saída, e o endereço IP do seu computador / navegador necessário para as regras de firewall.

Como prática recomendada, você definirá as variáveis ​​de ambiente em um arquivo, **myenv** , onde poderá recuperá-las facilmente, se necessário, usando o comando: **source myenv**

Crie o arquivo de origem para configurar e redefinir variáveis ​​de ambiente

1. Na janela do terminal **training\_vm** SSH, usando seu editor de linha de comandos preferido, crie e edite o arquivo para manter suas variáveis ​​de ambiente. Por exemplo:
2. cd ~
3. nano myenv

Identifique um projeto

One environment variable that you will set is \_\_$PROJECT\_ID\_\_ that contains the Google Cloud project ID required to access billable resources.

1. No console, no **menu Navegação** ( 7a91d354499ac9f1.png), clique em **Página inicial** . No painel com Informações do **projeto** , o **ID** do **projeto** é listado. Você também pode encontrar essas informações na guia Qwiklabs, em Detalhes da conexão, onde ela é rotulada como **ID do projeto do GCP** .
2. Adicione a variável de ambiente ao **myenv** para facilitar a referência.
3. PROJECT\_ID=<project ID>

Verificar um intervalo para o armazenamento temporário do Dataproc

O Dataproc pode usar um intervalo do Cloud Storage para organizar seus arquivos durante a inicialização. Você pode usar esse intervalo para preparar programas de aplicativos ou dados para uso pelo Dataproc. O bucket também pode hospedar scripts e saída de inicialização do Dataproc. O nome do intervalo deve ser globalmente exclusivo. A Qwiklabs já criou um intervalo para você que tem o mesmo nome da ID do projeto, que já é globalmente exclusiva.

1. No console, no **menu Navegação** ( 7a91d354499ac9f1.png), clique em **Armazenamento> Navegador** . Verifique se o balde existe. Observe a classe de armazenamento padrão e a localização (região) desse intervalo. Você estará usando as informações desta região a seguir.
2. Adicione a linha ao **myenv** para criar uma variável de ambiente chamada "BUCKET".
3. BUCKET=<project ID>

Você pode usar $ BUCKET nos comandos da CLI. E se você precisar inserir o nome do intervalo <your-bucket> em um campo de texto no Console, poderá recuperar rapidamente o nome com **echo $ BUCKET** .

Identifique uma região e zona

Você estará criando um cluster do Dataproc em uma região específica. O cluster do Dataproc e o bloco que ele usará para o armazenamento temporário devem estar na mesma região. Como o bloco que você está usando já existe, você precisará corresponder a variável de ambiente **$ MYREGION** à região do bucket.

1. Você pode encontrar a região na guia Qwiklabs em Detalhes da conexão, chamada **Região QL** .

A zona deve estar na mesma região.

1. Você pode encontrar a zona na guia Qwiklabs em Detalhes da conexão, chamada **Zona QL** .
2. Adicione as variáveis ​​de ambiente ao **myenv** para facilitar a referência.
3. MYREGION=<region>
4. MYZONE=<zone>

Identifique o endereço IP do navegador

Você usará o endereço IP do navegador para permitir que seu navegador local alcance o cluster do Dataproc.

1. Encontre o endereço IP do navegador do seu computador abrindo uma janela do navegador e visualizando <http://ip4.me/> Copie o endereço IP.
2. Adicione a linha ao **myenv** para criar uma variável de ambiente denominada **BROWSER\_IP** .
3. BROWSER\_IP=<your-browser-ip>
4. Depois de ter adicionado todas as três definições ao **myenv** e salvo o arquivo, use o comando source para criar as variáveis ​​de ambiente.

Definir e verificar variáveis ​​de ambiente

1. Defina as variáveis ​​de ambiente.
2. source myenv
3. Verifique se os valores estão definidos com eco.
4. echo $PROJECT\_ID
5. echo $MYREGION $MYZONE
6. echo $BUCKET
7. echo $BROWSER\_IP

Você pode usar as variáveis ​​de ambiente nos comandos e scripts da CLI. No entanto, o console não usa as variáveis ​​de ambiente. Por exemplo, se uma caixa de texto no Console exigisse um nome de bloco, você poderia usar echo $ BUCKET para recuperar o valor, mas seria necessário copiá-lo e colá-lo na caixa de texto no Console. Se você inseriu $ BUCKET na caixa de texto, ele seria usado literalmente e geraria um erro.

**Tarefa 2. Criar um cluster do Dataproc**

Criar um cluster do Dataproc Hadoop personalizado para usar a API do Google Cloud

Este laboratório faz várias alterações a partir dos valores padrão. Ele usa um mestre e três nós de trabalho. O mestre é um nó n1-padrão-2 com armazenamento de 100 GB e os nós do trabalhador são instâncias n1-padrão-1 com 50 GB de armazenamento em cada nó. Na produção, você selecionaria vCPUs e tamanhos de dados para refletir os requisitos da sua aplicação. Geralmente, mais recursos produzem resultados mais rapidamente e menos recursos custam menos.

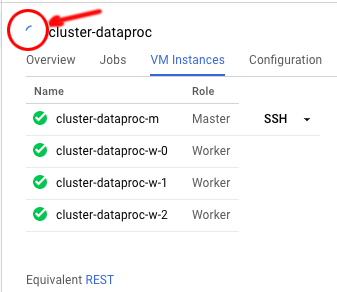
1. No Console, no **menu Navegação** ( 7a91d354499ac9f1.png), clique em **Dataproc> Clusters** .
2. Clique em **Criar Cluster** .
3. Especifique o seguinte e deixe as configurações restantes como seus padrões:

|  |  |
| --- | --- |
| **Propriedade** | **Valor**  (digite o valor ou selecione a opção conforme especificado) |
| **Nome** | cluster-dataproc |
| **Região** | <your region> |
| **Zona** | <your zone> |
| **Modo de cluster** | Standard (1 Master, N workers) |
| **(Nó mestre) Tipo de máquina** | n1-standard-2 |
| **(Nó mestre) Tamanho do disco principal** | 100 GB |
| **(Nós do trabalhador) Tipos de máquinas** | n1-standard-1 |
| **(Nós do trabalhador) Tamanho do disco principal** | 50 GB |
| **Nós (mínimo 2)** | 3 |

1. Clique nas **opções avançadas** .
2. Especifique o seguinte e deixe as configurações restantes como seus padrões:

|  |  |
| --- | --- |
| **Propriedade** | **Valor**  (digite o valor ou selecione a opção conforme especificado) |
| **Tags de rede** | hadoopaccess |
| **Balde de teste do Cloud Storage** | <your-bucket> |
| **Imagem** | 1.2 (Default) |
| **Acesso ao projeto** | [x] Allow API access to all Google Cloud services in the same project |

1. Clique em **Criar** .
2. O cluster levará vários minutos para se tornar operacional. No Console, no **menu Navegação** ( 7a91d354499ac9f1.png), clique em **Dataproc> Clusters** .
3. Clique no seu cluster, **cluster-dataproc** . Em seguida, clique na guia **Instâncias** da **VM** . As instâncias se tornarão operacionais antes que o software hadoop tenha concluído a inicialização. Quando uma marca de seleção em um círculo verde aparece ao lado do nome do cluster, ele fica operacional.



1. Levará alguns minutos a mais para a inicialização ser concluída.

**Tarefa 3. Ativar Acesso Seguro ao Cluster Dataproc**

Criar uma regra de firewall restritiva usando tags de destino, endereço IP e protocolo

Crie uma regra de firewall que permita acesso somente ao nó mestre do endereço IP do seu computador. Somente as portas 8088 (Interface de trabalho do Hadoop) e 9870 (interface do Administrador do Hadoop) serão permitidas.

A porta 8042 é a interface do usuário da web para o gerenciador de nó nos nós do trabalhador e a porta 8080 é a porta padrão do Datalab. O Datalab é um ambiente de desenvolvimento integrado baseado em notebook, derivado dos notebooks da Jupyter. É uma ferramenta comum para o desenvolvimento de aplicativos Dataproc. O curso Aprendizado de máquina sem servidor no GCP usa o Datalab extensivamente.

1. Recupere o endereço IP do navegador do seu computador para uso no Console.
2. echo $BROWSER\_IP
3. No console, no **menu Navegação** ( 7a91d354499ac9f1.png), clique em **Rede VPC** > **Regras de firewall** .
4. Clique em **Criar regra de firewall** .
5. Especifique o seguinte e deixe as configurações restantes como seus padrões:

|  |  |
| --- | --- |
| **Propriedade** | **Valor**  (digite o valor ou selecione a opção conforme especificado) |
| **Nome** | allow-hadoop |
| **Rede** | default |
| **Prioridade** | 1000 |
| **Direção de tráfego** | Ingress |
| **Ação no jogo** | Allow |
| **Alvos** | Specified target tags |
| **Tags alvo** | hadoopaccess |
| **Intervalos de IP de origem** | <your-IP>/32 |
| **Protocolos e portas especificados** | Verifique tcpe insira o número da porta 9870,8088 |

1. Clique em **Criar** .

Levará alguns minutos para que a regra de firewall se torne ativa.

Verifique se a tag de rede está definida no nó mestre

Verifique se a tag de rede "hadoopaccess" está definida no nó mestre. Isso aplicará a regra de firewall ao nó mestre, dando acesso ao seu laptop.

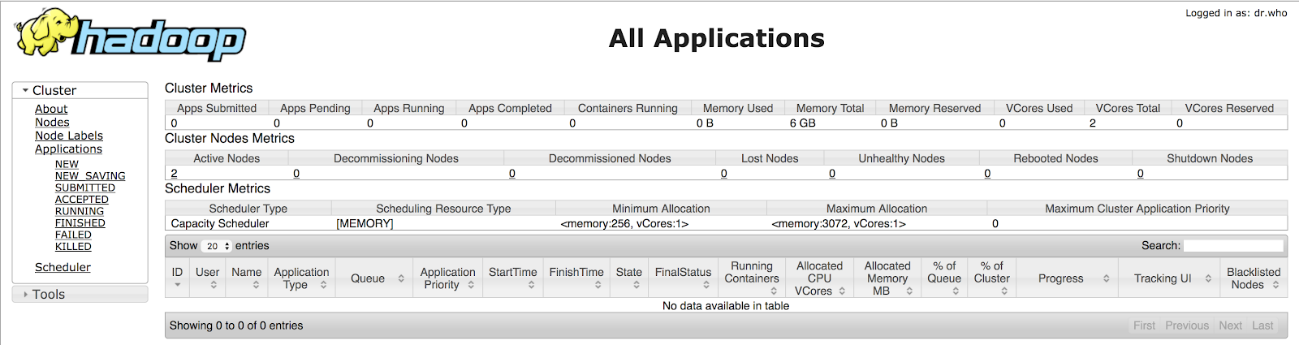
1. No console, no **menu Navegação** ( 7a91d354499ac9f1.png), clique em **Compute Engine** > **instâncias de VM** .
2. Clique no nó mestre, **cluster-dataproc-m** .
3. Verifique se, em Network Tags, ele lista **hadoopaccess** .
4. Se a tag não estiver lá, clique em **EDITAR** .
5. Em Network Tags, adicione a tag: **hadoopaccess**
6. Clique em **Salvar** .

**Tarefa 4. Explorar operações do Hadoop**

Navegue até o nó mestre

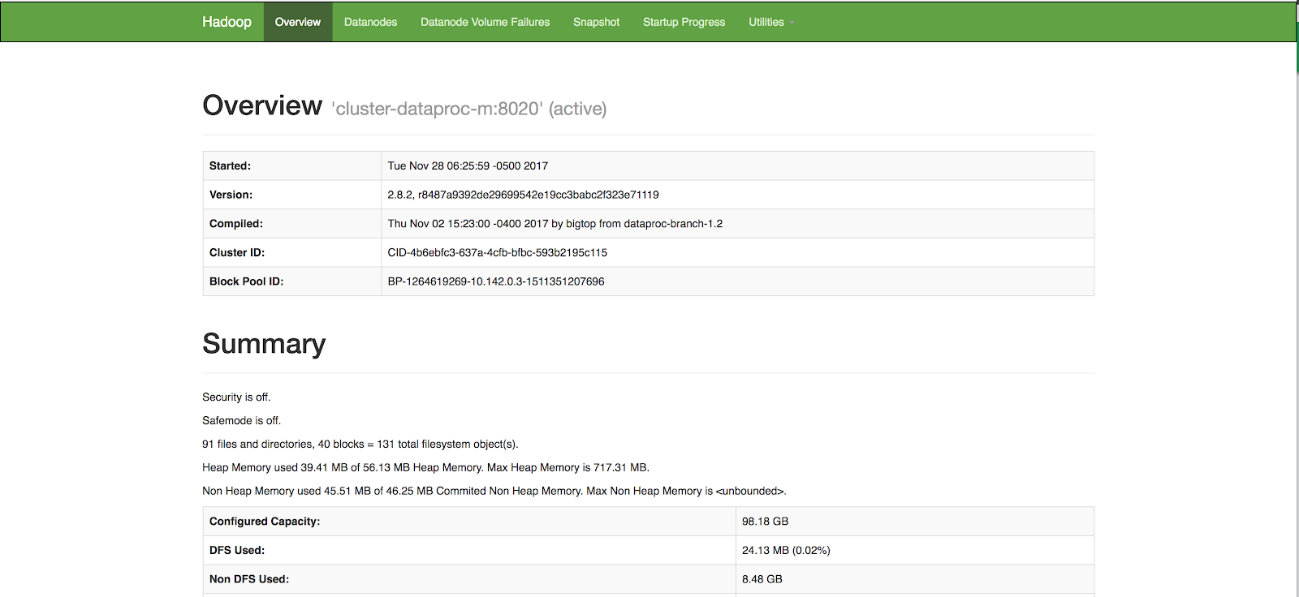
Nesta tarefa, você explorará os serviços integrados de operação e administração do Hadoop. Esta tarefa e as próximas duas ajudarão você a entender o relacionamento entre o Dataproc no console, a instância do Nó Principal e os serviços do Hadoop. Se você já trabalhou com o Hadoop antes, as interfaces de usuário da web serão familiares.

1. No Console, no **menu Navegação** ( 7a91d354499ac9f1.png), clique em **Compute Engine** > **Instâncias de VM** .
2. Na lista de instâncias de VMs, na linha do **cluster-dataproc-m** , destaque o **IP externo** e copie-o.
3. Abra uma nova aba ou janela do navegador e cole o IP externo. Adicione ": 8088" após o IP e pressione enter. Exemplo: <IP externo>: 8088 A página da web exibida é a interface do Hadoop Applications e deve ser algo como isto:



Você pode visualizar todos os trabalhos enviados ao cluster e seus status nessa página.

1. Abra uma nova guia ou janela do navegador. Cole o IP externo. Adicione ": 9870" após o IP e pressione Enter. Exemplo: <IP externo>: 9870 A página da Web exibida é a Interface de Administração do Hadoop e deve ser algo como isto:



1. Clique na guia **Datanodes** . Isso mostrará quanto de capacidade está sendo usada nos nós de trabalho HDFS (Hadoop Distributed File System) e quanto de capacidade permanece.
2. Clique em **Utilitários> Logs** . Isso mostra os arquivos de log do Hadoop para cada nó no cluster. É aqui que você pode investigar problemas com o Hadoop. Use o botão Voltar do seu navegador para retornar ao console de administração do Hadoop.
3. Clique em **Utilitários> Navegar no sistema de arquivos** . Após alguns instantes, o sistema de arquivos aparecerá na página do navegador. Você pode usar isso para navegar no sistema de arquivos. Na linha que diz Owner is **hdfs** e Group is **hadoop**, clique no link que diz **usuário** . Aqui você pode ver diretórios para todos os serviços do hadoop.

**Termine seu laboratório**

Quando você tiver concluído seu laboratório, clique em **Finalizar Laboratório** . O Qwiklabs remove os recursos que você usou e limpa a conta para você.

Você terá a oportunidade de avaliar a experiência do laboratório. Selecione o número de estrelas aplicável, digite um comentário e clique em **Enviar** .

O número de estrelas indica o seguinte:

* 1 estrela = muito insatisfeito
* 2 estrelas = insatisfeito
* 3 estrelas = neutra
* 4 estrelas = Satisfeito
* 5 estrelas = muito satisfeito

Você pode fechar a caixa de diálogo se não quiser fornecer feedback.

Para comentários, sugestões ou correções, use a guia **Suporte** .

Última atualização: 2018-09-11

© 2018 Google LLC Todos os direitos reservados. Google e o logotipo do Google são marcas registradas do Google LLC. Todos os outros nomes de empresas e produtos podem ser marcas registradas das respectivas empresas com as quais estão associados.