**Tarefa 1: preparação**

Você estará executando um simulador de sensores da VM de treinamento. Existem vários arquivos e algumas configurações do ambiente necessário.

**Abra o terminal SSH e conecte-se à VM de treinamento**

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, na coluna **Conectar** , clique em **SSH** para abrir uma janela de terminal.
4. Neste laboratório, você inserirá os comandos da CLI no **training-vm** .

**Verificar se a inicialização está completa**

1. O **training-vm** está instalando software em segundo plano. Verifique se a configuração está concluída, verificando se o diretório a seguir existe. Se não existir, aguarde alguns minutos e tente novamente.

ls /training

Aguarde até que a configuração esteja concluída antes de prosseguir. Você pode verificar a instalação do maven com o **mvn -version** e o JDK com **java -version** .

**Copie arquivos**

1. Um repositório foi baixado para a VM. Copie o repositório para o seu diretório inicial.

cp -r /training/training-data-analyst/ .

**Definir variáveis ​​de ambiente**

1. No terminal SSH **training-vm** , digite o seguinte:

source /training/project\_env.sh

Este script define as variáveis ​​de ambiente **$ DEVSHELL\_PROJECT\_ID** e **$ BUCKET** .

**Preparar arquivos de início rápido do HBase**

1. No terminal SSH **training-vm** , execute o script para baixar e descompactar os arquivos de início rápido (você os usará posteriormente para executar o shell do HBase).

cd ~/training-data-analyst/courses/streaming/process/sandiego

./install\_quickstart.sh

**Tarefa 2: Simular dados do sensor de tráfego no Pub / Sub**

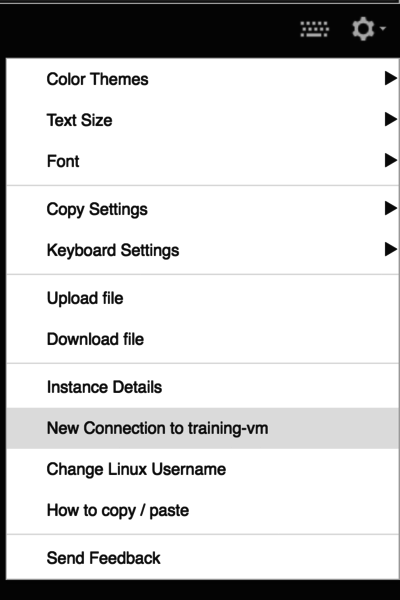
1. No terminal SSH **training-vm** , inicie o simulador do sensor. O script lê dados de amostra de um arquivo csv e os publica no Pub / Sub.

/training/sensor\_magic.sh

Este comando irá enviar 1 hora de dados em 1 minuto. Deixe o script continuar a rodar no terminal atual.

**Abra um segundo terminal SSH e conecte-se à VM de treinamento**

1. No canto superior direito do terminal SSH **training-vm** , clique no botão em forma de engrenagem ( 9649d58acf1c4e06.png) e selecione **New Connection to training-vm** no menu suspenso. Uma nova janela de terminal será aberta.



1. A nova sessão de terminal não terá as variáveis ​​de ambiente necessárias. Execute o seguinte comando para defini-los.
2. No novo terminal SSH **training-vm** , digite o seguinte:

source /training/project\_env.sh

**Tarefa 3: Iniciar Pipeline do Dataflow**

1. No segundo terminal SSH **training-vm** , navegue até o diretório deste laboratório. Examine o script no Cloud Shell ou usando o nano. **Não faça alterações no código.**

cd ~/training-data-analyst/courses/streaming/process/sandiego

nano run\_oncloud.sh

O que o script faz?

1. O script recebe 3 argumentos obrigatórios: id do projeto, nome do intervalo, nome da classe e possivelmente um quarto argumento: opções. Nesta parte do laboratório, usaremos a --bigtableopção que direcionará o pipeline para gravar no Cloud Bigtable.
2. Execute o seguinte script para criar a instância do Bigtable.

cd ~/training-data-analyst/courses/streaming/process/sandiego

./create\_cbt.sh

1. Execute o pipeline do Dataflow para ler o PubSub e gravar no Cloud Bigtable.

cd ~/training-data-analyst/courses/streaming/process/sandiego

./run\_oncloud.sh $DEVSHELL\_PROJECT\_ID $BUCKET CurrentConditions --bigtable

Exemplo de execução bem sucedida:

[INFO] ------------------------------------------------------------------------

[INFO] BUILD SUCCESS

[INFO] ------------------------------------------------------------------------

[INFO] Total time: 47.582 s

[INFO] Finished at: 2018-06-08T21:25:32+00:00

[INFO] Final Memory: 58M/213M

[INFO] ------------------------------------------------------------------------

**Tarefa 4: Explorar o pipeline**

1. Volte para a guia do navegador para o Console. No **menu Navegação** ( 7a91d354499ac9f1.png), clique em **Fluxo de dados** e clique no novo trabalho do pipeline. Confirme se o trabalho do pipeline está listado e verifique se ele está sendo executado sem erros.
2. Encontre a etapa **write: cbt** no gráfico de pipeline e clique na seta para baixo à direita para ver o gravador em ação. Clique em determinado escritor. Revise as **opções** do **Bigtable** no **resumo** da **etapa** .

**Tarefa 5: dados do Bigtable de consulta**

1. No segundo terminal SSH **training-vm** , execute o script **quickstart.sh** para ativar o shell do HBase.

cd ~/training-data-analyst/courses/streaming/process/sandiego/quickstart

./quickstart.sh

1. Se o script for executado com êxito, você estaria em um prompt de shell do HBase com a seguinte aparência:

hbase(main):001:0>

1. No prompt do shell do HBase, digite a consulta a seguir para recuperar 2 linhas de sua tabela do Bigtable que foi preenchida pelo pipeline.

scan 'current\_conditions', {'LIMIT' => 2}

1. Revise a saída. Observe que cada linha é dividida em colunas, registros de data e hora, combinações de valores.
2. Execute outra consulta. Desta vez, observe apenas a **coluna lane: speed** , limite a 10 linhas e especifique **padrões rowid** para as linhas inicial e final a serem verificadas.

scan 'current\_conditions', {'LIMIT' => 10, STARTROW => '15#S#1', ENDROW => '15#S#999', COLUMN => 'lane:speed'}

1. Revise a saída. Observe que você vê 10 da coluna, data e hora, combinações de valores, todas as quais correspondem à Highway 15. Observe também que a coluna está restrita a **lane: speed** .
2. Sinta-se à vontade para executar outras consultas se você estiver familiarizado com a sintaxe. Quando estiver satisfeito, quitsaia do shell.

quit

**Limpar**

1. Execute o script para excluir sua instância do Bigtable. Se solicitado, pressione Enter.

cd ~/training-data-analyst/courses/streaming/process/sandiego

./delete\_cbt.sh

1. Na sua página do Dataflow no seu Cloud Console, clique no nome do job do pipeline e clique no Stop jobno painel à direita.
2. Volte para a primeira guia do Cloud Shell com o editor e digite Ctrl+Cpara pará-lo.
3. Vá para o console do BigQuery e exclua as demonstrações de conjuntos de dados.

**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** .

Data da última atualização: 2018-09-25

Data do último teste: 2018-09-19

© 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.