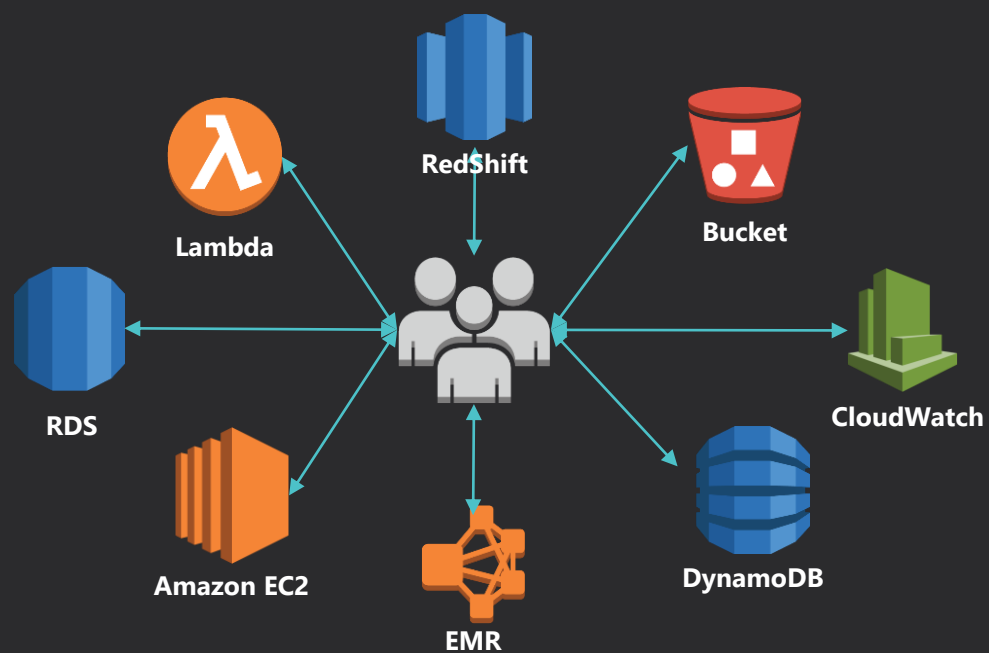


PowerOfData

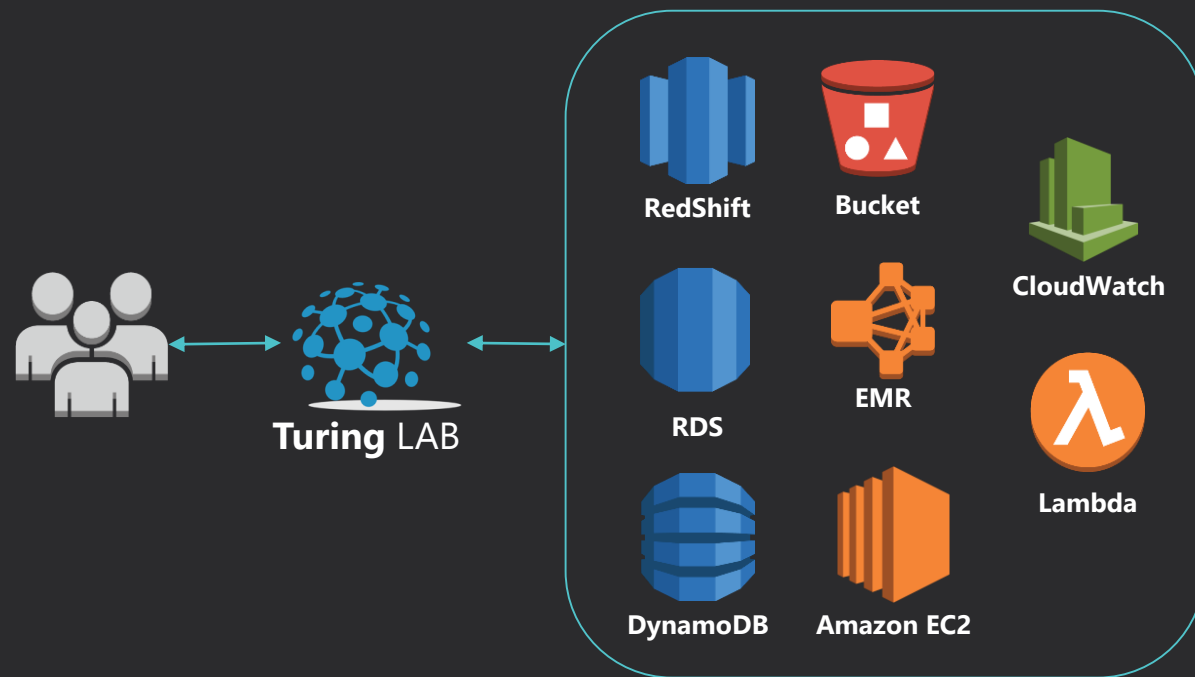
TURING LAB | Data + Analytics lab com alta produtividade

Procedures proprietárias + integração tecnológica amazon cloud

Amazon Cloud



Turing LAB




1) Faça Login na AWS.

Link: <https://983445661575.signin.aws.amazon.com/console>

User: turing-updata-treinamentos

Senha: Turing@BigData.2019



The image shows a screenshot of the AWS login page. At the top is the AWS logo. Below it, there are three input fields: 'Conta:' with the value '983445661575', 'Nome de usuário:' with the value 'turing-updata-treinamentos', and 'Senha:' with masked characters. A blue button labeled 'Fazer login' is positioned below the password field. At the bottom, there are two links: 'Fazer login usando as credenciais da conta raiz' and 'Esqueceu sua senha?'.

aws

Conta:

983445661575

Nome de usuário:

turing-updata-treinamentos

Senha:

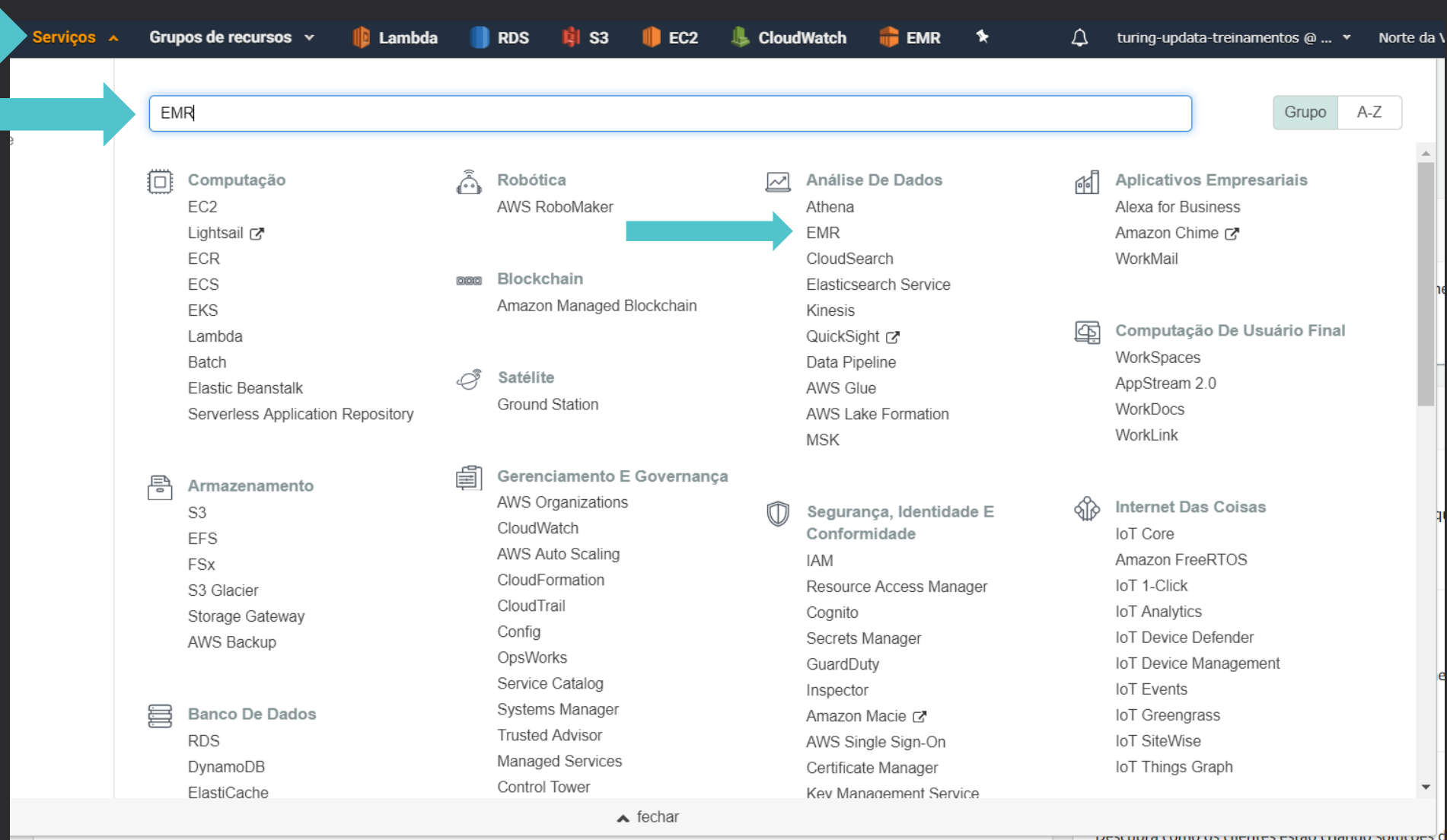
.....

Fazer login

Fazer login usando as credenciais da conta raiz

Esqueceu sua senha?

2) Clique em Serviços e procure por EMR



3) Selecione o cluster que tenha o nome da empresa que você trabalha e clique nele.

É possível usar o Catálogo de dados do AWS Glue como o metastore externo do Hive para cargas de trabalho do Apache Spark. Para começar, basta selecionar o Catálogo de dados do AWS Glue para obter os metadados da tabela ao criar o cluster.

[Criar cluster](#) [Visualizar detalhes](#) [Clonar](#) [Encerrar](#)

Filter: Clusters ativos 3 clusters (todos carregados)


	Nome	ID	Status
<input type="checkbox"/>	Treinamento - Turing Lab - Clube dos Dados	j-3VT6V5YXNWJ3R	Bootstrapping
<input type="checkbox"/>	Treinamento - Turing Lab - Basf	j-3A04GSHFXTOPQ	Bootstrapping
<input type="checkbox"/>	Treinamento - Turing Lab - PowerOfData	j-22Z3WDFWQKLSW	Bootstrapping

Ao clicar no cluster, a tela abaixo será exibida. Clique no link **Habilitar conexão web** e siga as instruções que vão aparecer.

[Clonar](#) [Encerrar](#) [Exportação de CLI da AWS](#)

Cluster: Treinamento - Turing Lab - PowerOfData **Aguardando** Cluster ready after last step completed.

[Resumo](#) [Histórico do aplicativo](#) [Monitoramento](#) [Hardware](#) [Configurações](#) [Eventos](#) [Etapas](#) [Ações de bootstrap](#)

 [Habilitar conexão da web](#) – Zeppelin, Servidor de histórico do Spark, Ganglia, JupyterHub, Gerenciador de recursos ... (Visualizar tudo)

DNS público principal [ec2-54-156-32-219.compute-1.amazonaws.com](#) [SSH](#)

Tags: Treinamento = Turma02 [Visualizar todas/Editar](#)

Resumo	Detalhes da configuração	Rede e hardware
<p>ID: j-22Z3WDFWQKLSW</p> <p>Data de criação: 2019-08-20 09:22 (UTC-3)</p> <p>Tempo decorrido: 16 minutos</p> <p>Encerramento automático: Não</p> <p>Proteção contra encerramento: Desativado Alterar</p>	<p>Rótulo da versão: emr-5.26.0</p> <p>Distribuição do Hadoop: Amazon 2.8.5</p> <p>Aplicativos: Spark 2.4.3, Ganglia 3.7.2, Zeppelin 0.8.1, Hive 2.3.5, Tez 0.9.2, Livy 0.6.0, Mahout 0.13.0, Oozie 5.1.0, JupyterHub 0.9.6</p> <p>URI do log: s3://turing-bkt-treinamentos/Logs/Clusters/ </p> <p>Visualização consistente do EMRFS: Desativado</p> <p>ID personalizado de AMI: --</p>	<p>Zona de disponibilidade: us-east-1a</p> <p>ID da sub-rede: subnet-0a42009b008e416ea </p> <p>Principal: Running 1 c5.9xlarge Spot (max \$0.8/hr)</p> <p>Serviços: --</p> <p>Tarefa: --</p>

4) Configurando conexão com PuTTY (Windows)

Habilitar conexão da web

Configurar conexão da web

O Hadoop, o Ganglia e outros aplicativos publicam interfaces de usuário como sites hospedados no nó principal. Por razões de segurança, esses sites estão disponíveis apenas no servidor web local do nó principal.

Para se conectar às interfaces da web, você deve estabelecer um túnel SSH com o nó principal, usando o encaminhamento de portas dinâmicas ou locais. Se você estabelecer um túnel SSH usando o encaminhamento de portas dinâmicas, também deverá configurar um servidor de proxy para visualizar as interfaces da web.

Etapa 1: Abrir um túnel SSH para o nó principal do Amazon EMR - [Saiba mais](#)

Windows

Mac/Linux

1. Faça download de PuTTY.exe para seu computador de: <http://www.chiark.greenend.org.uk/~sgtatham/putty/download.html>
2. Inicie o PuTTY.
3. Na lista Categoria, clique em Sessão
4. No campo Nome de host, digite **hadoop@ec2-54-156-32-219.compute-1.amazonaws.com**
5. Na lista Categoria, expanda Conexão > SSH > Autenticar
6. Para autenticação no arquivo de chave privada, clique em Procurar e selecione o arquivo de chave privada (**TuringSSH_Treinamento.ppk**) usado para ativar o cluster.
7. Na lista Categoria, expanda Conexão > SSH e depois clique em Túneis.
8. No campo Porta de origem, digite **8157** (uma porta local não utilizada e escolhida aleatoriamente).
9. Selecione as opções Dinâmica e Automática.
10. Deixe o campo Destino vazio e clique em Adicionar.
11. Clique em Abrir.
12. Clique em Sim para descartar o alerta de segurança.

Etapa 2: Configurar uma ferramenta de gerenciamento de proxy - [Saiba mais](#)

Chrome

Firefox

1. Faça download da versão padrão de FoxyProxy e instale-a de: <http://foxyproxy.mozdev.org/downloads.html>
2. Reinicie o Chrome depois de instalar o FoxyProxy.

Fechar

Senha: labturing

5) Configurando conexão no Mac/Linux

Habilitar conexão da web

Configurar conexão da web

O Hadoop, o Ganglia e outros aplicativos publicam interfaces de usuário como sites hospedados no nó principal. Por razões de segurança, esses sites estão disponíveis apenas no servidor web local do nó principal.

Para se conectar às interfaces da web, você deve estabelecer um túnel SSH com o nó principal, usando o encaminhamento de portas dinâmicas ou locais. Se você estabelecer um túnel SSH usando o encaminhamento de portas dinâmicas, também deverá configurar um servidor de proxy para visualizar as interfaces da web.

Etapas:

Windows

Mac/Linux

1. Abra uma janela de terminal. No Mac OS X, escolha Applications > Utilities > Terminal. Em outras distribuições do Linux, o terminal é normalmente encontrado em Applications > Accessories > Terminal.
2. Digite o seguinte comando para estabelecer um túnel SSH com o nó principal usando o encaminhamento de portas dinâmicas. Substitua ~/TuringSSH_Treinamento.pem pelo local e nome de arquivo da chave privada (.pem) usados para ativar o cluster.

```
ssh -i ~/TuringSSH_Treinamento.pem -ND 8157 hadoop@ec2-54-156-32-219.compute-1.amazonaws.com
```
3. Digite yes para descartar o aviso de segurança.

Senha: labturing

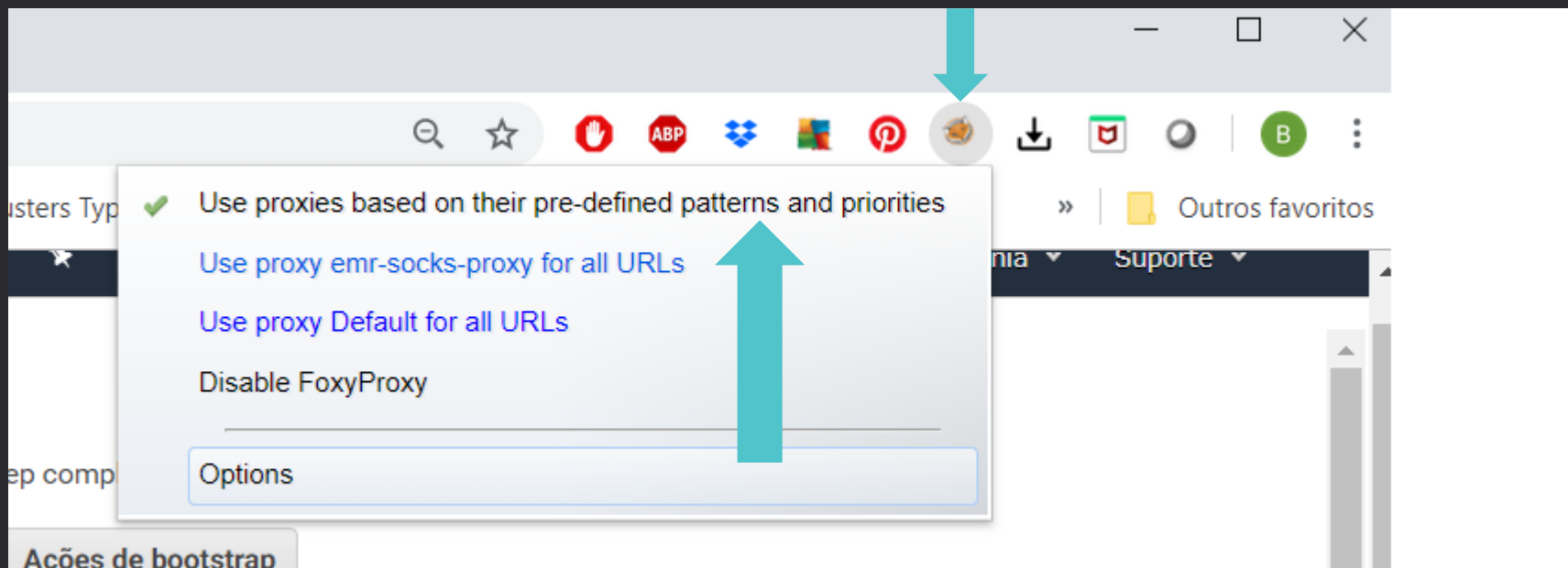
Após configurar, e conectar no cluster a tela de commando ficará bloqueada sem nenhum retorno, não se preocupe, está tudo correto.

Lembre-se que a chave .pem deve estar salva em diretório sem espaço no nome das pastas e após salvar execute o comando: `chmod 400 TuringSSH_Treinamento.pem`

6) Habilitando FoxyProxy no navegador Chrome

<https://chrome.google.com/webstore/detail/foxyproxy-standard/gcknhkkoolaabfmInjonogaaifnjlfnp?hl=pt-BR>

4. Clique no ícone FoxyProxy na barra de ferramentas e selecione Opções.
5. Clique em Importar/Exportar.
6. Clique em Escolher arquivo, selecione foxyproxy-settings.xml e clique em Abrir.
7. Na caixa de diálogo Importar configurações do FoxyProxy, clique em Adicionar.
8. Na página superior da página, para o Modo de proxy, escolha Usar Use proxies com base em padrões e prioridades predefinidos



Selecione a opção Use proxies based on their pre-defined patterns and priorities

7) Verifique se os aplicativos foram habilitados

Clonar Encerrar Exportação de CLI da AWS

Cluster: Treinamento - Turing Lab - PowerOfData **Aguardando** Cluster ready after last step completed.

Resumo Histórico do aplicativo Monitoramento Hardware Configurações Eventos Etapas Ações de bootstrap

Conexões: [Zeppelin](#), [Servidor de histórico do Spark](#), [Ganglia](#), [JupyterHub](#), [Gerenciador de recursos](#) ... (Visualizar tudo)

DNS público principal [ec2-54-156-32-219.compute-1.amazonaws.com](#) [SSH](#)

Tags: Treinamento = Turma02 [Visualizar todas/Editar](#)

8) Caso não tenha aparecido esta tela, clique no ícone de refresh e verifique novamente.