Consultas SQL Avançadas

**visão global**

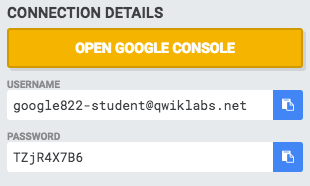
Um [conjunto de dados público do](https://cloud.google.com/bigquery/public-data/github) BigQuery tem informações sobre todas as confirmações do GitHub. Neste laboratório, você usará esses dados e métodos SQL avançados para responder à pergunta "Em que linguagens de programação os programadores de código aberto programam nos finais de semana?"

* Use campos aninhados, expressões regulares, com instrução e grupo e tendo
* Extrair informações de programação sobre commits de código

Para cada laboratório, você obtém um novo projeto de GCP e um conjunto de recursos por um tempo fixo, sem nenhum custo.

1. Certifique-se de ter entrado no Qwiklabs usando uma **janela anônima** .
2. Observe o tempo de acesso do laboratório (por exemplo, img / time.pnge certifique-se de que você pode terminar nesse bloco de tempo.

Não há recurso de pausa. Você pode reiniciar, se necessário, mas você deve começar no começo.

1. Quando estiver pronto, clique em img / start_lab.png.
2. Anote suas credenciais de laboratório. Você os usará para fazer login no Cloud Platform Console. 
3. Clique em **Abrir o Google Console** .
4. Clique em **Usar outra conta** e copie / cole credenciais para **este** laboratório nos prompts.

Se você usar outras credenciais, receberá erros ou **incorrerá em cobranças** .

1. Aceite os termos e pule a página de recursos de recuperação.

Não clique em **Terminar, a** menos que tenha terminado o laboratório ou queira reiniciá-lo. Isso limpa o seu trabalho e remove o projeto.

**Tarefa 1. Obter Informações sobre Confirmações de Código**

1. No console, no **menu Navegação** ( 7a91d354499ac9f1.png), clique em **BigQuery** .
2. Quando o console abrir, você vai alterar a interface clássica. Clique em **Ir para a interface clássica** .
3. O console do BigQuery será aberto em uma nova guia do navegador. Agora, selecione seu projeto Qwiklabs clicando na seta para baixo ao lado de **Recursos Qwiklabs** , selecionando **Alternar para projeto> Seu projeto Qwiklab** .
4. Clique no botão **Compor Consulta** no canto superior esquerdo.
5. Clique em **Mostrar opções** e verifique se você está usando o SQL padrão. Você está usando o SQL Padrão se a caixa de seleção **Usar Legacy SQL** estiver desmarcada (você está usando o SQL Padrão).
6. Clique em **Ocultar opções** .
7. Na janela **Nova Consulta** , digite (ou copie e cole) a seguinte consulta:

SELECT

author.email,

diff.new\_path AS path,

author.date

FROM

`bigquery-public-data.github\_repos.commits`,

UNNEST(difference) diff

WHERE

EXTRACT(YEAR

FROM

author.date)=2016

LIMIT 10

1. Jogue um pouco com a consulta acima para entender o que está fazendo. Por exemplo, em vez de author.email, tente apenas autor. Que tipo de campo é autor?
2. Mude diff.new\_pathpara difference.new\_path. Por que isso não funciona? Substitua difference.new\_path by difference[OFFSET(0)].new\_path. Isto funciona? Por quê? O que está UNNESTfazendo?

**Tarefa 2. Extrair linguagem de programação**

Nesta tarefa, você usará expressões regulares. Considere o caso em que uma extensão de nome de arquivo identifica a linguagem de programação. Portanto, um arquivo que termina em ".py" possui a linguagem de programação "py". Como você poderia usar uma expressão regular para extrair a extensão do caminho?

1. **Execute** a consulta a seguir. Lembre-se de desmarcar **Legacy SQL** .

SELECT

author.email,

LOWER(REGEXP\_EXTRACT(diff.new\_path, r'\.([^\./\(~\_ \- #]\*)$')) lang,

diff.new\_path AS path,

author.date

FROM

`bigquery-public-data.github\_repos.commits`,

UNNEST(difference) diff

WHERE

EXTRACT(YEAR

FROM

author.date)=2016

LIMIT

10

1. Modifique a consulta acima para usar somente lang se o idioma consistir exclusivamente de letras e tiver um tamanho inferior a 8 caracteres.
2. Modifique a consulta acima para agrupar por idioma e liste em ordem decrescente do número de confirmações.

Aqui está uma solução:

WITH

commits AS (

SELECT

author.email,

LOWER(REGEXP\_EXTRACT(diff.new\_path, r'\.([^\./\(~\_ \- #]\*)$')) lang,

diff.new\_path AS path,

author.date

FROM

`bigquery-public-data.github\_repos.commits`,

UNNEST(difference) diff

WHERE

EXTRACT(YEAR

FROM

author.date)=2016 )

SELECT

lang,

COUNT(path) AS numcommits

FROM

commits

WHERE

LENGTH(lang)<8

AND lang IS NOT NULL

AND REGEXP\_CONTAINS(lang, '[a-zA-Z]')

GROUP BY

lang

HAVING

numcommits > 100

ORDER BY

numcommits DESC

**Tarefa 3. Fim de semana ou dia da semana?**

Agrupe os commits com base em se eles aconteceram ou não em um final de semana. Como você faria?

1. Modifique a consulta anterior para extrair o dia da semana de author.date. Dias 2 a 6 são dias da semana.

Aqui está uma solução:

WITH

commits AS (

SELECT

author.email,

EXTRACT(DAYOFWEEK

FROM

author.date) BETWEEN 2

AND 6 is\_weekday,

LOWER(REGEXP\_EXTRACT(diff.new\_path, r'\.([^\./\(~\_ \- #]\*)$')) lang,

diff.new\_path AS path,

author.date

FROM

`bigquery-public-data.github\_repos.commits`,

UNNEST(difference) diff

WHERE

EXTRACT(YEAR

FROM

author.date)=2016)

SELECT

lang,

is\_weekday,

COUNT(path) AS numcommits

FROM

commits

WHERE

lang IS NOT NULL

GROUP BY

lang,

is\_weekday

HAVING

numcommits > 100

ORDER BY

numcommits DESC

Ignorando extensões de arquivos que não correspondem a linguagens de programação, parece que as linguagens de programação de final de semana mais populares são JavaScript, PHP e C (\* .h é o arquivo de cabeçalho C), Java e Python.

**Agradecimento** : Este laboratório é baseado em um artigo de Felipe Hoffa: [As principais linguagens de programação de final de semana de acordo com o código do GitHub](https://medium.com/@hoffa/the-top-weekend-languages-according-to-githubs-code-6022ea2e33e8#.8oj2rp804)

**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: 2018-09-17

© 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.