Olá! Este foi um projeto acadêmico com a finalidade de praticar conceitos sobre BI e Data Warehouse, neste projeto populei o banco de dados postgresSQL e o Redshift na nuvem da AWS, porém dei prioridade de apresentar os dados na nuvem. Vamos a uma breve explicação do Data Warehouse no postgres. No PostgresSQL eu criei dois schemas um para o modelo relacional e outro para o modelo dimensional:

Interface gráfica do usuário, Aplicativo

Descrição gerada automaticamente

No modelo relacional temos as tabelas de clientes, itensvendas, produtos, vendas e vendedores. No modelo dimensional temos as tabelas DimensaoClientes, DimensaoProduto, DimensaoTempo, DimensaoVendedores e fatovendas. Vamos supor que o modelo relacional já tenha os dados, e que já estamos com a modelagem star schema pronta, então é hora de popular as tabelas dimensões e a fato. O SQL é muito poderoso e podemos fazer essa importante tarefa através dele!

Para popular as dimensões eu usei um select, basicamente foi só alterar as tabelas :

Interface gráfica do usuário, Texto, Aplicativo, Email

Descrição gerada automaticamente

Hora de carregar a fato! Observação: Nas tabelas de dimensão e na fato temos duas colunas adicionais que são colunas para auxiliar a fato a ter um “histórico”, ou seja, vamos supor que todo mês temos uma nova carga de dados na fato se algum dado for alterado no modelo relacional, quando esse dado for para o modelo dimensional ele terá a data em que de inicio em que ocorreu a primeira carga e data de fimvalidade, por exemplo, a primeira carga foi no dia 01/06/2023 essa é a data de inicio se esse dado foi alterado no dia 20/06/2023 ele terá essa data como fimvalidade, vamos ver a primeira carga e depois a segunda carga com o dado alterado:

Interface gráfica do usuário, Texto, Aplicativo, Email

Descrição gerada automaticamente

Agora supondo um update no modelo relacional:

Texto

Descrição gerada automaticamente

Realizando um select nos dados alterados:

Interface gráfica do usuário, Aplicativo, Tabela

Descrição gerada automaticamente com confiança média

O que acontece é que o Data Warehouse armazena o momento em que um cliente deixou de ser um cliente silver e passou para gold. Utilizando o mesmo script da primeira imagem populando a tabela fato podemos continuar populando mês a mês basta mudar os números dos meses ou utilizamos no ultimo comando “Where date\_part('MONTH', V.DATA) > 01”.

Hora de subir as tabelas para nuvem da AWS! Importante destacar que eu importei as tabelas em formato csv, mas poderia ser em .sql já que vamos subir para o s3.

De inicio vamos criar um cluster do redshift, eu utilizei um cluster provisionado de apenas 1 nó, na região de São Paulo pois era o local que eu achei o free trial de um cluster provisionado, o tipo do nó era o dc2.large.

Para acessar é bem fácil, após a criação do cluster e o status estiver “Available”, basta clicar no “editor de consultas v2”.

Interface gráfica do usuário, Texto, Aplicativo

Descrição gerada automaticamente

Ao acessar ele abre o editor de query em uma nova aba, basta criar o database e as tabelas.

Tela de celular

Descrição gerada automaticamente

Criando as tabelas:

Tela de computador com texto preto sobre fundo branco

Descrição gerada automaticamente

Ao total foram 8 tabelas criadas, agora basta popular as tabelas! Vamos fazer isso subindo as tabelas para o bucket do s3!

Importante destacar que o nome do bucket do s3 ele tem que ser único no mundo todo! A criação é bem rápida e prática, e para carregar os dados basta sabe em qual pasta estão é bem rápido com poucos cliques os dados já estão na nuvem.

Interface gráfica do usuário, Aplicativo

Descrição gerada automaticamente

Não podemos nos esquecer de criarmos uma credencial para o nosso usuário sem as chaves da credencial não conseguimos puxar os dados do s3! Basta ir no IAM > Credenciais de segurança e rolar a página toda para baixo e ir em chaves de acesso e criar uma. Não se esqueça de fazer o download ou copiar para um arquivo local pois precisará desta chave, são criadas duas uma é secreta e ela é essencial, por isso guarde-a bem!

Chaves de acesso criadas vamos ao Redshift Importar os dados!

Tela de vídeo game

Descrição gerada automaticamente

No comando importamos os dados para as tabelas, basta indicar qual o bucket que vamos puxar os dados, as chaves de acesso(credenciais) e a região do bucket

Tela de computador com jogo

Descrição gerada automaticamente

Podemos inclusive criar uma tabela a partir de um select!

Tela de celular com aplicativo aberto

Descrição gerada automaticamente

Visualizando a nova tabela

Tela de computador com letras e números em fundo preto

Descrição gerada automaticamente

O Redshift é um excelente banco de dados para se armazenar um Data Warehouse, é escalável e bem fácil de mexer (para quem sabe SQL), em poucos minutos criamos uma tabela dos vendedores armazenando o total de vendas de 2022. Além de ser um banco extremamente rápido não podemos deixar de citar que podemos criar tabelas externas para ambientes de testes!