



Palavras-chave: Postgresql. Citus. Colunar. Economia recursos financeiros. Economia recursos computacionais

Introdução/Objetivo:

Diversos sistemas necessitam armazenar dados históricos para consultas, como em casos de auditoria ou diversos outros cenários. O banco de dados de histórico de acessos às redes WIFI (Eduroam e Unicamp-visitante) se encaixa neste cenário, sendo o maior backup que o CCUEC administra incluindo outras unidades da Unicamp. Os dados da rede WIFI, em virtude do Marco Civil da Internet, precisam ser armazenados por pelo menos um ano para efeito de auditoria. O tamanho total do banco de dados é de aproximadamente 670GB, com uma estimativa de carga diária de 2 milhões de registros referentes aos logs de acesso às redes WIFI.

Metodologia:

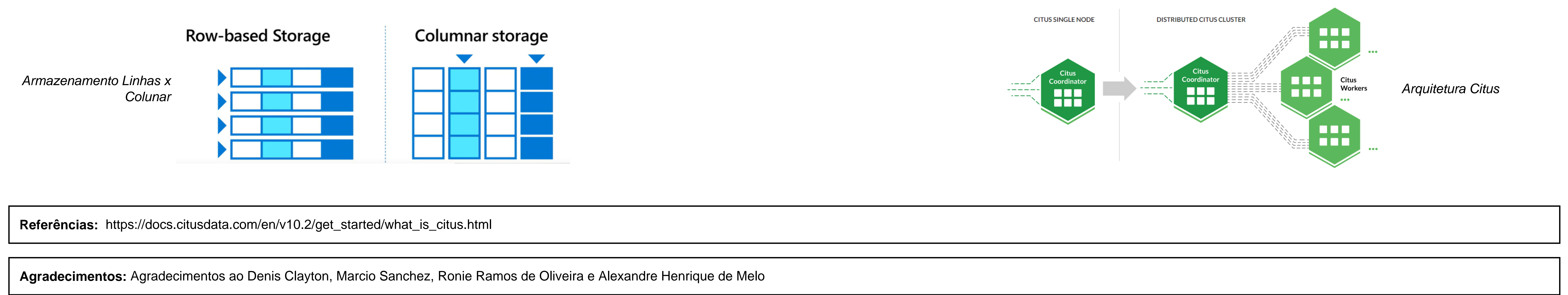
Foram realizados estudos com a extensão do PostgreSQL chamada Citus, sendo que com essa extensão é possível transformar tabelas específicas do banco de dados para tipo Colunar divididas por partições. Tabelas do tipo Colunar através do Citus, até o momento, não suportam deleções e atualizações em registros, somente inserções. O fato da limitação de deleções e atualizações não impactam tabelas do tipo histórico, pois elas somente sofrem inserções e se houver necessidade de remoção de dados, a mesma deverá ser feita removendo a partição específica onde os registros residem. Foram realizados testes de conversão para tabelas colunares em ambiente de homologação para medição do tempo necessário, sendo que, após os testes concluídos foi agendada a janela para produção.

Resultados:

Foi realizada a instalação do Citus no banco de dados onde residem as tabelas históricas, sendo que essas tabelas foram particionadas de forma Colunar por ano, ou seja, para cada ano os registros são armazenados em partições diferentes. Os resultados foram os abaixo: Redução do tamanho do banco de dados foi de aproximadamente 70%, passando de 670 GB para 205 GB; Tempo de backup diário do banco de dados passou de 10 horas para 3 horas; Tempo de execução de consultas ficaram cerca de 2 vezes mais rápidas; Grande economia de recursos financeiros por utilizar menor espaço em cartuchos de backup para retenção dos dados, exigindo assim uso reduzido dos cartuchos que é um recurso custoso. A utilização dos cartuchos é um recurso para tratar Disaster Recovery e aumentar o tempo de retenção dos dados que precisam ser armazenados; Os ganhos de performance e de espaço em disco para as tabelas utilizando a forma de acesso Colunar através da extensão Citus podem variar, levando em consideração o hardware e também a quantidade de índices presentes na tabela, sendo que o espaço em disco referentes à índices não sofreram redução.

Conclusão:

Os ganhos que a extensão Citus para PostgreSQL trazem para ambientes de dados históricos são claros. Além do ganho de espaço em disco e tempo de backup, temos ganhos em performance nas consultas que, segundo à documentação oficial, podem chegar até 10 vezes mais rápidas do que tabelas tradicionais. Este trabalho auxiliou a Diretoria de Operações do CCUEC a resolver problemas de concorrência na execução de rotinas de backup que afetavam diferentes unidades clientes deste serviço prestado pelo CCUEC.



Referências: https://docs.citusdata.com/en/v10.2/get_started/what_is_citus.html

Agradecimentos: Agradecimentos ao Denis Clayton, Marcio Sanchez, Ronie Ramos de Oliveira e Alexandre Henrique de Melo