



SEFIC2018
UNILASALLE

CIÊNCIA E TECNOLOGIA PARA A
REDUÇÃO DAS DESIGUALDADES

22 A 27
DE OUTUBRO

COMPARAÇÃO DE PERFORMANCE ENTRE MODELOS TABULAR E MULTIDIMENSIONAL

Guilherme Gonçalves Nascimento Silveira
Alexandre Giordani Andreoli
Universidade La Salle

Área Temática: Engenharias e Computação

Resumo: Ao observar os acontecimentos recentes na área do empreendedorismo é possível constatar o quão valiosos são os dados produzidos por uma organização. Novos modelos de negócio vêm surgindo tendo o mínimo de infraestrutura e custo mas com a proposta de conectar pessoas que querem determinados serviços com quem os provisiona e desta maneira lucrando de maneira exponencial e dominando o mercado global. A integração do time de Tecnologia da Informação com a área administrativa cria um novo modelo de negócio onde neste o objetivo é criar modelagens que saibam ler, interpretar, tratar e nos entregar os dados de forma que responda dúvidas do passado, analise o presente e prediga o futuro. Com muitas opções no mercado este presente artigo se atém à duas tecnologias para modelagem de dados: Multidimensional e Tabular do SQL Server Analysis Services (SSAS). Com o objetivo de averiguar se o modelo Tabular é viável para grandes projetos de inteligência de dados é criado uma estrutura em nuvem para o banco de dados e o serviço SSAS. Há medidas que podem ser feitas tanto em colunas de valores numéricos como em colunas de valores de texto. Para este trabalho foram realizadas medidas iguais em cada modelo para avaliar as diferenças de obter o mesmo resultado em diferentes contextos de dados. Para medidas de soma é produzido apenas o resultado de um somatório dos valores reais ou inteiros de uma coluna de dados numéricos e este resultado retorna um dado de tipo real ou inteiro. Para medidas de contagem distinta é feita uma contagem da quantidade de valores distintos em dada coluna, é uma medida muito utilizada para valores de chave primária em tabelas fato. Com estas medidas é possível traçar um comparativo entre como ambas tecnologias lidam com o tipo de dado que lhe é entregue. Com base nos resultados obtidos, o modelo Tabular é consideravelmente mais rápido que o cubo Multidimensional para conjuntos de dados de até 500 milhões de linhas quando temos um baixo nível de granularidade de dados. O modelo Tabular obteve variações mínimas para efetuar cálculos nas bases de dados de diferentes tamanhos em linhas e mesmo com variações no tamanho de até 64 colunas no conjunto de 10 milhões de linhas. Analisando o cubo Multidimensional verifica-se que ele chega a valores superiores a 10 mil vezes em relação ao Tabular para calcular uma medida, passando até mesmo de um minuto conforme o conjunto de dados aumenta. Na avaliação de performance pelo tamanho de colunas do conjunto é registrado uma possível interferência significativa da conexão com a rede ao realizar uma pergunta ao modelo.

Palavras-Chave: Modelagem de dados, Tabular, Multidimensional