

Escalonador Dinâmico para Frameworks de Processamento de Stream para Big Data em Ambientes Multitenant

Doutorado

Cleverton Vicentini, Altair Olivo Santin

Contexto: A computação em nuvem é caracterizada por sua alta escalabilidade, sendo os recursos computacionais providos de acordo com a necessidade do cliente. A computação em nuvem é um ambiente intrinsecamente multi-locatário (*multi-tenant*), que compartilha os recursos físicos (processador, rede, disco e memória) entre os clientes, geralmente usando a virtualização do *hardware*. Ferramentas de processamento para *streaming* de dados fazem o uso destes ambientes para ganhar escalabilidade e elasticidade, devido a demanda de processamento de dados em larga escala e tempo real. *Multi-tenant* tem como ponto negativo a interferência que um cliente pode sofrer devido ao compartilhamento do mesmo *hardware* físico. Isso significa que um cliente executando em *hardware* virtualizado pode obter dados incorretos a respeito do estado físico desse *hardware*, devido ao uso excessivo de outro cliente sobre os recursos físicos compartilhados. Desta forma, para obter o estado real do *hardware* virtualizado, torna-se necessário um mecanismo que promova um diagnóstico mais preciso do estado físico dos recursos do *hardware* virtualizado. Porém, atualmente, esta informação não está disponível em nenhuma plataforma de computação em nuvem. **Objetivo Geral:** Desenvolver um escalonamento de demandas computacionais para processamento de *streaming* de dados, de tempo real e baseado em computação em nuvem. **Questão de Pesquisa:** Aprimorar o processo de escalonamento no âmbito de *frameworks* de processamento de *Stream* para Big Data, fornecendo poder de decisão para o escalonador proposto. Esse deve avaliar o estado físico de cada instância de *hardware* virtualizado e direcionar o processamento para onde há baixas taxas de utilização. Isso permitirá realizar a distribuição de tarefas baseada nos dados de utilização, mais próximo do estado real do *hardware* físico. A coleta das informações de utilização dos recursos de *hardware* ocorrerá em intervalos regulares. **Método de Pesquisa:** Este trabalho propõe um escalonador dinâmico de tarefas para *frameworks* de processamento de *streaming*, chamado de *Dynamic Scheduler* (DyS). Por se tratar de processamento em tempo real, o escalonador avalia o estado do *hardware* virtualizado usando ferramentas de *benchmark*, que servem de referencial para obter o estado real do *hardware* virtualizado. Para prova de conceito, será utilizado o *framework Apache Storm*, em um ambiente de nuvem computacional, operando em modo *multi-tenant*. **Resultados Preliminares:** O resultado esperado para esse ambiente é que as tarefas executadas sobre a plataforma *Storm*, usando o DyS,

apresentem um desempenho geral superior em comparação com o *Default Scheduler* do *Storm*, que usa o algoritmo de *Round-Robin*.

Palavras-chave: Big Data, Multi-tenant cloud computing, Dynamic Scheduler.