Escalonamento Adaptativo para o Apache Hadoop

Guilherme Weigert Cassales[1] **e Andrea Schwertner Charão**[1]

[1]Laboratório de Sistemas de Computação
Universidade Federal de Santa Maria
Santa Maria – RS – Brazil

{cassales,andrea}@inf.ufsm.br

Resumo. *O Apache Hadoop, distribuição mais conhecida de MapReduce, requer uma etapa de configuração com interação humana. A qual, num ambiente pervasivo, deve ser reduzida ao máximo. Uma técnica que possibilita a redução é a sensibilidade ao contexto. Nesse âmbito esse trabalho propõe melhorias no processo de escalonamento de tarefas, mediante aquisição de informações sobre os recursos disponíveis com objetivo de tornar o ambiente adaptável.*

1. Introdução

A quantidade de dados gerados atualmente excede o poder de processamento da maioria dos computadores quando o resultado do processamento possui um prazo pequeno para ser apresentado. Uma das alternativas para este problema é a utilização de *clusters* e *frameworks* especializados para determinados tipos de ambiente – como ambientes distribuídos ou de arquitetura *multicore* – e atividade desejados.

Um exemplo de *framework* especializado para processamento de grande volume de dados em ambiente distribuído é o Apache Hadoop, que utiliza o paradigma de MapReduce e foi projetado para utilização em *clusters* homogêneos. Esta decisão de projeto tem sérias repercussões, pois sabe-se que a performance de um *framework* está intimamente ligada com o escalonador de tarefas [Kumar et al. 2012]. Logo, quando submetido a um ambiente diferente do previsto o desempenho tende a degradar.

Quando o ambiente diverge da suposição inicial – no caso de alguns nós possuírem mais/menos recursos do que aqueles descritos no arquivo *default* – o sistema acaba desperdiçando capacidade de processamento ou sobrecarregando os recursos, fazendo uma má utilização da capacidade disponível em ambas situações. Este trabalho propõe uma melhoria no processo de escalonamento de tarefas através de uma abordagem pervasiva sensível ao contexto para o Apache Hadoop.

2. Proposta

Por padrão o Hadoop é configurado por meio de propriedades descritas em arquivos XML, contendo uma configuração *default*. Para que o usuário modifique estas propriedades, é necessária uma edição de um arquivo específico para cada nó do *cluster*, caso esse arquivo não seja criado/editado as informações do arquivo *default* serão utilizadas.

Além da possível utilização de configurações *default*, as informações passadas ao escalonador no momento da inicialização são alteradas somente com a utilização dos recursos pelo próprio Hadoop. Se algum outro processo for iniciado em um nó, os recursos utilizados por ele estarão constando como disponíveis para o escalonador do Hadoop.

Nota-se que os dados utilizados são estáticos e que a alteração requer grande quantidade de interação humana.

Este trabalho propõe uma mudança nas políticas de escalonamento de tarefas quanto a avaliação da performance e um conjunto de modificações que possibilitam a coleta de informações sobre os recursos disponíveis (memória, cores, carga de CPU) em determinado momento e a transmissão de dados atualizados para o escalonador. Sabe-se que uma nova tarefa é lançada caso a performance de outra esteja inferior à média do cluster. Para esta avaliação é necessário possuir informações sobre o nó que a tarefa está sendo executada para que seja possível aplicar a heurística.

3. Resultados Preliminares e Esperados

A solução já foi parcialmente implementada. A implementação corresponde à coleta de informações sobre os recursos em cada nó e subsequente transmissão destas informações para o escalonador no momento de inicialização. Testes preliminares (execuções de TeraSort com conjunto de dados de 5GB) com estas alterações demonstram que o tempo de execução melhorou em até 47% quando comparado ao tempo de execução utilizando as configurações *default* num mesmo ambiente [Cassales et al. 2014].

Espera-se que após concluídas todas as modificações, o *framework* seja capaz de reagir a mudanças na disponibilidade de recursos em tempo real. O escalonador deverá receber mensagens periódicas de todos os nós – contendo dados atualizados sobre a situação dos recursos – e redistribuir a carga para que os nenhum nó fique sobrecarregado ou subutilizado. Para que isso seja possível, é necessário criar um novo canal de comunicação entre os nós, o qual optou-se por utilizar a API ZooKeeper, outro projeto da Apache que fornece um ambiente seguro e robusto de escrita e leitura distribuída [Zookeeper 2014].

Se corretamente implementada, esta técnica proporcionará uma melhor adaptação em ambientes onde a configuração dos nós pode variar bastante de um nó para outro, bem como em casos onde os recursos são compartilhados com outras aplicações e podem, no decorrer de uma tarefa, sofrer uma diminuição na quantidade de recursos disponíveis. Para avaliação do desempenho utiliza-se *Gantt Charts* consolidados por nó, contendo informações do número de processos em execução a cada momento. Além do tempo geral de execução, serão avaliados: a utilização dos recursos e a capacidade de adaptação do *framework* aos diferentes ambientes aos quais for submetido.

Referências

Cassales, G., Charao, A., Kirsch-Pinheiro, M., Souveyet, C., and Steffenel, L. (2014). Bringing context to apache hadoop. In *Proceedings of The Eighth International Conference on Mobile Ubiquitous Computing, Systems, Services and Technologies*, IARIA '14, pages 52–61.

Kumar, K. A., Konishetty, V. K., Voruganti, K., and Rao, G. V. P. (2012). Cash: context aware scheduler for hadoop. In *Proceedings of the International Conference on Advances in Computing, Communications and Informatics*, ICACCI '12, pages 52–61, New York, NY, USA. ACM.

Zookeeper, A. (2014). Zookeeper. `http://zookeeper.apache.org`, Accessado em Dezembro 2014.