Uma Arquitectura para Múltiplas Tabelas de Hash Distribuídas em Ambiente Cluster

José Rufino
Instituto Politécnico de Bragança
rufino@ipb.pt

António Pina
Universidade do Minho
pina@di.uminho.pt

Quando as aplicações necessitam de lidar com grandes volumes de dados, para lá da capacidade de armazenamento do sistema computacional hospedeiro, deixa de ser possível a utilização de estruturas de dados convencionais (centralizadas), sendo necessário o recurso a *Estruturas de Dados Distribuídas* (DDSs), realizadas através da cooperação de vários nós computacionais.

Idealmente, a realização de uma DDS deveria ser feita no quadro de uma política global de rentabilização dos recursos dos nós computacionais de suporte, que leva em linha de conta a sua eventual (ou melhor, inevitável) *heterogeneidade*.

Essa heterogeneidade decorre, em primeira instância, de perfis de *hardware* eventualmente diferentes – *heterogeneidade física*. Dependendo do ambiente distribuído, essa heterogeneidade pode ser em maior ou menor grau (*e.g.*, ambientes *Peer-to-Peer* (P2P) versus ambientes *cluster*), mas é relativamente *estática*, pois as configurações de hardware dos nós, que definem as suas *capacidades*, mudam pouco e lentamente.

Outra fonte de heterogeneidade é a execução concorrente, em cada nó, de múltiplas aplicações com requisitos específicos e dinâmicos, originando flutuações na *disponibilidade* dos seus recursos. Esta heterogeneidade, que classificamos de *heterogeneidade lógica* (porque originada ao nível do *software*) é inerentemente *dinâmica*, levantando por isso questões relativamente complexas.

Para além do tratamento da heterogeneidade do ambiente de execução, é também desejável suportar a co-existência de múltiplas instâncias de DDSs, eventualmente partilhadas por diversas aplicações. Assim, são necessários serviços capazes de gerir, de forma integrada, a participação simultânea de um nó computacional em diversas DDSs ao mesmo tempo que, em colaboração uns com os outros, asseguram uma utilização balanceada do conjunto de recursos que as sustentam.

Certas classes de aplicações estão particularmente interessadas em *dicionários distribuídos*, repositórios distribuídos em que o acesso a cada registo se faz por uma chave única. Neste contexto, Árvores e Tabelas de Hash distribuídas são, provavelmente, os paradigmas mais populares.

Nesta comunicação descrevem-se os aspectos gerais da proposta de uma arquitectura que, no âmbito de um ambiente de computação paralela e distribuída do tipo cluster, pretende suportar a realização, exploração e gestão de múltiplas *Tabelas de Hash Distibuídas* DHTs, levando em conta a heterogeneidade dos nós computacionais que as sustentam.