UDESC UNIVERSIDADE DO ESTADO DE SANTA CATARINA
Seminário de Iniciação Científica
Universidade do Estado de Santa Catarina
29º SIC UDESC

ANÁLISE DO COMPORTAMENTO DA TAXA DE CHEGADA DE SUBMISSÕES DE TAREFAS NOS ANOS 2015 A 2018 NA GRID5000

Gustavo Bandeira da Silva[1], Maurício Aronne Pillon[2]
[1] Acadêmico do Curso de Bacharelado em Ciência da Computação - BCC bolsista PROBIC/UDESC
[2] Orientador, Departamento de Ciência da Computação – DCC – mauricio.pillon@udesc.br

Palavras-chave: Escalonamento de tarefas, Grades Computacionais e *GRID5000.*

Grades computacionais são plataformas distribuídas de larga escala constituídas pela agregação de recursos computacionais de instituições geograficamente distribuídas. Recursos são disponibilizados de forma dinâmica, flexível e modelados em tarefas. Em uma tarefa, o usuário da grade especifica as necessidades de recursos computacionais (e.g.: número de processadores, quantidade de memória), tempo estimado de uso dos recursos, a aplicação que deseja executar, entre outras. Assim que a tarefa é executada, o usuário passa a ser proprietário da totalidade de recursos computacionais especificados no momento da submissão pelo período de tempo requisitado. O escalonador de recursos, associado a um Sistema de Gerenciamento de Recursos (RMS), determina quando e onde a tarefa executará. Algoritmos de escalonamento sofrem com a imprevisibilidade das requisições (e.g.: taxa de chegada, quantidade de recursos, tempo estimado de uso).

Caracterizar o comportamento de tarefas e de usuários de grades computacionais permite a adequação de algoritmos de escalonamento, otimizando-se a utilização dos recursos ou minimizando-se o tempo de execução das aplicações. A proposta deste trabalho é a análise de comportamento das submissões de tarefas de uma grade computacional privada dedicada a pesquisa, a *Grid5000*[1]. Em funcionamento desde meados dos anos *2000*, a *Grid5000* conta atualmente com 12.000 núcleos de processamentos distribuídos em 8 sítios principais, localizados na França. Com dados reais, obtidos diretamente do OAR[2], a análise deste trabalho restringiu-se as submissões dos últimos quatro anos (2015 a 2018), de um único agregado, o Taurus, situado no sítio de Lyon. A métrica em foco foi a distribuição de chegada das tarefas.

Observando-se a distribuição de chegada das tarefas espera-se estabelecer períodos diários de comportamentos que permitam a adequação de algoritmos de escalonamentos da *Grid5000*. Os resultados podem ser observados na Figura 1, a qual tem em seu eixo *x*, o período de 24h e, no eixo *y*, o total do número de tarefas submetidas por hora, nos anos analisados. Os gráficos estão todos na mesma escala, permitindo concluir que o ano de 2016 teve a menor taxa de submissão de tarefas de todos os anos analisados. Todavia, o comportamento, em função da hora do dia, se repete ao longo dos anos. É possível observar que a maior taxa de submissão é entre 9h e 11h e, as menores, entre 3h e 6h. Os destaques individuais das submissões ao longo dos anos são: (i) Em 2015, o pico de tarefas ocorreu às 10h, totalizando 5436, e o menor valor as 5h, com 819; (ii) Em 2016, o pico se repete, mas o total de tarefas baixou em 43%, atingindo 3539, e o menor valor as 4h, com 599; (iii) Em 2017, os valores dispararam novamente, superando 2015 em 54%, porém neste ano o horário de maior taxa de submissão foi às 9h, com 5469, e o menor valor, 6h, com 2448; e (iv) Em 2018, o pico voltou a ser às 10h e o total chegou a 4620, e o menor valor 6h, com 1445.

---

1 Disponível em https://www.grid5000.fr/ .
2 Disponível em http://oar.imag.fr/ .

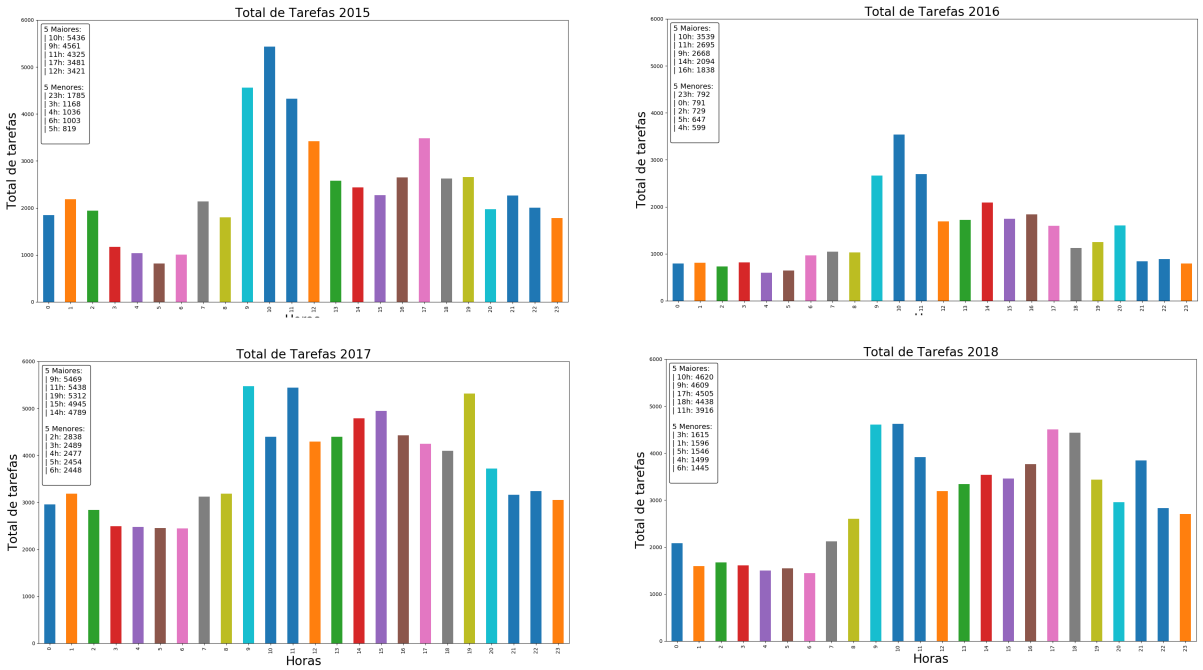

Figura 1: *Total de submissões ao agregado Taurus/Lyon (Grid5000) entre os anos 2015 e 2018.*

O gráfico ainda destaca os dez resultados extremos, os cinco maiores e os cinco menores, buscando auxiliar na definição de períodos. Embora os gráficos apresentados compreendam os totais dos anos, ao longo da pesquisa, fez-se a análise diária e mensal dos períodos, que se assemelham ao comportamento das totalidades. Na lista dos cinco maiores, os horários da manhã têm maior frequência, aparecendo em todos os anos na lista (9h, 10h e 11h), com algumas exceções que mostram horários do meio da tarde (14h, 15h ou 16h). No outro extremo, a tendência é clara, destacando horários entre 2h e 6h da manhã. Este comportamento pode ser explicado pela jornada de trabalho dos pesquisadores europeus, que iniciam seus trabalhos próximo das 8h, fazem um intervalo para o almoço, e deixam seus postos no final da tarde.

Outro comportamento cíclico anual observado é o aumento na taxa de chegada de tarefas as 19h. Este comportamento pode ser explicado por outro fato, a política de uso da *Grid5000*[3]. O tempo de uso dos recursos no período "comercial", determinado entre 7h as 19h, é contabilizado e restrito ao uso total máximo de cada agregado de 2h por usuário. Embora a restrição ainda represente muito poder computacional, usuários costumam priorizar a submissão fora do horário comercial como política de "boa vizinhança". Com este trabalho foi possível definir períodos de alta (Pa1: 9h as 12h, Pa2: 14h às 17h e Pa3: 19h) e baixa (Pb1: 0h às 8h) utilizações. Os algoritmos de escalonamento podem se beneficiar desta caracterização de uso em períodos para aplicar políticas diferenciadas mais adequadas ao comportamento.

Referências:

LELONG, J.; REIS, V.; TRYSTRAM, D. Tuning EASY-Backfilling queues. In: SPRINGER. Workshop on Job Scheduling Strategies for Parallel Processing. [S.l.], 2017. p. 43–61.

FOSTER, I.; KESSELMAN, C. "The Grid 2: Blueprint for a New Computing Infrastructure". San Francisco, CA, USA: Morgan Kaufmann Publishers Inc., 2003. ISBN 1558609334.

3 Disponível em https://www.grid5000.fr/w/Grid5000:UsagePolicy.