

Proposta para Dissertação de Mestrado em Engenharia Informática

Título:

Optimização de base de dados distribuída: o caso do EPOS

Orientador:

Rui Fernandes

Co-Orientadores:

Rui Cardoso

João Muranho

Contexto:

Esta proposta de trabalho de investigação e desenvolvimento corresponde a uma componente a desenvolver no projeto de investigação europeu EPOS (European Plate Observing System) a decorrer. O projeto EPOS, que envolve várias instituições europeias do ensino superior e de investigação. O trabalho de investigação e desenvolvimento a realizar integra-se neste projeto internacional, contribuindo para melhorar o desempenho da base de dados que suporta os dados e metadados disponibilizados aos utilizadores.

Sumário:

Os Sistemas de Gestão de base de Dados (SGBD) existem já há bastante tempo, sendo apenas em 1986 que a Structured query language (SQL) se tornou-se um padrão, ISO / IEC. A optimização do desempenho de base de dados tem tornado-se cada vez mais relevante à medida que os SGBD suportam cada vez maior quantidade de informação, torna-se necessário cada vez mais efectuar a sua optimização [1] para assegurar um bom desempenho. A tarefa de optimização envolve várias vertentes, desde reescrever as consultas para as mais eficientes. Mas todo o processo de optimização de base de dados envolve também outras áreas [2][3][4][5], como a indexação, o particionamento de tabelas, o planeamento de consultas, a configuração do servidor, a adaptação do hardware às necessidades do sistema nomeadamente através do aumento da memória ou da largura de banda disponível, mas também pode implicar mudar completamente de SGBD para outro diferente por exemplo optar por base de dados NoSQL (cada vez mais utilizadas em cenários em que existe um grande volume de informação a considerar e ou quando se destinam a ser usada em portais que disponibilizam dados em tempo real. Este trabalho de investigação tem como principal objectivo desenvolver um conjunto de recomendações para optimizar base de dados distribuídas utilizando como caso de estudo e de validação a base de dados do EPOS, uma base de dados distribuída em constante crescimento e para a qual é necessário assegurar e ou melhorar o desempenho em face às solicitações dos seus utilizadores, fazendo uso de uma perspectiva holística em que todas as partes do SGBD são susceptíveis de serem relevantes para melhorar o desempenho.

Objetivos

O principal objetivo desta dissertação consiste na especificação e implementação de um conjunto de tarefas de modo a conseguir otimizar uma base de dados utilizando como caso de estudo a base de dados EPOS.

Tarefas a Realizar

No âmbito do trabalho de investigação e de desenvolvimento a realizar elaboração da dissertação de mestrado:, são propostas as seguintes tarefas a efectuar:

- Tarefa 1. Estudo e familiarização com os principais conceitos subjacentes à optimização de base de dados distribuídas.
- Tarefa 2. Estudo e comparação de soluções de implementação de optimização existentes (por exemplo, baseadas em virtualização nativa de hardware e em infraestruturas virtualizadas de elevada disponibilidade, baseadas em soluções SQL e ou em NoSQL).
- Tarefa 3. Especificação e implementação de uma plataforma de testes usando como “*proof of concept*” a base de Dados do EPOS em desenvolvimento no C4G/SEGAL/UBI.
- Tarefa 4. Execução de testes e validação da base de dados em várias plataformas.
- Tarefa 5. Elaboração de um estudo sobre o processamento de optimização baseado no caso de estudo experimental baseado na base de dados EPOS.
- Tarefa 6. Escrita de um artigo científico sobre o trabalho de investigação realizado.
- Tarefa 7. Escrita da dissertação de mestrado.

Cronograma

A lista seguinte representa a calendarização prevista para a execução das várias tarefas ao longo de 10 meses:

- Tarefa 1. Set, Out.
- Tarefa 2. Out, Nov, Dez.
- Tarefa 3. Nov, Dez, Jan, Fev, Mar.
- Tarefa 4. Abr, Mai.
- Tarefa 5. Mai, Jun.
- Tarefa 6. Jun.
- Tarefa 7. Mai, Jun.

Escrita da Dissertação em Língua Inglesa

A dissertação de mestrado resultante da realização do plano de trabalho proposto deverá ser escrita em língua inglesa, tendo em vista a divulgação internacional do trabalho científico desenvolvido. O título da dissertação em língua inglesa deverá ser o seguinte: “Optimizing distributed databases performance: The EPOS Case”.

Referências

- [1] M. Jarke and J. Koch, “Query optimization in database systems,” ACM Computing surveys (CSUR), vol. 16, no. 2, pp. 111-152, 1984.
- [2] C. Vicknair, M. Macias, Z. Zhao, X. Nan, Y. Chen, and D. Wilkins, “A comparison

of a graph database and a relational database: a data provenance perspective,” in Proceedings of the 48th annual Southeast regional conference, p. 42, ACM, 2010.

[3] Y.-S. Cui, Y. Zhang, C. Zeng, J.-H. Feng, and C.-X. Xing, “Database physical structure optimization technology,” Ruanjian Xuebao/Journal of Software, vol. 24, no. 4, pp. 761-780, 2013.

[4] S. Chaudhuri, A. K. Gupta, and V. Narasayya, “Compressing sql workloads,” in Proceedings of the 2002 ACM SIGMOD international conference on Management of data, pp. 488-499, ACM, 2002.

[5] S. Agrawal, E. Chu, and V. Narasayya, “Automatic physical design tuning: workload as a sequence,” in Proceedings of the 2006 ACM SIGMOD international conference on Management of data, pp. 683-694, ACM, 2006.