

**Projeto em ciência de
dados com soluções
para processamento
paralelo e distribuído
de dados**



Monitoramento e depuração de programas paralelos. Sistemas de armazenamento de dados distribuídos


Bloco 1

Marcelo Tavares de Lima





► Objetivos

- Apresentar conceitos e definições associados a monitoramento e depuração de programas paralelos.
 - Apresentar conceitos fundamentais de bancos de dados.
 - Apresentar conceitos fundamentais de armazenamento de dados distribuídos.
- 



► Conceitos básicos de bancos de dados

- Ricarte (1996, p. 1) afirma que “um banco de dados é uma coleção de dados relacionados que pode ser armazenada sob alguma forma física”.




► Conceitos básicos de bancos de dados

- Associado a essa definição, o autor afirma ainda que “os dados armazenados em um banco de dados representam algum aspecto específico do mundo real”. (RICARTE, 1996, p. 1)




► Depuração de sistemas paralelos

- Surgiu com a intenção de redução do tempo de processamento de dados.
 - Surgiu na década de 90.
 - Quanto menor o tempo de resolução dos problemas, mais rapidamente se toma decisões importantes para os negócios.
 - O processamento também precisa ser confiável.
- 



► Depuração de sistemas paralelos

- Na prática, é o uso de mais de uma unidade de processamento (CPU) para a execução da resolução conjunta de um problema.
 - Divide o problema em problemas (tarefas) menores.
 - Exige investimento em hardware e software.
 - Processamento de alto desempenho.
- 



► Depuração de sistemas paralelos

- A principal motivação para o surgimento de sistemas paralelos, segundo Ricarte (1996), foi a execução de programas científicos, com uma grande capacidade de processamento numérico.



► Depuração de sistemas paralelos

Segundo Schnorr (2014):

- O objetivo da análise de desempenho de aplicações paralelas é a identificação de regiões do programa que tem uma baixa exploração dos recursos computacionais. (SCHNORR, 2014, p. 1)




► Depuração de sistemas paralelos

- É usual partir para a análise de desempenho de programas paralelos, após sua implementação livre de qualquer possível erro, momento em que o desenvolvedor ou administrador do banco de dados está apto para realizar esta etapa do trabalho.



► Depuração de sistemas paralelos

Duas fases para uma análise de desempenho:

- Coleta de observações e registros.
 - Análise de dados.
- 



► Armazenamento de dados distribuídos

Segundo Varajão (2016):

- Os sistemas de informação distribuídos (SID) ficaram mais populares depois da explosão da Internet em 1993 e, desde então, estes sistemas não param de crescer. (VARAJÃO, 2016, p. 12)



► Armazenamento de dados distribuídos

Propriedades de um sistema de armazenamento de dados distribuídos:

1. Consistência.
2. Disponibilidade.
3. Tolerância à partição.

Monitoramento e depuração de programas paralelos. Sistemas de armazenamento de dados distribuídos


Bloco 2

Marcelo Tavares de Lima





► Depuração de sistemas paralelos

- Fase de coleta: ocorre ao longo da execução da aplicação e tem como objetivo o registro de informações importantes sobre o comportamento do sistema.
- 



► Depuração de sistemas paralelos

- Fase de análise: tem como objetivo a identificação de problemas de desempenho durante a execução da aplicação, assim como a identificação das causas desses possíveis problemas. Dentre as técnicas utilizadas para análise dos dados de desempenho, podemos citar as técnicas de visualização de rastros e a análise estatística.



► Depuração de sistemas paralelos

- Abordagem on-line.
- Abordagem off-line.



► Armazenamento de dados distribuídos

- Teorema CAP (Consistência, Disponibilidade e Tolerância a Partição – *Consistency, Availability, Partition tolerance*): apenas duas propriedades ocorrem de forma simultânea.

► Armazenamento de dados distribuídos

Figura 1 – Teorema CAP



Fonte: Almeida e Dutra (2019).

PÓS-GRADUAÇÃO

Teoria em prática


Bloco 3

Marcelo Tavares de Lima





► Teoria em prática

- Suponha que você faça parte de uma equipe de Tecnologia da Informação (TI) de uma grande empresa, com várias filiais espalhadas por todo Brasil. No entanto, na unidade em que você trabalha, estão centralizados os trabalhos de construção de sistemas de dados e de acompanhamento do desempenho dos mesmos. Portanto, você e seus colegas são responsáveis por essa importante atividade de trabalho, essencial para a troca de informações e dados da empresa.
- 



► Teoria em prática

- Existe intenção de melhoria na rede de distribuição de dados para torná-la mais eficiente, segura e mais rápida na troca de informações, na consulta aos dados e na produção de relatórios de dados que possam auxiliar na tomada de decisões, tanto internas quanto externas. No entanto, para isso, é preciso realizar uma análise de desempenho do sistema, atualmente instalado para buscar os pontos de deficiência e de fragilidade que, potencialmente, possam existir. Tal tarefa ajudará a convencer seus superiores a investirem na ideia de sua equipe.

Dica do professor

Bloco 4

Marcelo Tavares de Lima



► Indicação de artigo

Figura 2 – Artigo

Armazéns de Dados em MongoDB vs SQL Server

Uma análise comparativa da performance de consultas

Data Warehouses in MongoDB vs SQL Server

A comparative analysis of the query performance

Daniel Pereira

Instituto Superior de Engenharia do Porto – IPP
Rua Dr. António Bernardino de Almeida, 431
4249-015 Porto, Portugal
1121279@isep.ipp.pt

Paulo Oliveira e Fátima Rodrigues

Instituto Superior de Engenharia do Porto – IPP e GECAD
Rua Dr. António Bernardino de Almeida, 431
4249-015 Porto, Portugal
{PJO,MFC}@isep.ipp.pt

Fonte: Pereira (2015).



► Referências

ALMEIDA, D. S.; DUTRA, D. B. Armazenamento de big data: uma abordagem didática big data storage – a didactic survey. **Perspectivas em Ciências Tecnológicas**, v. 8, n. 8, p. 168-194, 2019. Disponível em: [http://fatece.edu.br/arquivos/arquivos%20revistas/perspectiva/volume8/Days e%20Silveira%20de%20Almeida;%20Danilo%20Borges%20Dutra.pdf](http://fatece.edu.br/arquivos/arquivos%20revistas/perspectiva/volume8/Days%20Silveira%20de%20Almeida;%20Danilo%20Borges%20Dutra.pdf). Acesso em: 05 fev. 2020.

PEREIRA, D.; OLIVEIRA, P.; RODRIGUES, F. Armazéns de dados em MongoDB vs SQL Server. **CISTI (Iberian Conference on Information Systems & Technologies/ Conferência Ibérica de Sistemas e Tecnologias de Informação) Proceedings**. v. 1, p. 610-616, 2015.

► Referências

RICARTE, I. L. M. **Sistemas paralelos de bancos de dados**: arquiteturas e algoritmos. Campinas: Faculdade de engenharia elétrica e de computação, 1996. Dissertação Disponível em: <ftp://ftp.dca.fee.unicamp.br/pub/docs/ricarte/apostilas/spbdaa.pdf>. Acesso em: 05 fev. 2020.

SCHNORR, L. C. **Análise de desempenho de programas paralelos**. Porto Alegre: Instituto de Informática. Universidade Federal do Rio Grande do Sul. 2014. Disponível em: <http://www.inf.ufrgs.br/~schnorr/download/talks/erad2014-minicurso-texto.pdf>. Acesso em: 03 fev. 2020.

VARAJÃO, F. **Sistemas distribuídos**. Apostila do curso de bacharelado em sistema de informação da FEUC, 2016. Disponível em: <https://varajao.com.br/disciplinas/SD/SD%20-%20Apostila%20-%20SISTEMAS%20DISTRIBUIDOS.pdf>. Acesso em: 03 fev. 2020.

