

# Implantação de um banco de dados não relacional e distribuído em um cluster de SoC's

Lucas Ferreira da Silva<sup>1</sup>, João V. F. Lima<sup>1</sup>

<sup>1</sup> Programa de Pós-Graduação em Ciência da Computação (PPGCC)  
Universidade Federal de Santa Maria  
Santa Maria – RS – Brasil

{lferreira, jvlima}@inf.ufsm.br

**Resumo.** *Este trabalho apresenta um estudo sobre a utilização de um cluster de SoC's (System-on-a-chip) enquanto alternativa para a implantação do banco de dados não relacional e distribuído Cassandra. O cluster de SoC's é composto por quatro Raspberry Pis e gerenciado por meio do Docker Swarm. Os resultados preliminares demonstram que mesmo em um ambiente não comumente utilizado para essa finalidade e com uso de recursos limitados, o banco de dados operou com estabilidade.*

## 1. Introdução

O crescente aumento do volume de dados, provenientes de diversas fontes e dispositivos, tem gerado uma grande demanda por ferramentas e tecnologias cada vez mais poderosas para o armazenamento e processamento de tais dados. Assim, abordagens como o uso de bancos de dados não relacionais e distribuídos tem se mostrado promissoras, já que características como escalabilidade horizontal, alta disponibilidade e estruturas de armazenamento otimizadas, trazem vantagens que se adaptam bem a esse novo paradigma. Entretanto, com o aumento do volume de dados, conseqüentemente há o aumento da demanda de capacidade computacional dos *datacenters*, que passam a exigir estruturas de hardware também mais robustas, ocasionando no aumento do investimento, espaço e consumo energético, aspectos alvos de diversas discussões atualmente.

Nesse sentido, estudos apontam que a utilização de componentes de hardware como os sistemas baseados em SoC (*System-on-a-chip*) podem ser uma opção para sanar os problemas de infraestrutura dos *datacenters*, pois são recursos que encapsulam memória RAM, CPU, GPU entre outros componentes em um mesmo chip, possibilitando, conseqüentemente, que se reduzam o tamanho, o consumo energético e o custo do hardware em questão, além de proverem uma boa relação entre poder computacional e preço, quando comparados com estruturas de processamento de dados tradicionais. [Wolf et al. 2008]

Pensando nisso, o presente trabalho apresenta a implantação de um banco de dados não relacional e distribuído em um cluster de dispositivos SoC, visando explorar os aspectos de viabilidade do uso de tal abordagem como alternativa aos ambientes de armazenamento e processamento de dados tradicionais.

## 2. Materiais e Métodos

O cluster de SoC's é composto por 4 Raspberry Pis, dos quais dois são Raspberry Pi 3 (SoC Broadcom BCM2837, CPU ARM Cortex-A53 64-bit, 4 núcleos, 1.2 GHz, 1GB de

memória RAM) e 2 Raspberry Pi 4 (SoC Broadcom BCM2711, CPU ARM Cortex-A72 64-bit, 4 núcleos, 1.5 GHz, 4GB de memória RAM). A interconexão entre os dispositivos é realizada via rede, sendo que cada Raspberry está conectado a um *switch* de 1Gbps de velocidade. O Sistema Operacional utilizado em cada Raspberry é o HypriotOS 1.11.

Ao optar-se pelo uso de dispositivos não tão comuns para a composição de um ambiente distribuído, é comum o surgimento de alguns obstáculos, como dificuldades para gerenciar um ambiente distribuído formado por dispositivos com características de hardware heterogêneas e a dificuldade de deploy de uma aplicação, pois há bibliotecas e dependências de software distintas para cada arquitetura. Pensando nisso, no intuito de minimizar tais problemas foi utilizado o Docker Swarm (Docker versão 19.03) para o gerenciamento e orquestração dos recursos, que abstrai algumas tarefas de gerenciamento dos nós do cluster, bem como facilita a instalação, configuração e dimensionamento dos recursos.

Como banco de dados, foi escolhido o Apache Cassandra (versão 3.11.4), por ser um banco de dados não relacional altamente escalável e distribuído, que se enquadra como sendo do tipo AP (*Availability / Partition Tolerance*) segundo o teorema de CAP [Gilbert e Lynch 2002], provendo escalabilidade horizontal linear e tolerância a falhas em diversos tipos de infraestruturas. Além disso, experiências prévias como as apresentadas por [Richardson et al. 2017] mostram o potencial de uso do Cassandra aliado ao Raspberry Pi.

### 3. Resultados Preliminares e Trabalhos futuros

O ambiente proposto no trabalho encontra-se atualmente operacional no cluster. Dadas as restrições de hardware dos dispositivos, foi necessária uma configuração de uso de recursos de hardware diferente da padrão do Cassandra, sendo limitados o consumo de memória RAM e o número de instâncias do banco de dados por nó do cluster. Porém, mesmo com recursos limitados às características dos dispositivos, o banco de dados distribuído funcionou de forma estável durante inserções e consultas de teste realizadas. Com o uso do Docker Swarm, o processo de criar novas instâncias do Cassandra foi simplificado, bastando um único comando para escalar o número de instâncias e nós executando o serviço.

Futuramente pretende-se coletar e avaliar, através de *benchmarks* e testes, indicadores quantitativos que representem as vantagens e desvantagens do uso dessa estratégia como alternativa às comumente utilizadas. Assim, trabalhos a serem desenvolvidos investigarão aspectos como consumo energético, escalabilidade, desempenho e balanceamento de carga.

### Referências

- Gilbert, S. e Lynch, N. (2002). Brewer's conjecture and the feasibility of consistent available partition-tolerant web services. In *ACM SIGACT News*, page 2002.
- Richardson, D. P., Lin, A. C., e Pecarina, J. M. (2017). Hosting distributed databases on internet of things-scale devices. In *2017 IEEE Conference on Dependable and Secure Computing*, pages 352–357.
- Wolf, W., Jerraya, A. A., e Martin, G. (2008). Multiprocessor system-on-chip (mpsoc) technology. *Trans. Comp.-Aided Des. Integ. Cir. Sys.*, 27(10):1701–1713.