



## Análise Comparativa entre MySQL e MariaDB em Ambiente de Alta Disponibilidade

Gabriell Henrique Rodrigues do Nascimento, Flaudísio Tolentino Filho, Ana Claudia Moura Laurentino, Allysson Steve Mota Lacerda

### INTRODUÇÃO

Sabe-se que sistemas computacionais não estão isentos de falhas e indisponibilidades, que podem decorrer de panes elétricas ou de hardware e até mesmo de falhas do sistema operacional ou algum software base para a execução dos sistemas, dentre vários outros itens. Com o intuito de prover maior segurança e confiança, técnicas de alta disponibilidade são empregadas para que os sistemas computacionais estejam disponíveis aos seus usuários pela maior quantidade de tempo possível. Tais técnicas, no entanto, por envolverem a cópia de dados, por motivo de segurança, de um computador para outro, garantindo assim a disponibilidade dos dados em caso de falhas em um dos computadores, pode acarretar a perda de desempenho do sistema, deixando a sensação de lentidão ao usuário final. Desse modo, este artigo tem como objetivo comparar o desempenho de sistemas gerenciadores de bancos de dados (SGBD) configurados como clusters de alta disponibilidade. Os SGBDs são serviços básicos para diversos tipos de sistemas computacionais e os dados neles armazenados são muitas vezes sensíveis e extremamente necessários para o funcionamento correto do sistema. Por esses motivos os SGBDs precisam ser confiáveis, consistentes, seguros e velozes, mesmo nas situações adversas.

### MATERIAL E MÉTODOS

#### A. MySQL e MariaDB

MySQL e MariaDB são SGBDs relacionais que utilizam a Structured Query Language (SQL) como interface para acesso aos dados que armazenam. Criado por um dos fundadores do MySQL, o MariaDB surgiu como um fork do projeto do MySQL, após a aquisição deste pela Oracle. Para manter a compatibilidade entre os SGBDs, o desenvolvimento do MariaDB, até o lançamento da versão 10, seguiu o processo de desenvolvimento do MySQL e a partir de então o MariaDB recebeu funcionalidades novas e diferentes das encontradas naquele SGBD [1] [2].

MySQL e, por conseguinte, MariaDB possuem diversos recursos, sendo um dos mais importantes o suporte a diferentes engines de armazenamento, cada qual oferecendo características distintas, que podem influenciar em desempenho de escrita e/ou leitura de dados, funcionalidades, relacionamento entre os dados no banco de dados, entre outros [2].

A base de usuários do MySQL é extensa, no entanto empresas como a Wikipédia já optaram por trocar o MySQL pelo MariaDB. Ambos os SGBDs possuem suporte a replicação de dados, possibilitando, desse modo, seu uso para a configuração de um ambiente de alta disponibilidade.

#### B. Galera Cluster

O Galera Cluster é um software para replicação de dados de SGBDs que funciona como uma camada única de comunicação entre os servidores nos quais está configurado o cluster. Um cluster Galera é um conjunto de máquinas que possui este software instalado e configurado para se comunicar com sua instância nas outras máquinas do cluster [3]. Sendo assim, todas as máquinas do cluster podem enviar e receber mensagens sobre o estado dos dados de seus respectivos SGBDs e, a partir disso, realizar procedimentos para que os dados de seus SGBDs estejam idênticos aos dados de todos os outros SGBDs do conjunto, mantendo, desse modo, a consistência dos dados.

O controle do estado dos dados dos SGBDs é feito através do uso de Identificadores Globais de Transação (Global Transaction Identifiers – GTID). Utilizando-se de GTIDs é possível identificar, de forma única em todo o cluster, as mudanças de estado de cada SGBD do conjunto [4]. O estado, então, refere-se aos dados manipulados pela aplicação, que, no caso dos SGBDs, refere-se ao seu conteúdo, isto é, os dados que armazena, cuja manipulação se dá com grande frequência. Esse modo de replicação utilizado pelo Galera é chamado de Replicação Baseada em Certificação (Certification Based Replication) [4].

O Galera Cluster pode ser utilizado em um ambiente de alta disponibilidade que utiliza como SGBD tanto o MySQL quanto o MariaDB.



### C. Sysbench

O Sysbench é uma ferramenta de benchmarking capaz de executar repetidos testes em serviços específicos com o intuito de avaliar o desempenho de tais serviços. Essa ferramenta é modular e possui suporte à execução de múltiplas threads [5]. Entre os serviços nos quais os testes do Sysbench podem ser executados estão os SGBDs.

Nos SGBDs, o Sysbench permite executar testes que verificam o desempenho de escrita e leitura nas bases de dados, utilizando diferentes engines de armazenamento, números de linhas nas tabelas e, ainda, especificando a utilização ou não de transações, entre outras opções [5]. Desse modo, essa ferramenta se torna de valia para mensurar se há impacto no desempenho dos SGBDs MySQL e MariaDB ao se utilizar o Galera Cluster como o instrumento para prover alta disponibilidade.

### D. Ambiente de testes

Os SGBDs MySQL e MariaDB, por serem compatíveis, puderam ser executados com o mesmo arquivo de configuração que, dentre outras configurações, tinha variáveis relacionadas a desempenho definidas de acordo com a Tabela 1. Todos os testes nos SGBDs realizados pelo Sysbench foram executados utilizando os mesmos parâmetros, conforme Fig. 1, com exceção das opções `--oltp-table-name`, `--mysql-db`, `--mysql-user` e `--mysql-password`. Os testes foram executados por 5 vezes, durante 60 segundos cada, utilizando 16, 32 e 128 threads. Dos resultados colhidos foi feita uma média do valor de transações por segundo dos 5 testes realizados para cada configuração. O valor de transações por segundo resultante da média calculada para cada configuração foi o utilizado para fins de comparativo.

## RESULTADOS

Os resultados obtidos foram coletados e analisados e, a partir da média da quantidade de transações por segundo que cada SGBD alcançou, foi gerado o gráfico da Fig. 1. Esses resultados indicam a média de transações por segundo em um intervalo de 60 segundos.

Com base nos resultados, é possível afirmar que, no caso do MySQL, houve perda de desempenho por parte do uso do Galera Cluster, mesmo existindo apenas um único nó no cluster. Por outro lado, o MariaDB obteve ganho no desempenho ao se utilizar o Galera Cluster – um fato inusitado, de início.

Analisando-se os dados dos SGBDs quando não utilizando o Galera Cluster nota-se que ambos se equivalem em desempenho. Ao considerar, além dos valores das médias, os valores máximos e mínimos de transações por segundo, percebe-se que há interseção entre os valores obtidos nos testes realizados para os dois SGBDs. Desse modo, o desempenho dos SGBDs é similar utilizando-se a configuração definida na Tabela 1, sem o uso do Galera Cluster.

## DISCUSSÃO

Inicialmente era esperado que o Galera Cluster influenciasse o desempenho dos SGBDs de forma negativa – como ocorrido com o MySQL –, pelo fato de que haveria uma camada adicional no SGBD. O fato de o MariaDB ter alcançado um desempenho melhor ao se utilizar o Galera Cluster é intrigante. Uma hipótese que poderia justificar esse resultado é a de que a configuração do Galera Cluster, ao se utilizar em conjunto com o MariaDB, define, neste último, variáveis que impactam no desempenho do SGBD, ou seja, o MariaDB Galera Cluster já tem configurações padrão que aumentam o desempenho desse SGBD. Essa hipótese, no entanto, deve ser melhor estudada a partir de pesquisa com base na configuração do MariaDB Galera Cluster.

Quanto aos outros resultados, era esperado um desempenho similar entre MySQL e MariaDB e um desempenho pior do MySQL quando utilizando o Galera Cluster. Esses dados foram comprovados pela pesquisa realizada.

## CONCLUSÃO

O uso de um software para a replicação de dados com vistas a garantir alta disponibilidade pode impactar o desempenho de certos SGBDs. Todavia, há de se verificar se a perda de desempenho que se terá é justificada pela segurança a mais obtida pelo uso de um software que garanta a alta disponibilidade dos dados armazenados no SGBD. A necessidade ou não do uso de tal ferramenta dependerá da finalidade e dos requisitos que terá a aplicação final.

## REFERÊNCIAS

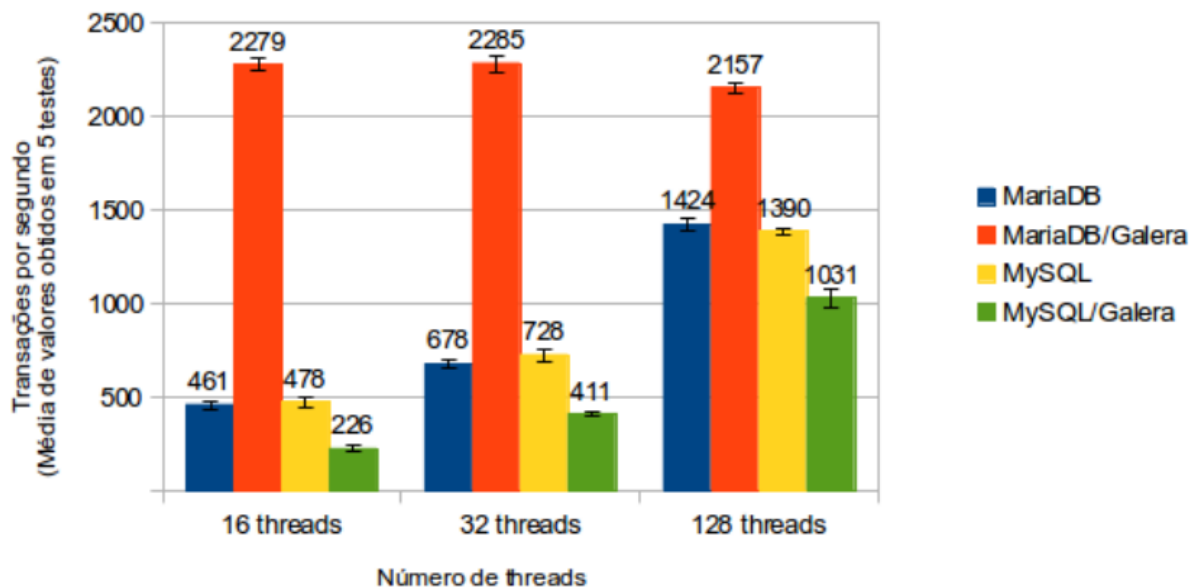
- [1] MARIADB KNOWLEDGE BASE. MariaDB versus MySQL – Compatibility. 2014. Disponível em: <https://mariadb.com/kb/en/mariadb/mariadb-vs-mysql-compatibility/>. Acesso em: 22 Jul. 2014.
- [2] MARIADB KNOWLEDGE BASE. MariaDB versus MySQL – Features. 2014. Disponível em: <https://mariadb.com/kb/en/mariadb/mariadb-vs-mysql-features/>. Acesso em: 22 Jul. 2014.
- [3] GALERA CLUSTER. Galera Cluster Documentation. 2014. Disponível em: <http://galeracluster.com/documentation-webpages/index.html>. Acesso em: 7 Ago. 2014.
- [4] GALERA CLUSTER. Certification Based Replication. 2014. Disponível em: <http://galeracluster.com/documentation-webpages/certificationbasedreplication.html>. Acesso em: 7 Ago. 2014.
- [5] KOPYTOV, A. sysbench User Manual: SYSBENCH(1). 2014.

**Tabela 1.** Configuração utilizada no MySQL e no MariaDB durante os testes realizados pelo Sysbench.

| Variável                          | Valor  |
|-----------------------------------|--------|
| --default_storage_engine          | InnoDB |
| --query_cache_limit               | 4M     |
| --query_cache_size                | 32M    |
| --innodb_additional_mem_pool_size | 16M    |
| --innodb_buffer_pool_size         | 256M   |
| --innodb_commit_concurrency       | 0      |
| --innodb_flush_log_at_trx_commit  | 1      |
| --innodb_lock_wait_timeout        | 50     |
| --innodb_log_buffer_size          | 8M     |
| --innodb_log_file_size            | 32M    |
| --innodb_log_files_in_group       | 2      |
| --innodb_max_dirty_pages_pct      | 90     |
| --innodb_sync_spin_loops          | 20     |
| --innodb_thread_concurrency       | 8      |
| --join_buffer_size                | 1M     |
| --max_heap_table_size             | 64M    |
| --tmp table size                  | 64M    |

Fonte: Dados da pesquisa

**Figura 1.** Comparativo de transações por segundo processadas pelos SGBDs. O gráfico indica a média dos valores obtidos de 5 execuções de 60 segundos de cada teste juntamente do desvio padrão de cada resultado.



Fonte: Dados da pesquisa