Database Performance for Data Engineers

Peter Zaitsev
Founder at PERCONA

ConFoo.CA
DEVELOPER CONFERENCE
February 26-28, 2025

All those Titles

Database Administrator

Database Architect

Database Reliability Engineer

Full Stack Developer

Site Reliability Engineer

DevOps

… Growing number of engineers have to deal with database but it is not the only part of the job

Poll

How many of you are…

Database is a Black Box

You can connect to the Database Service Point, Quickly

Run Queries you need to run

Meaning

Queries

1

Run them without errors

2

Run them with correct results

3

Run them with required response time

Performance

Performance is about Response Time you get for your Queries

Response Time – Database View

“I see database responds to queries in 5ms in average”

Response Time – Business View

All Users have outstanding performance experience with all their application interactions

Downtime

Very Bad Performance is indistinguishable from downtime

Over Time

LOOK AT RESPONSE TIME TRENDS OVER TIME

MINOR SLOWDOWN OFTEN HAPPENS BEFORE POOR PERFORMANCE “DOWNTIME”

PERFORMANCE CAN BE WORSE AT CERTAIN TIMES – BACKUPS, BATCH JOBS, MAINTENANCE

What Drives Performance ?

Performance Depends on

Amount of data which needs to be processed

Processing Speed

PERCONA

Logical – Actually needed to deliver result

Physical – Which end up needing to be processed

Amount of Data

Impacting Logical Data Size

- **Schema**
- **Query Plan**
- **Summarized, Pre-Generated Data**
- **Caching**

PERCONA

Impacting Physical Data Size

- Row store vs Column Store
- Normalization and De-Normalization
- Compression, Dictionaries
- “Indexes”

PERCONA

Sort Index , Multiple Sorting

Skip Indexes (ie Min/Max Value based)

Bloom Filters

Indexes - Beyond BTree

Vectors

Enabling AI Applications – Vectors and Vector Search

Does not work well with conventional indexes

Special Support in the database is required

PERCONA

Algorithms

Hardware

Resource Sharing

Processing Speed

Algorithms

Most Advanced Algorithms

High Level and Low Level Optimization

Lots of innovation constantly happening

PERCONA

Hardware

- Processor speed
- GPU Speed And Capabilities
- Vectorized Execution Capability
- Memory amount and Bandwidth
- Network Bandwidth
- Storage Bandwidth
- Number of CPU Cores/Threads

PERCONA

Execution

Vectorized Execution

Parallelized Execution

Distributed Execution

PERCONA

Data Storage

In Memory

On Disk

PERCONA

Deployment Architecture

Fixed Resources

Dynamic Resources

Mixed (Dynamic Storage)

PERCONA

Analytical
Use Case

Sustained

Periodic

PERCONA

Performance and Cost

Cost Per Environment vs Cost Per Execution

Architecture Trends

Serverless

Separation of Storage and Compute

Cloud Native

PERCONA

Tradeoffs

No Perfect Solution, All about finding optimal tradeoffs

Tools worth mentioning

Percona Monitoring and Management (PMM)

https://www.percona.com/software/database-tools/percona-monitoring-and-management

PERCONA

coroot :~#

PERCONA

https://www.linkedin.com/in/peterzaitsev/

https://twitter.com/PeterZaitsev

http://www.peterzaitsev.com

Q&A
Thank you!