Leveraging object storage: Tiered Storage for ClickHouse

Arthur Ansquer

Agenda

- What is ClickHouse?
- ClickHouse as a service
- Problem
- Solution: Tiered Storage

What is ClickHouse?

Arthur Ansquer - OSA CON 2023

What is ClickHouse?

- High-performance, column-oriented SQL database management system (DBMS) for online analytical processing (OLAP)
- Developed by Yandex, open-source since 2016

Why is it gaining momentum?

- Column-oriented
- Incredible compression
- Fast ⚡
- SQL-based
- Distributed database management system
- Can talk to pretty much every technology!

ClickHouse use cases

- Real-time Analytics
- Business Intelligence

Even ...

- Logs & Traces
- GenAI with vector Search
- Machine learning

It is just **THAT good!**

ClickHouse as a service

Arthur Ansquer - OSA CON 2023

Who used ClickHouse and why?

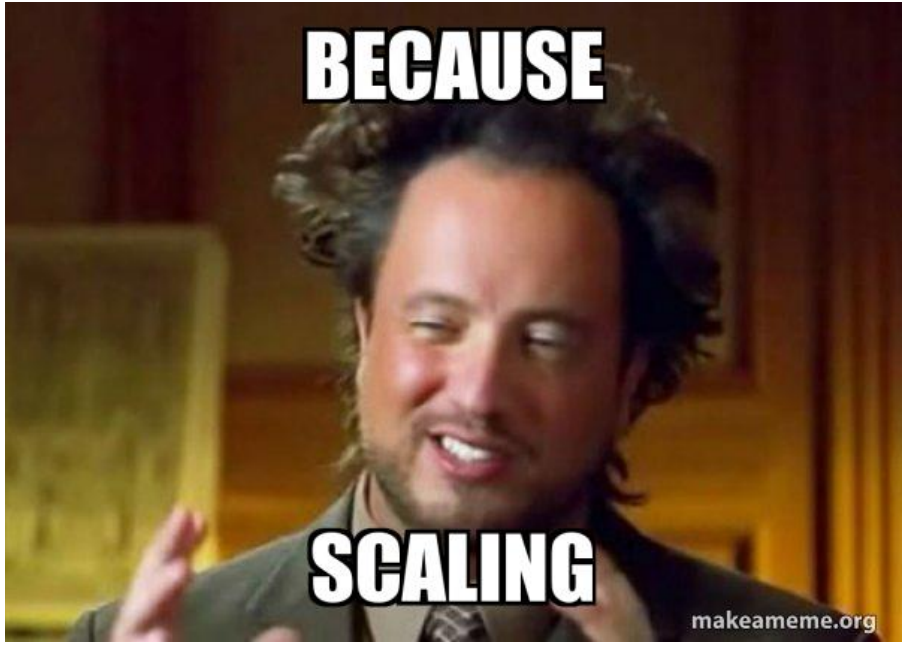
- Anyone with **analytics workloads** ...

- That needs **very high performance** (sub-second latency)
- Or queries over **large amount** (billions) of **rows**

- No matter where your data may be.

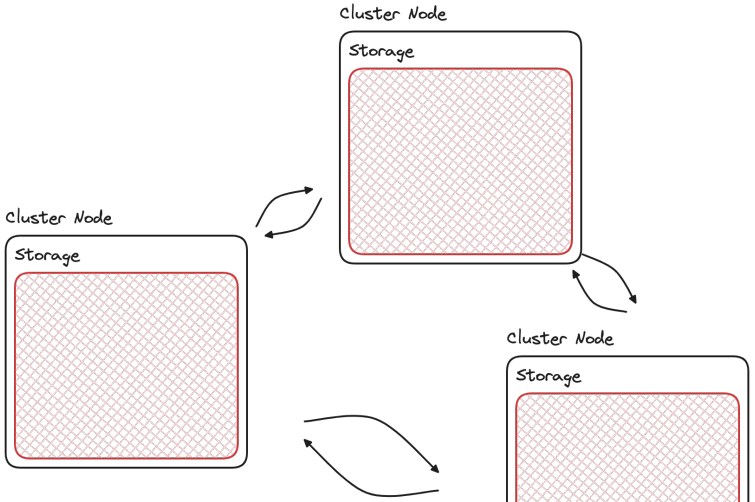
Who failed to use it and why?

- No **compelling scaling** story



Problem: Storage is a finite resource

Problem #1: Scale



- We need to make sure we **support use cases of any size**
- And storage **can't scale independently**

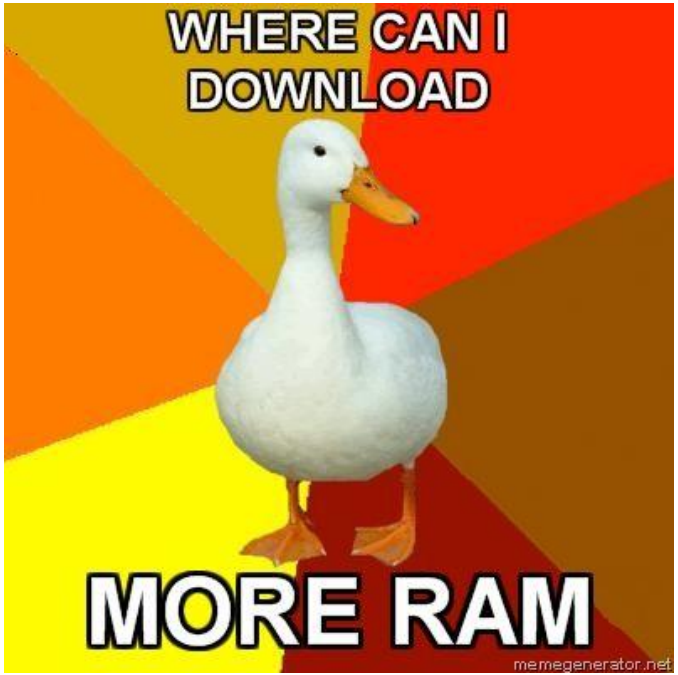
Problem #2: Cost

- **SSDs can get expensive**

Problem #3: Performance

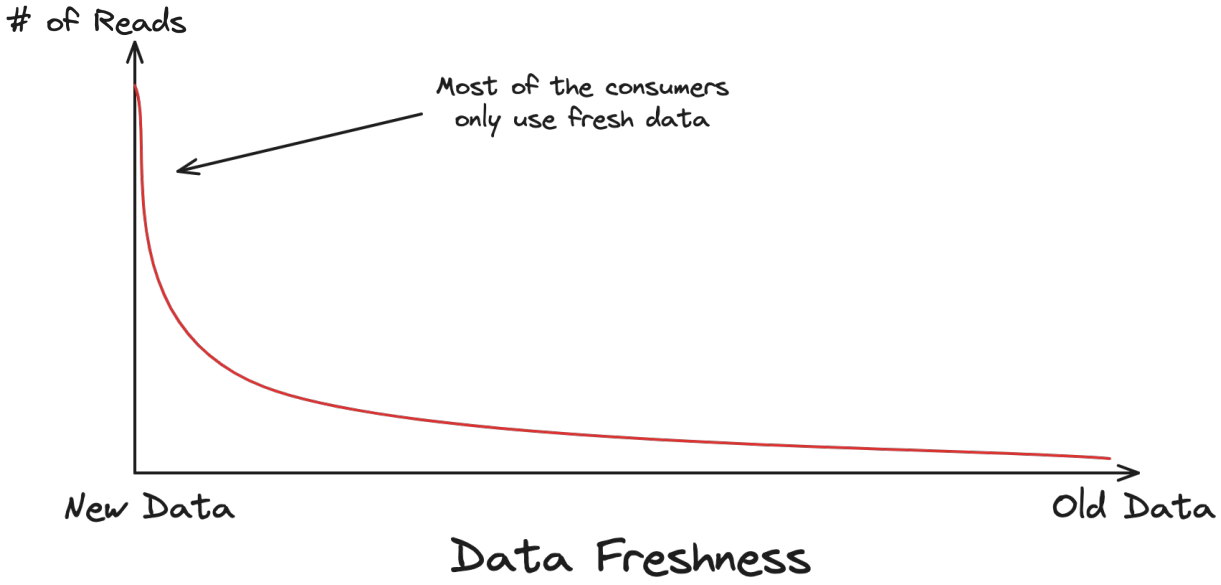
- **Avoid performance hit** at all cost
- **Errors** can be costly when pushing local storage too hard

In need of a more elegant solution…

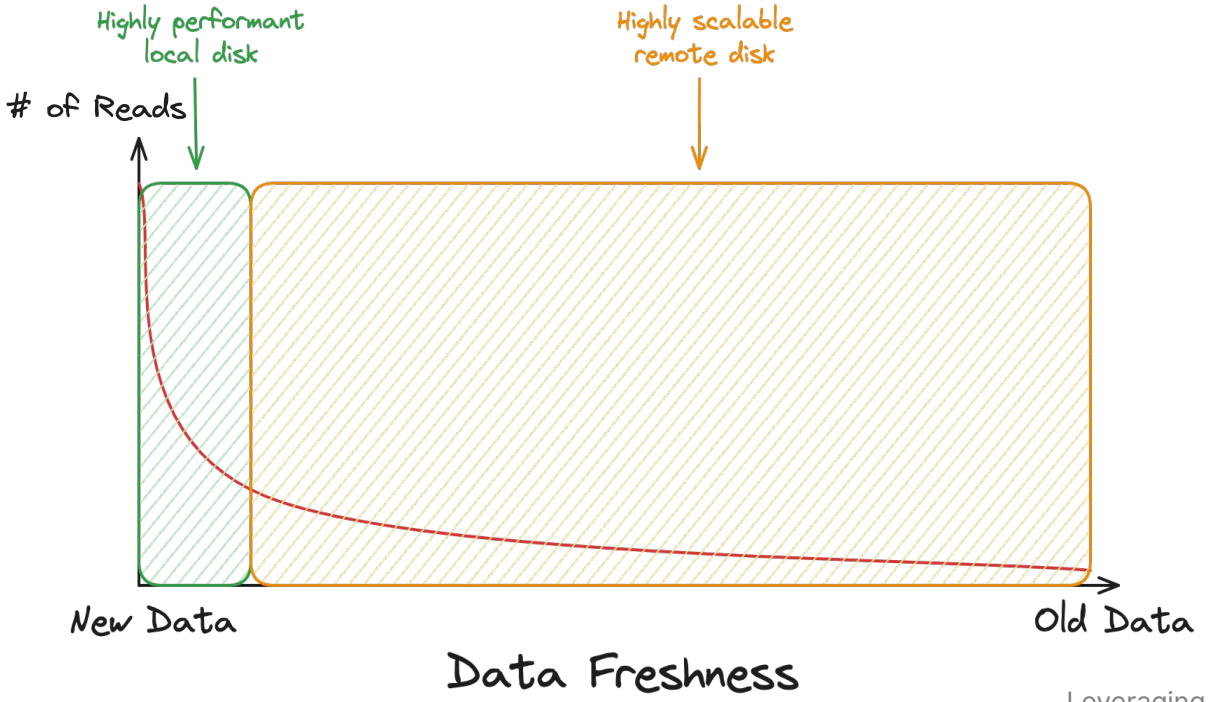


Solution: Tiered Storage

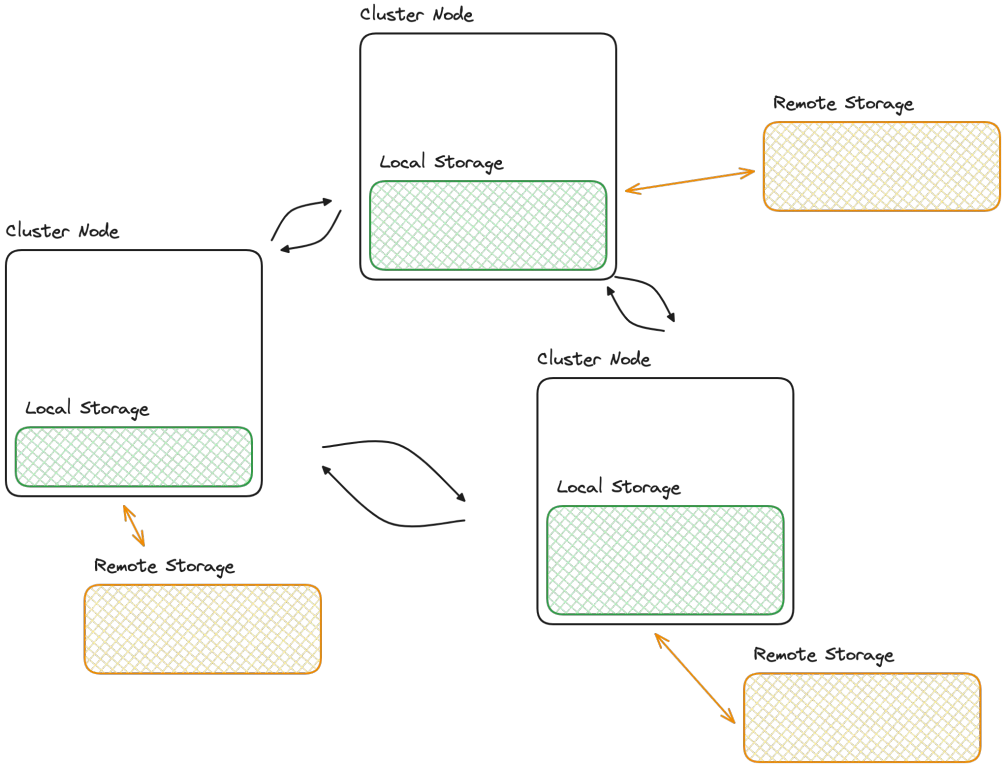
Understand your data access patterns



What is Tiered Storage?



What is Tiered Storage?



Why is it so great?

- Scale
- Cost
- Performance

Wrapping up


WE HEARD YOU LIKED STORAGE
SO WE ADDED STORAGE TO YOUR STORAGE
imgflip.com

Thank you!