

МАСШТАБОВАНІСТЬ БАЗ ДАНИХ: ВИКЛИКИ ТА РІШЕННЯ

Database Scalability



01

Поняття масштабованості

02

Вертикальне масштабування

03

Горизонтальне масштабування

04

Реплікація

05

Індексування

06

Summary

Маштабованість

властивість системи, мережі, або процесу, яка свідчить про здатність системи обробити більший обсяг роботи, даних або бути легко розширеною

Застосування:

- зростання обсягу даних --- із зростанням обсягу продажів збільшується обсяг даних
- збільшення кількості користувачів --- соціальні мережі, онлайн-ігри, веб-сайти
- висока доступність --- доступ до даних 24/7
- пікові навантаження --- сезонні, періодичні навантаження
- масштабування для віддалених регіонів --- компанії з глобальною присутністю

Основні недоліки:

- складність реалізації --- значні зусилля від команди розробників
- збільшення витрат --- обладнання, ПЗ, підтримка
- системна складність --- ускладнення управління та підтримки системи
- ризик безпеки даних --- втрата/виток інформації

ОСНОВНІ ВИДИ МАШТАБУВАННЯ БД

Вертикальне

оновлення сервера
бази даних до машини
з більшою пам'яттю,
швидкістю процесора/
ядрами, ємністю
введення/виведення
тощо

Горизонтальне

збільшення потужності
системи шляхом
додавання додаткових
машин (вузлів)

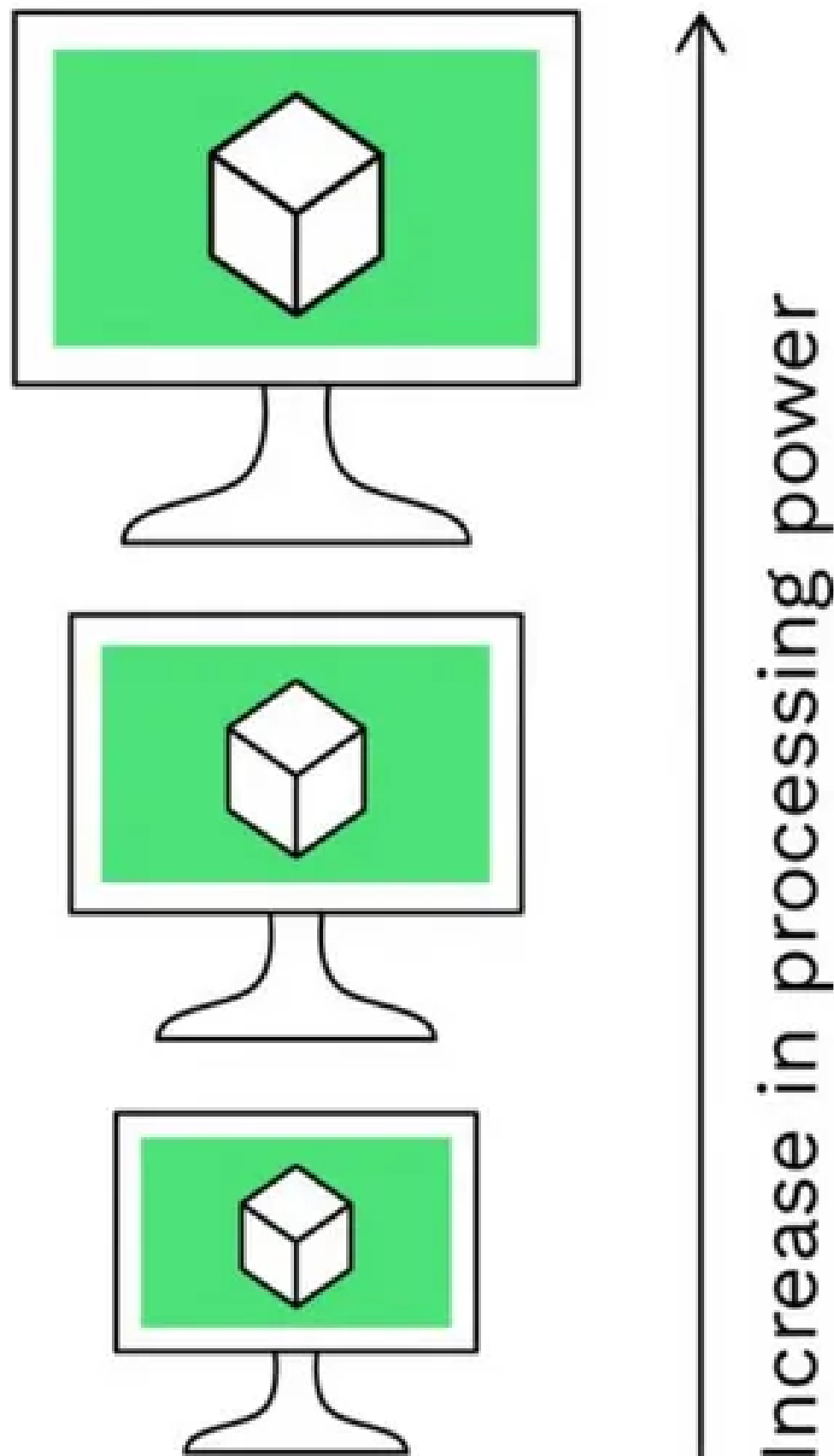
Реплікація

процес створення та
підтримки копій баз
даних на різних вузлах

Індексування

створити структуру
даних, яка дозволяє
ефективно виконувати
пошук і фільтрацію
записів без
необхідності
перегляду всіх записів
у базі даних

Vertical scaling



Вертикальне масштабування:

оновлення фізичної машини,
на якій працює система

Мета використання:

підвищити продуктивність, можливості та потужність
обробки даних на існуючому сервері

Основні недоліки:

- межі розширення є обмеженими
- складність адміністрування
- вартість --- збільшення потужності сервера вимагає великих витрат

USECASE:

уявімо інтернет-магазин, який продає тисячі різних товарів. З часом кількість клієнтів та обсяги продаж можуть значно зрости, що призведе до збільшення обсягу даних в базі.

Горизонтальне масштабування:

збільшення потужності системи шляхом додавання додаткових машин

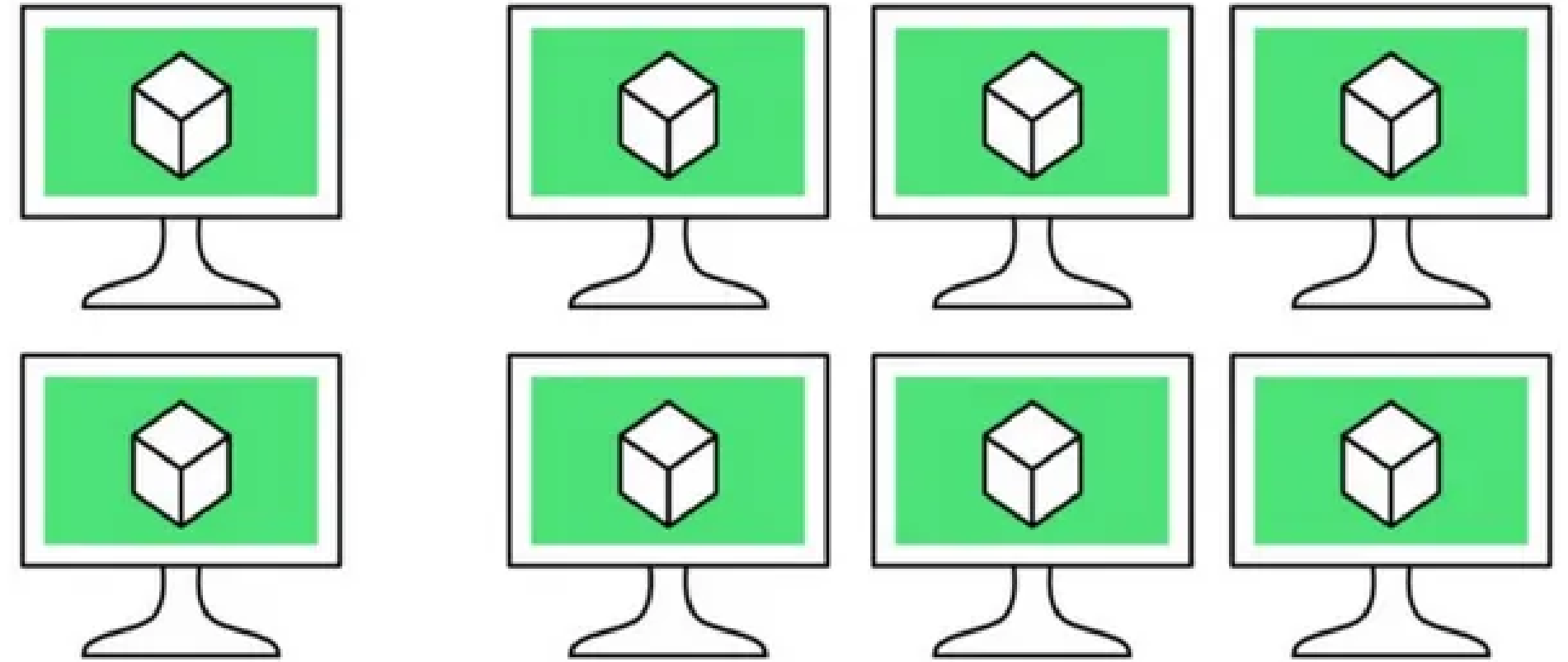
Переваги:

- висока доступність --- розподіл навантаження без ризиків відмов системи
- легкість розширення --- додання нових серверів без переривання роботи
- економічність --- кількість серверів збільшується за потреби
- гнучкість

Недоліки:

- складність програмування
- більший обсяг даних --- складність керування, синхронізації
- мережева завадостійкість --- система може стати більш вразливою до мережевих збоїв

Horizontal scaling



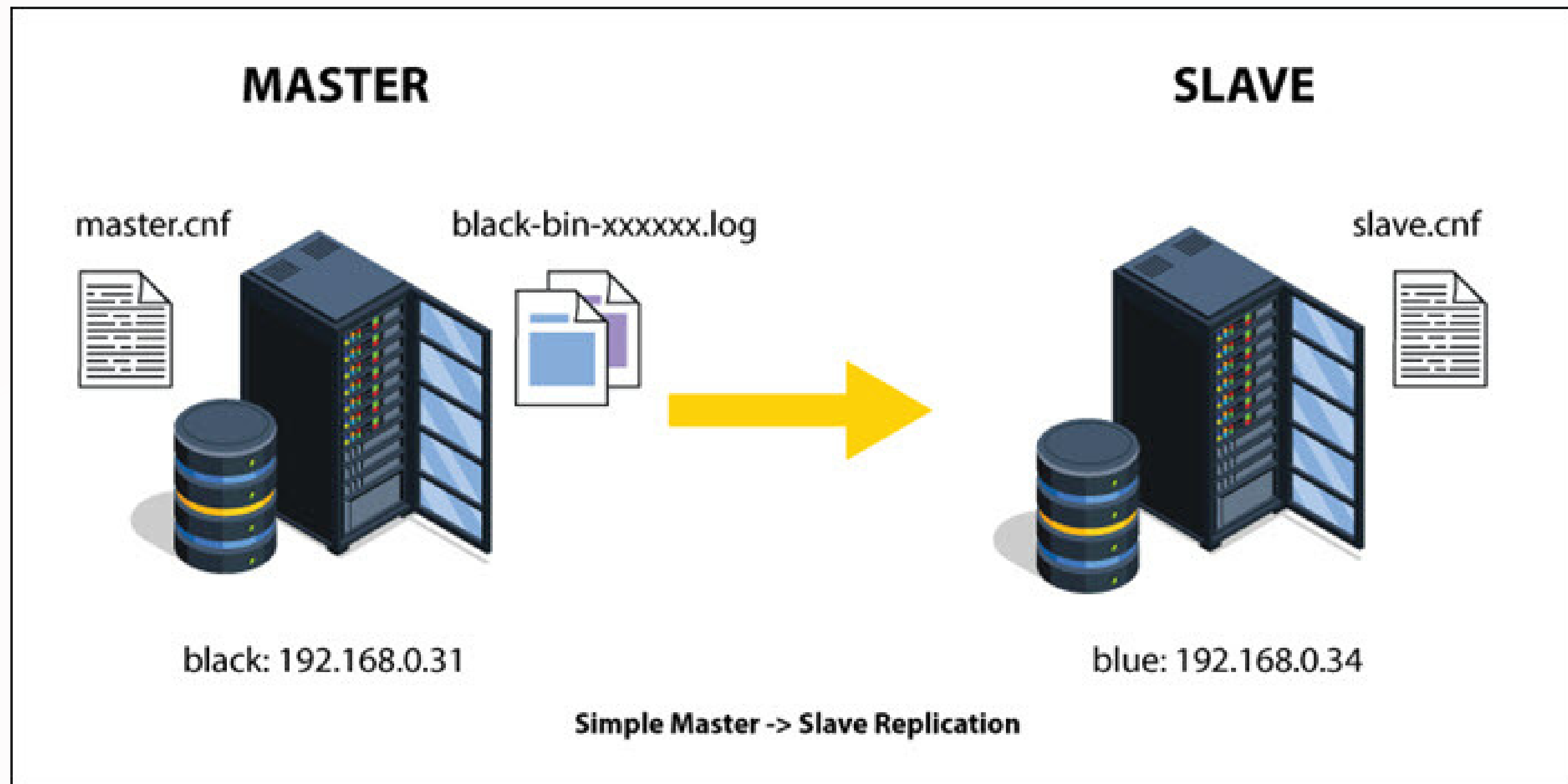
Increase in number of machines

Приклади використання: веб-програми, обробка даних, хмарні сервіси, мультимедійні сервіси

Реплікація

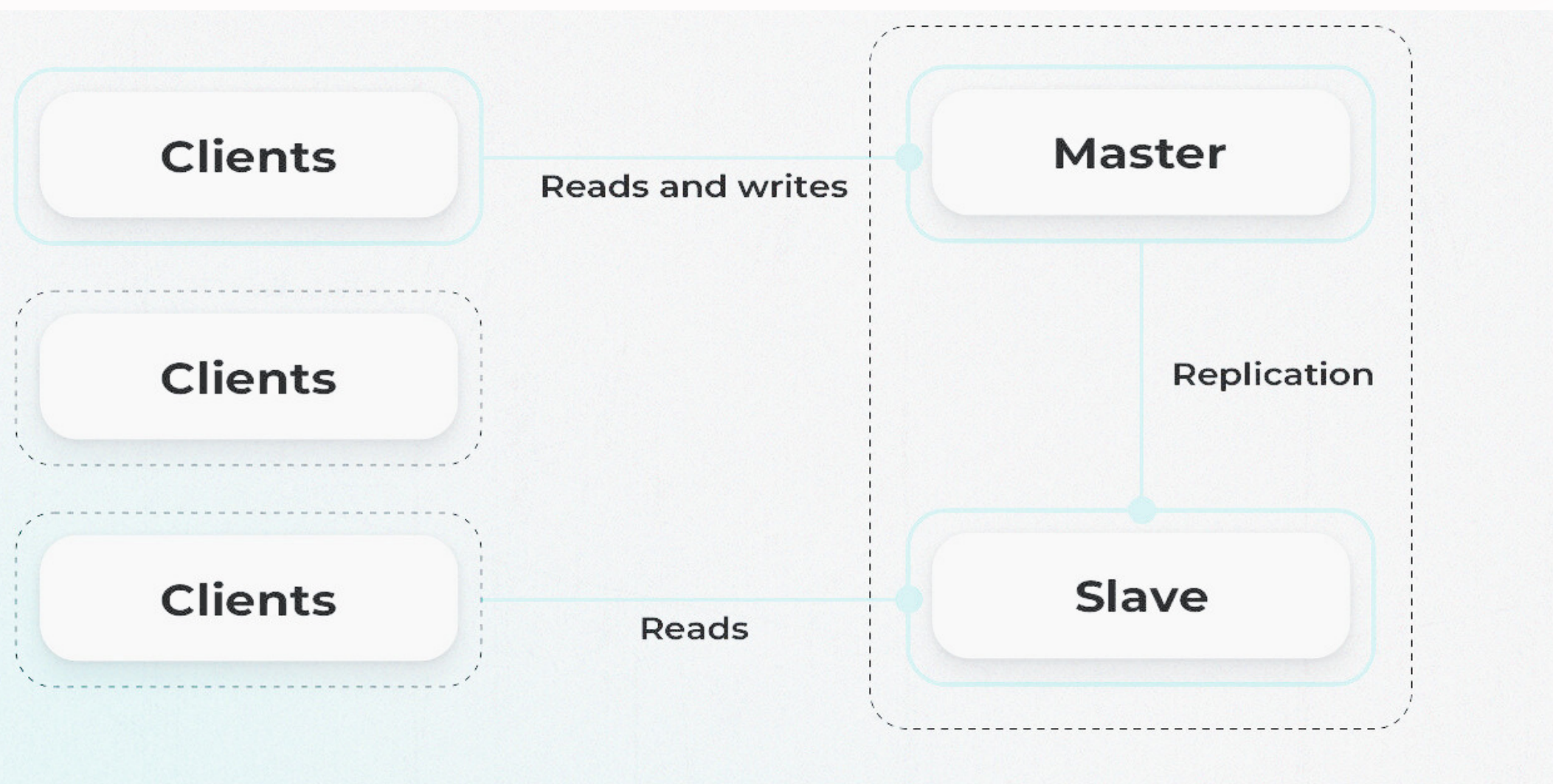
це процес створення та підтримки копій баз даних на різних вузлах з метою забезпечення високої доступності, збереження даних та розподілу навантаження.

Реплікація має дві схеми роботи: Master/Slave і Master/Master.



Master/Slave:

Це підхід, у якому є головний (Master) вузол, який копіює дані на одну або декілька реплік (Slave).

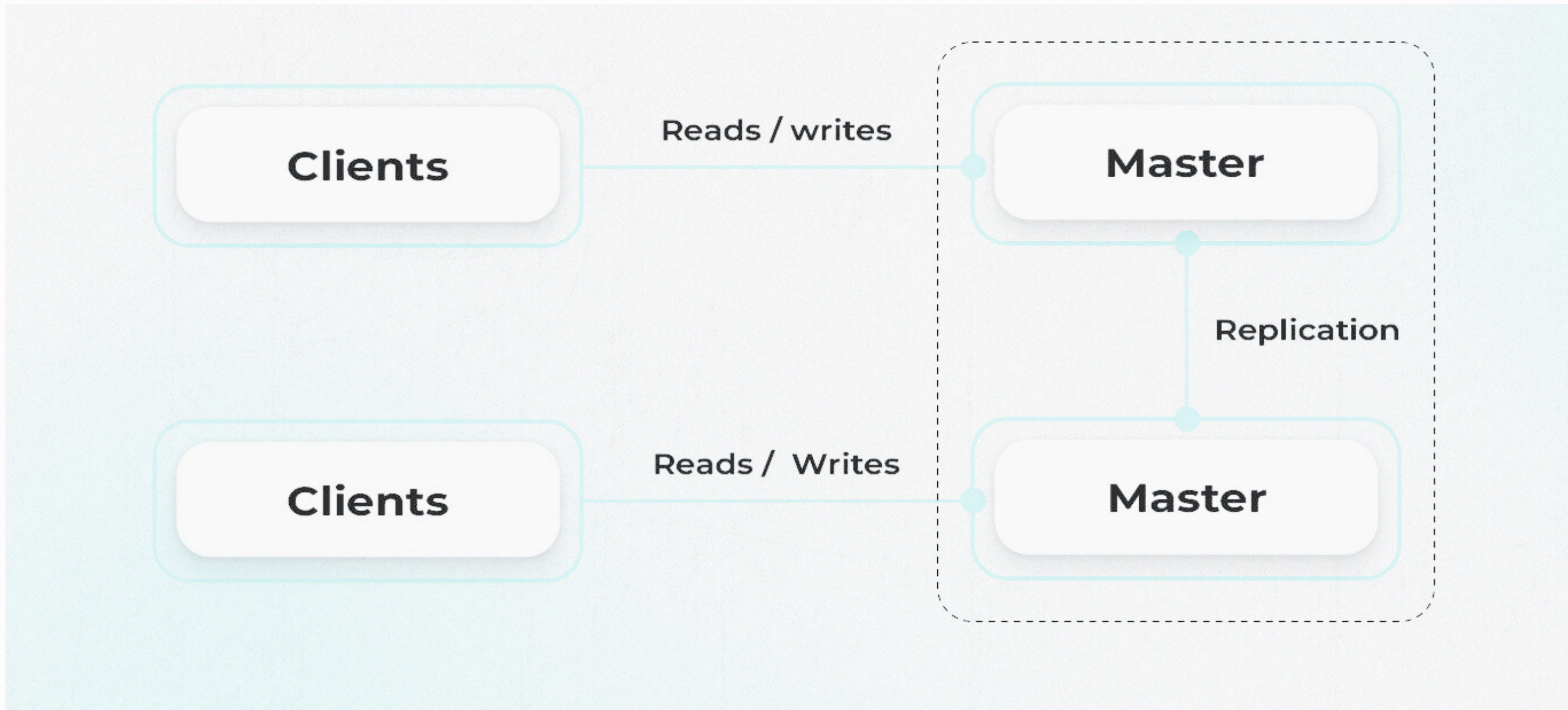


- Зміни (такі як insert/update/delete тощо) на головному вузлі.
- Звідти їх отримують та зберігають репліки.

Master відповідає за запис та читання,
а Slave — тільки за читання.

Підхід реплікації Master/Master корисний, коли вашій системі потрібно виконувати операції запису у великій кількості.

Master/Master:



Підхід Master/Master допоможе розв'язати цю проблему синхронізації даних.

Недоліки реплікації

Логіка в коді



Затримка оновлення
даних



Збільшення
інфраструктури



Проте стратегія реплікації є корисною для
подолання пікових навантажень.



Техніка індексування

Індексування баз даних є важливою стратегією для оптимізації швидкості доступу до даних в масштабованих системах.

01

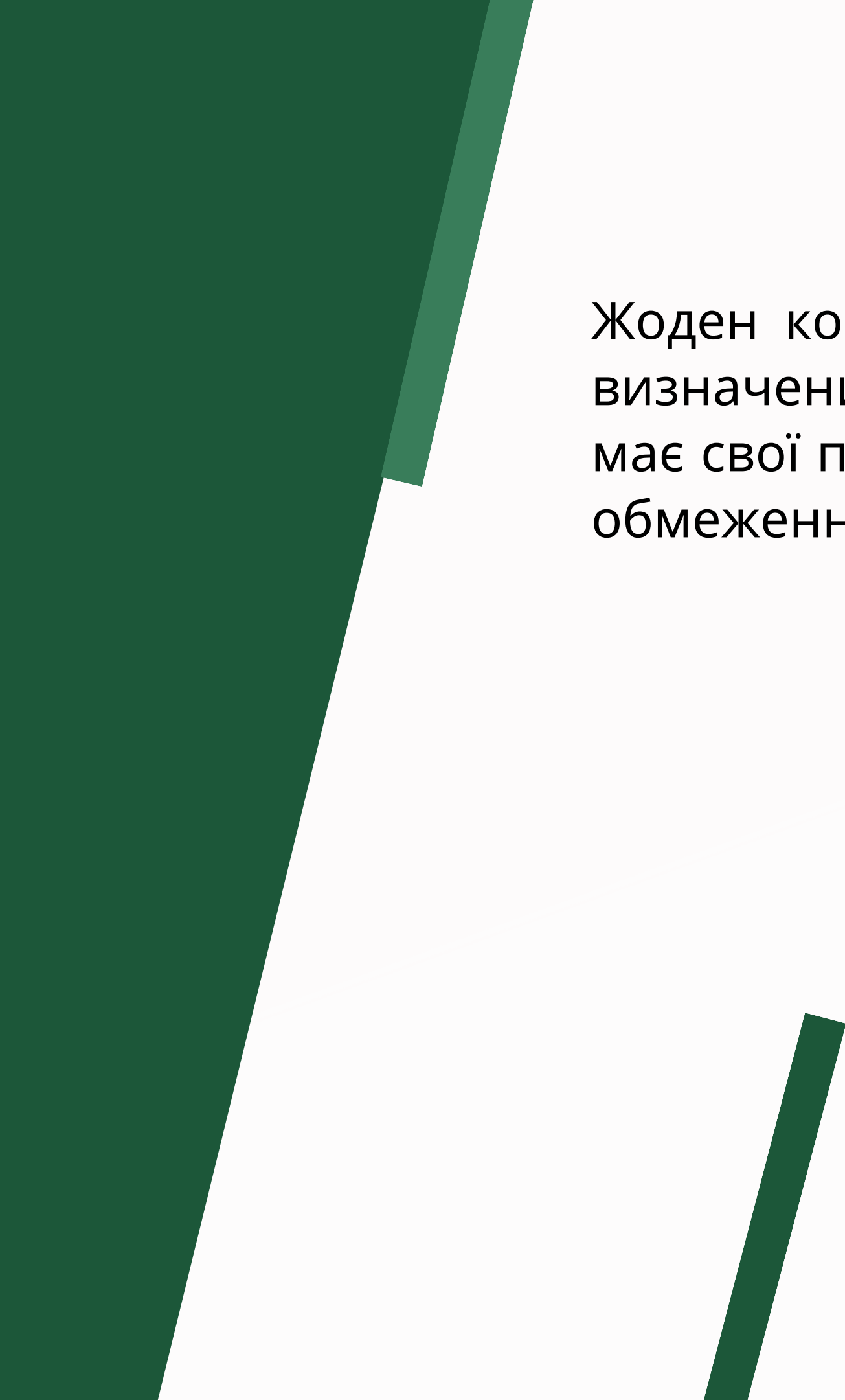
Кластерне індексування:
групування записів зі схожими
значеннями в одну область
пам'яті

Вторинне індексування: Цей підхід
полягає в створенні окремих індексів
для полів, які часто
використовуються для фільтрації або
сортування даних.

02

03

Багаторівневе індексування: Цей підхід
використовується для розділення
індексів на рівні, де кожен рівень вказує
на певний діапазон значень.



Жоден конкретний підхід масштабування баз даних не може бути визначений як "найкращий" у всіх випадках, оскільки кожен з них має свої переваги і недоліки, а також відповідає на різні потреби та обмеження.

На практиці часто використовується комбінація цих підходів в залежності від конкретних потреб, обмежень та вимог проекту.

Наприклад, горизонтальне масштабування може бути використане для обробки великого обсягу транзакцій, а реплікація - для забезпечення стійкості та доступності.