System Design

Databases e Indexação de Dados

$ whoami

Matheus Fidelis

Engenheiro de $RANDOM

@fidelissauro

https://fidelissauro.dev

https://linktr.ee/fidelissauro

OBJETIVOS

- Definição e papel do Database
- Tipos de Databases
- Níveis de Consistêncai
- Modelos de Dados
- Armazenamento e Indexação
- Cenários Arquiteturais

1
Definindo um Databases

Definições e Conceitos

Definindo um Banco de Dados

- Organização de Abstração de Dados
- Camada de Software entre o Cliente e o Armazenamento
- Considerações em Sistemas Distribuídos
- Replicação, Geo-Distribuição, Consistência

2. Tipos de Bancos de Dados

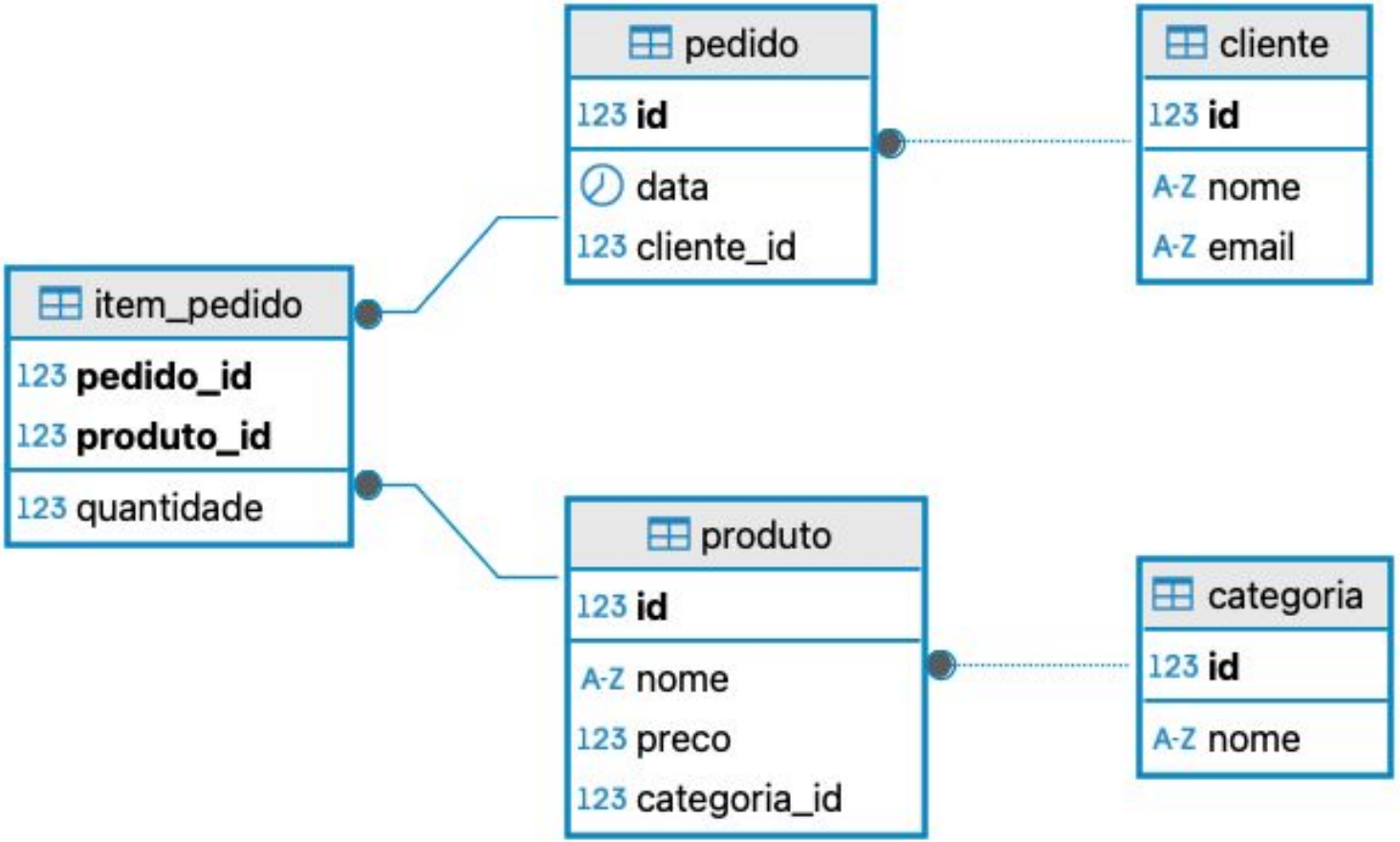
Definições e Conceitos

Tipos de Bancos de Dados

- Racional de Features
- Performance, Consistência, Custo
- Bancos Relacionais
- Bancos Não-Relacionais
- Bancos NewSQL
- Bancos de Dados em Memória
- Timeseries Databases

Bancos de Dados Relacionais (SQL)

- Modelo de Codd (1970) - Artigo
- Tabelas, Tuplas (linhas) e Colunas
- Schema Rígido e Declarativo
- Consistência Forte
- ACID
- Integridade Referencial Forte
- Cenários Transacionais
- Foco em Integridade e Durabilidade


pedido
123 id
data
123 cliente_id
cliente
123 id
A-Z nome
A-Z email
item_pedido
123 pedido_id
123 produto_id
123 quantidade
produto
123 id
A-Z nome
123 preco
123 categoria_id
categoria
123 id
A-Z nome

Bancos de Dados Relacionais (SQL)

- MySQL
- MariaDB
- PostgreSQL
- SQLServer
- Oracle
- Etc…

Bancos de Dados Não-Relacionais (NoSQL)

- Schemas Flexíveis
- Consistência Eventual
- Diversos Formatos de Dados
- Documentos, Chave-Valor, Grafos, Colunar
- Escala Horizontal Otimizada
- Sem Garantia de Integridade e Atomicidade
- Dados Semi-Estruturados

```
{
  "_id": ObjectId("60f5a2d1a2e9b5f1d4c8e918"),
  "nome": "Ana Silva",
  "email": "ana.silva@exemplo.com",
  "pedidos": [
    {
      "pedidoId": "PED12345",
      "data": "2025-07-27T14:35:00Z",
      "itens": [
        {
          "produto": {
            "id": ObjectId("60f5a3e8a2e9b5f1d4c8e91a"),
            "nome": "Camiseta Manga Curta",
            "preco": 79.90
          },
          "quantidade": 2
        },
        {
          "produto": {
            "id": ObjectId("60f5a3f2a2e9b5f1d4c8e91b"),
            "nome": "Calça Jeans",
            "preco": 149.90
          },
          "quantidade": 1
        }
      ]
    },
    {
      "pedidoId": "PED12346",
      "data": "2025-07-28T09:20:00Z",
      "itens": [
```

Bancos de Dados Não-Relacionais (NoSQL)

- MongoDB
- Elasticsearch
- Cassandra
- DynamoDB
- CosmosDB
- ScyllaDB
- Redis
- Memcached
- Etc…

Bancos de Dados NewSQL

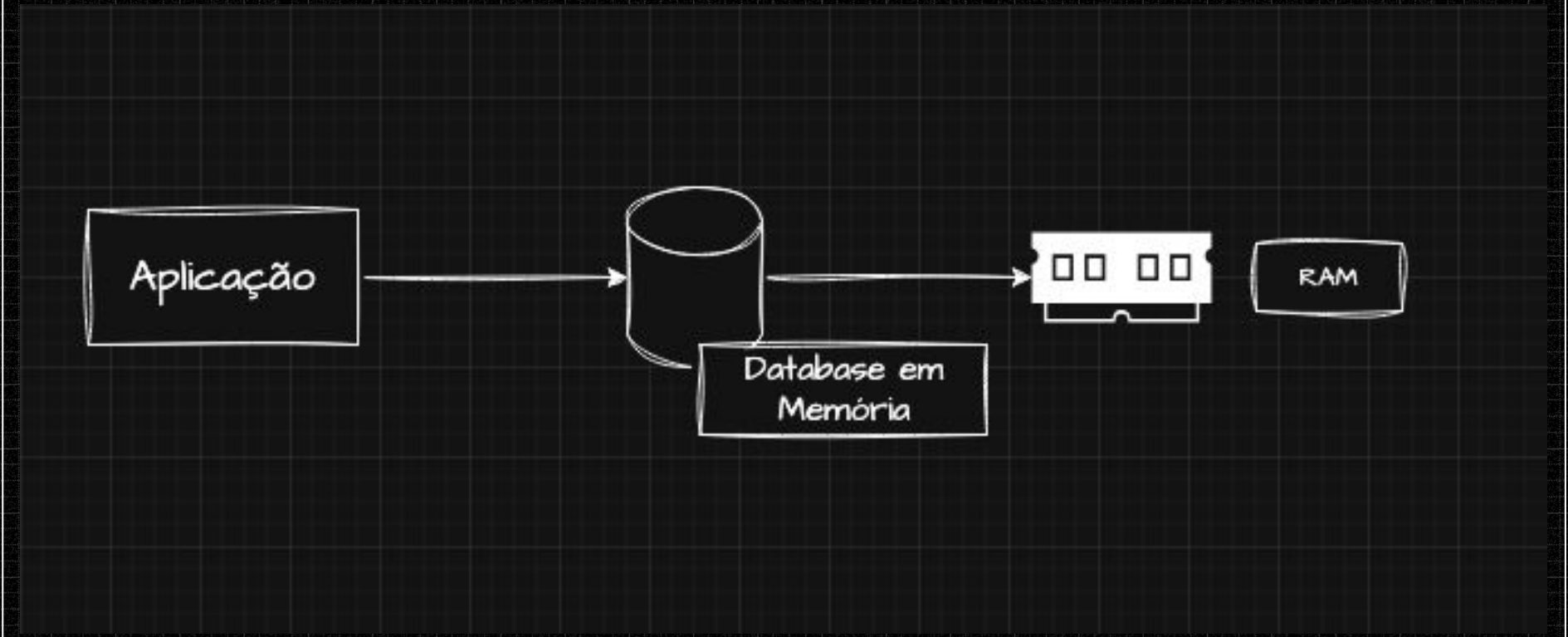
- ACID + Escalabilidade Horizontal
- Sharding e Replicação
- Alto acoplamento no protocolo de consenso
- RAFT/Paxos

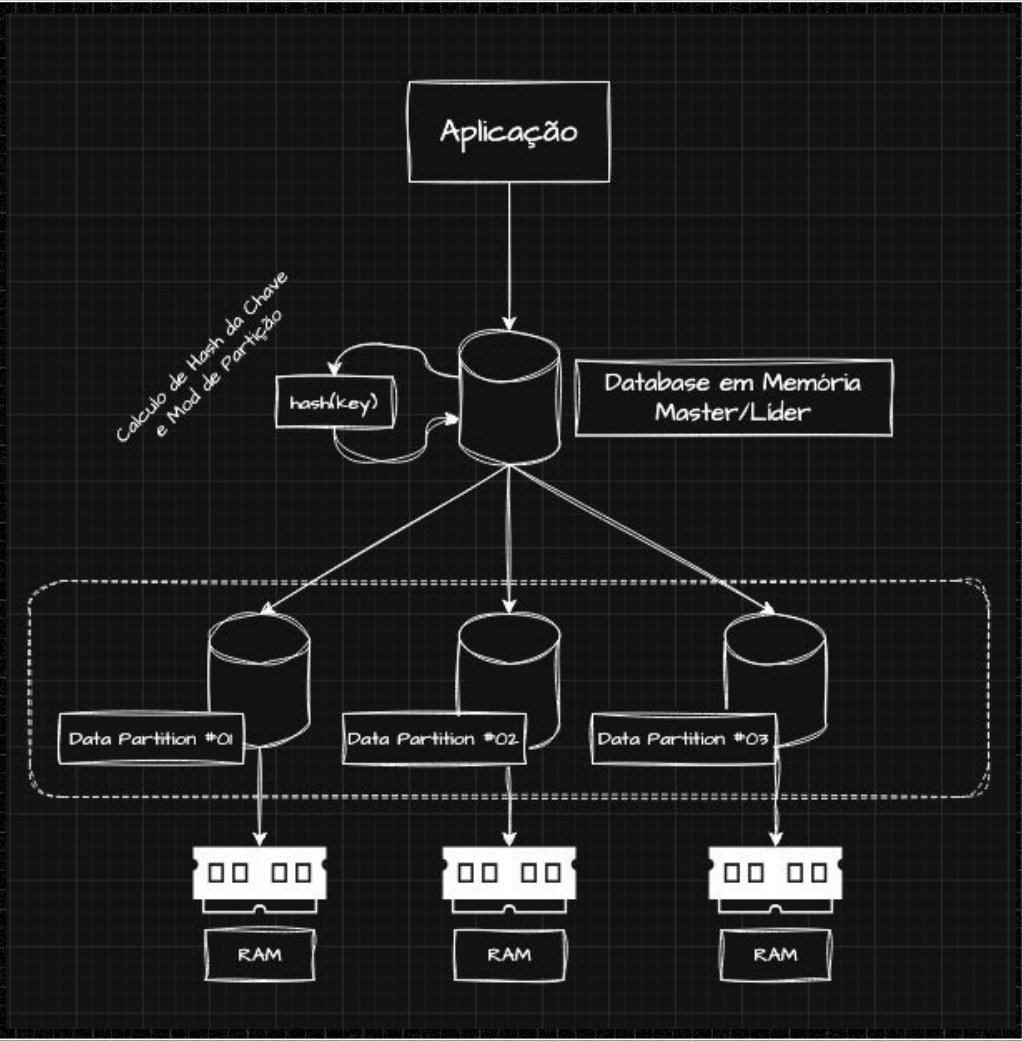
Bancos de Dados NewSQL

- CockroachDB
- Google Cloud Spanner
- MemSQL
- AltiBase
- VoltDB

Bancos de Dados Em Memória (In-Memory)

- Uso de Memória Volátil
- Latência de Nanosegundos em RAM
- Modelo Key-Value
- Modelo Não Estruturado
- Volatilidade e Performance
- Escalabilidade por Hash Consistente
- Camadas de Cache e Read Intensive
- Dados Precisam Ser Reconstituíveis


Aplicação
Database em Memória
RAM

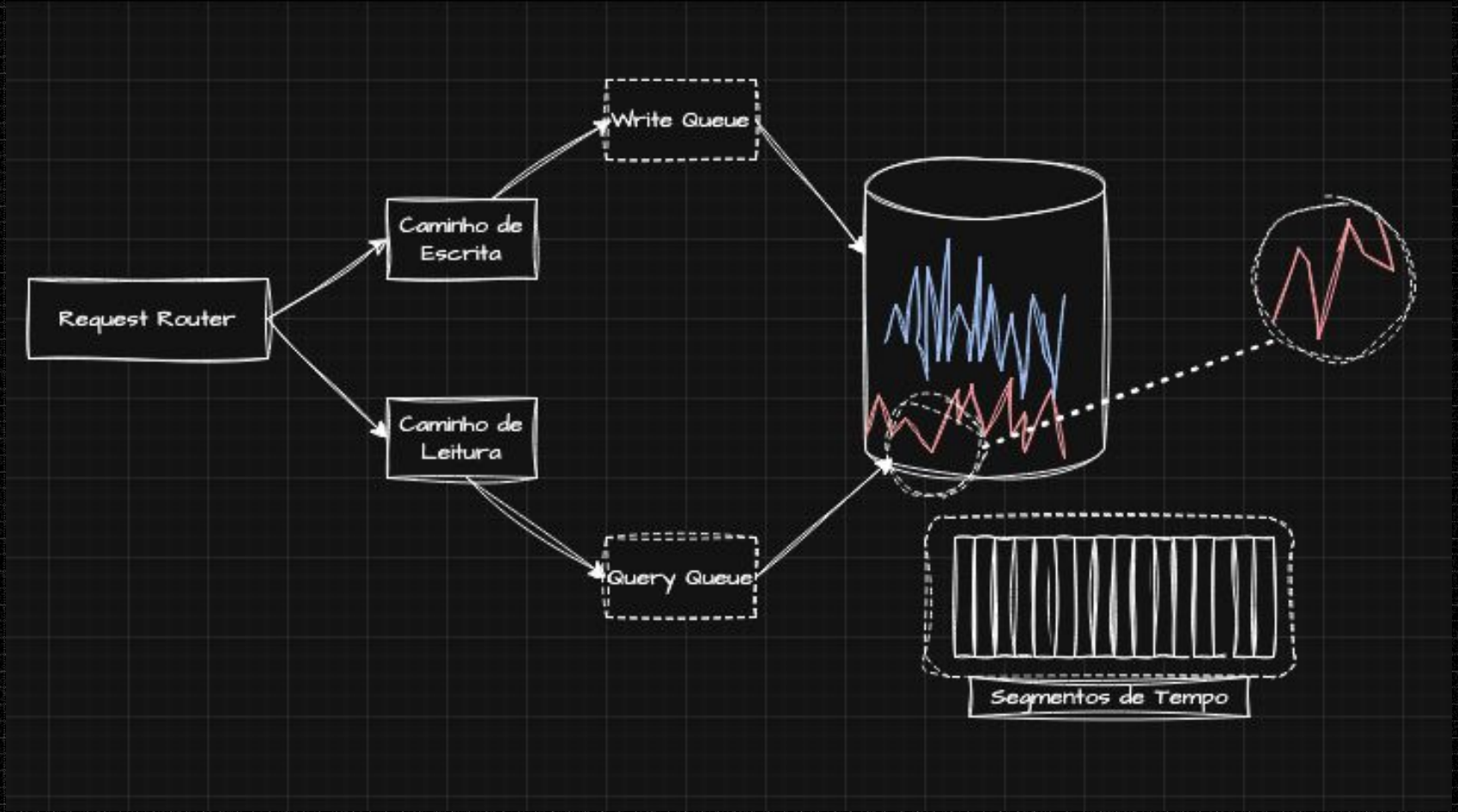

Aplicação
Calculo de Hash da Chave e Mod de Partição
hash(key)
Database em Memória Master/Lider
Data Partition #01
Data Partition #02
Data Partition #03
RAM
RAM
RAM

Bancos de Dados Em Memória (In-Memory)

- Redis
- Memcached
- Valkey
- Apache Ignite
- Aerospike

Bancos de Dados Time-Series (TSDB)

- Time-Series Databases
- Append Only
- Indexação Temporal
- Alta Ingestão
- Buscas Analiticas
- Métricas e Logs
- Expurgo Automático
- Alta volumetria de escrita
- Operações Matemáticas em Vôo


Write Queue
Caminho de Escrita
Request Router
Caminho de Leitura
Query Queue
Segmentos de Tempo

Bancos de Dados Time-Series (TSDB)

- Timescale
- InfluxDB
- Prometheus
- VictoriaMetrics
- Graphite

3. Níveis de Consistência

Consistência
Forte e Eventual

Níveis de Consistência

- Consistencia Forte vs Eventual
- Impacto na Experiência do Cliente
- Confiabilidade vs Performance
- Integridade vs Escalabilidade
- Níveis de Integridade

Consistência Forte

- Linearidade
- Quorum Síncrono
- Paxos, Raft
- Latência de Escrita
- Maior Confiabilidade
- Atomicidade e Transações
- Estado Consistente -> Estado Consistente
- Fluxo de Commit Sincrono
- Operações Críticas e Atômicas

Consistência Forte

- MySQL
- MariaDB
- PostgreSQL
- SQL Server
- Oracle
- Cassandra (Quorum ALL)
- ScyllaDB (Quorum ALL)
- DynamoDB (COnsistency True)

Consistência Eventual

- Replicação Assincrona
- Alta Disponibilidade
- Tolerância a Partições
- Estratégias de Resolução de Conflitos
- Last-Write-Wins, CRDT's
- Alto Throughput e Baixa Latência
- Requer um tempo para refletir em todos os nodes.

Consistência Eventual

- ScyllaDB (Quorum)
- Cassandra (Quorum Baixo)
- DynamoDB
- CouchDB
- MongoDB
- Elasticsearch

4. Modelos de Dados

Definições e Conceitos

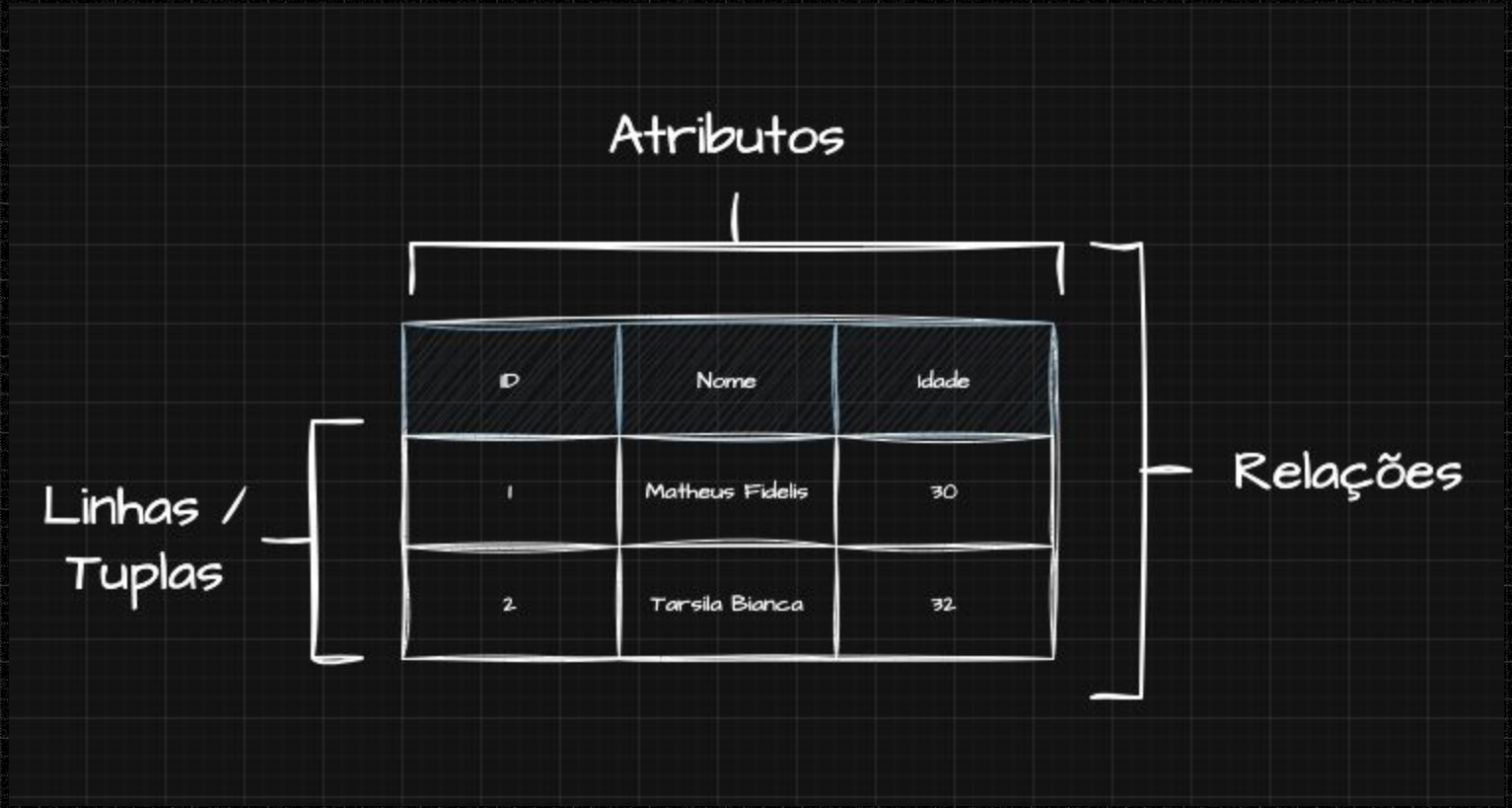
Modelos de Dados

- Tuplas (Row-Oriented)
- Colunares (Column-Oriented)
- Documentos (Document-Oriented)
- Coluna Larga (Wide-Column)
- Chave-Valor (Key-Value)
- Grafos
- Tradeoffs de Escrita e Leitura
- Tradeoffs de Complexidade de Consultas

- Tradeoffs Transacional, Analitico, Agregações, Relacionamentos…

Row-Oriented - Linhas e Colunas

- Tuplas COmpletas
- OLTP
- Caches de Página
- Baixa Latência por Linha
- Agrupam Atributos da Mesma Entidade Fisicamente no mesmo bloco


Atributos
ID
Nome
Idade
1
Matheus Fidelis
30
2
Tarsila Bianca
32
Linhas / Tuplas
Relações

Row-Oriented - Linhas e Colunas

- MySQL
- MariaDB
- PostgreSQL
- SQLServer
- Oracle
- Etc…

Document-Oriented

- Schemas Flexíveis
- JSON/BSON
- Indexação Invertida
- Full-Text Search
- Evoluir o Schema sem Migrações Complexas
- Sem relacionamentos

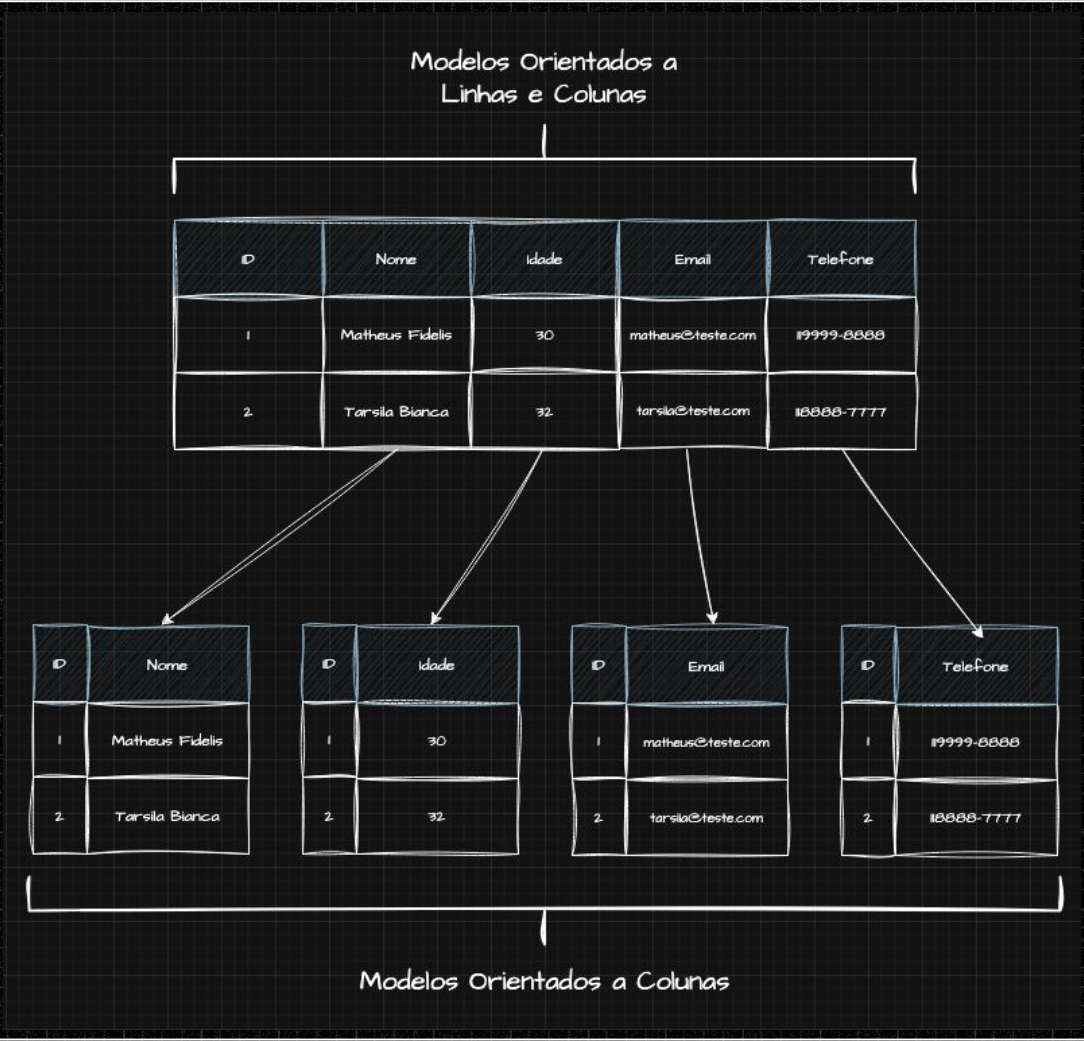
```
{
   "$jsonSchema":{
      "bsonType":"object",
      "required":[
         "nome",
         "preco",
         "categoria"
      ],
      "properties":{
         "_id":{
            "bsonType":"objectId"
         },
         "nome":{
            "bsonType":"string"
         },
         "preco":{
            "bsonType":"decimal"
         },
         "categoria":{
            "bsonType":"object",
            "required":[
               "id",
               "nome"
            ],
            "properties":{
               "id":{
                  "bsonType":"string"
               },
               "nome":{
                  "bsonType":"string"
               }
            }
         }
      }
   }
}
```

Document-Oriented

- MongoDB
- CouchDB
- Coachbase
- Elasticsearch

Column Oriented

- Colunas Contíguas
- Compressão
- Analytics
- Familias de Colunas
- Heavy Scans de Valores Específicos

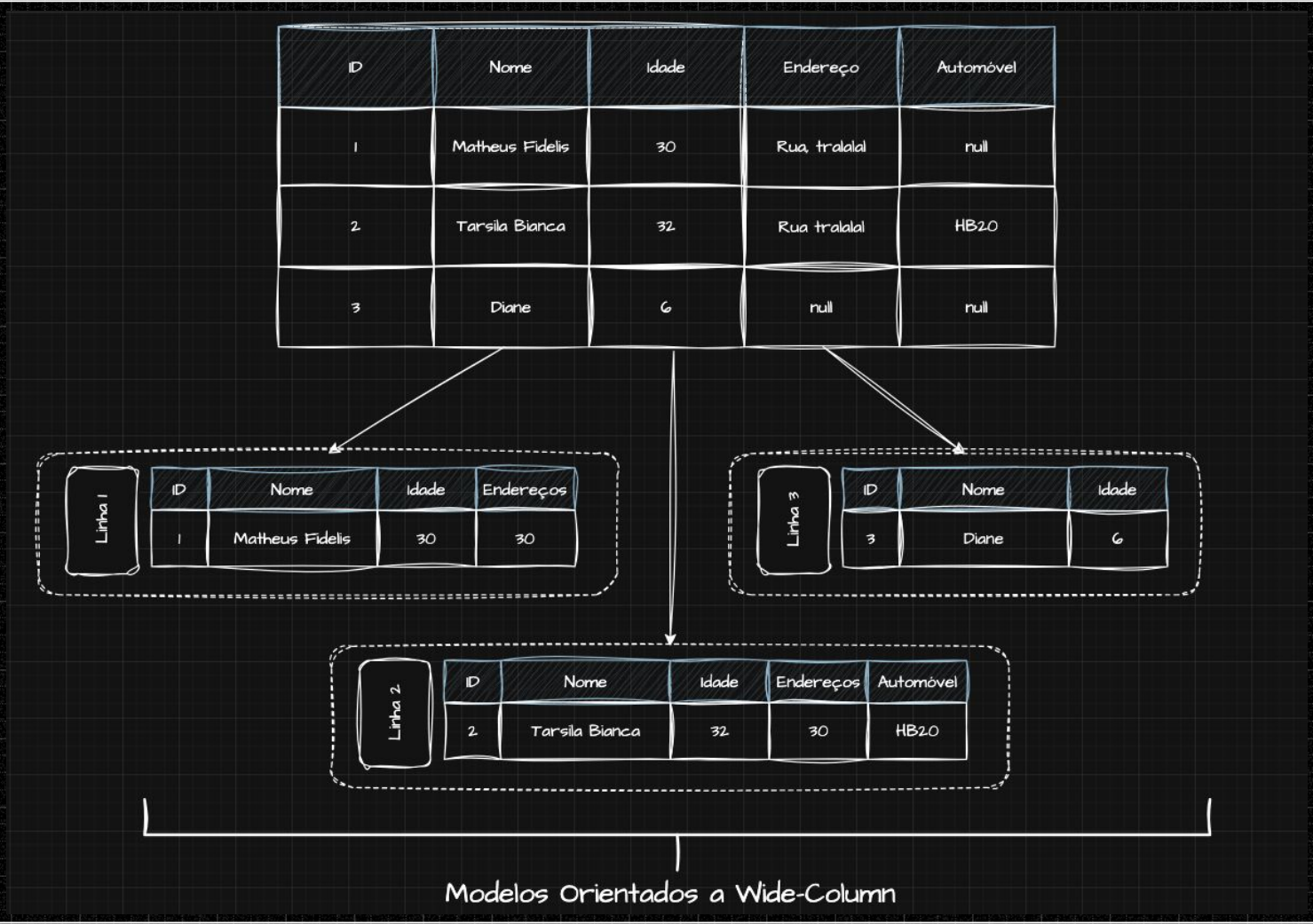

Modelos Orientados a
Linhas e Colunas
ID
Nome
Idade
Email
Telefone
1
Matheus Fidelis
30
matheus@teste.com
19999-8888
2
Tarsila Bianca
32
tarsila@teste.com
18888-7777
ID
Nome
1
Matheus Fidelis
2
Tarsila Bianca
ID
Idade
1
30
2
32
ID
Email
1
matheus@teste.com
2
tarsila@teste.com
ID
Telefone
1
19999-8888
2
18888-7777
Modelos Orientados a Colunas

Column Oriented

- Amazon Redshift
- Google BigQuery
- Snowflake
- MariaDB ColumnStore

Wide-Column Oriented (Coluna Larga)

- Famílias de Coluna Por Linha
- Schema Flexivel por Linha (entidade)
- Cada registro pode conter seu próprio conjunto de colunas.
- Agrupado por Famílias de Colunas

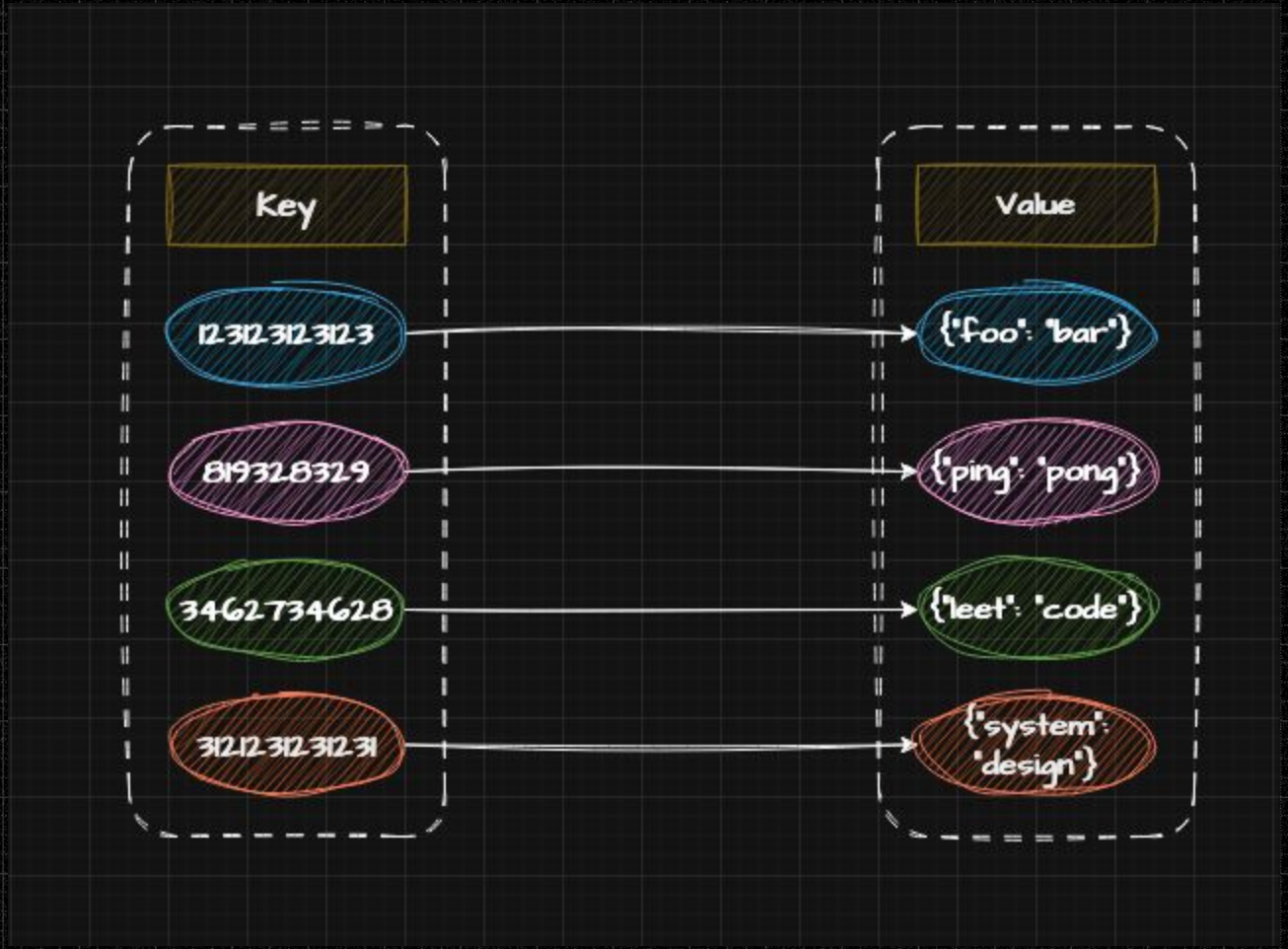

ID
Nome
Idade
Endereço
Automóvel
1
Matheus Fidelis
30
Rua, tralalal
null
2
Tarsila Bianca
32
Rua tralalal
HB20
3
Diane
6
null
null
Linha 1
ID
Nome
Idade
Endereços
1
Matheus Fidelis
30
30
Linha 3
ID
Nome
Idade
3
Diane
6
Linha 2
ID
Nome
Idade
Endereços
Automóvel
2
Tarsila Bianca
32
30
HB20
Modelos Orientados a Wide-Column

Wide-Column Oriented (Coluna Larga)

- CassandraDB
- ScyllaDB
- HBase
- DynamoDB
- CosmosDB

Key-Value (Chave-Valor)

- Lookup Direto
- Hashing de Paridade
- Chave (Identificador Único)
- Valor (Dado Desestruturado)
- Strings, Números, Booleanos, JSON e Blobs
- Extrema Facilidade de Indexação

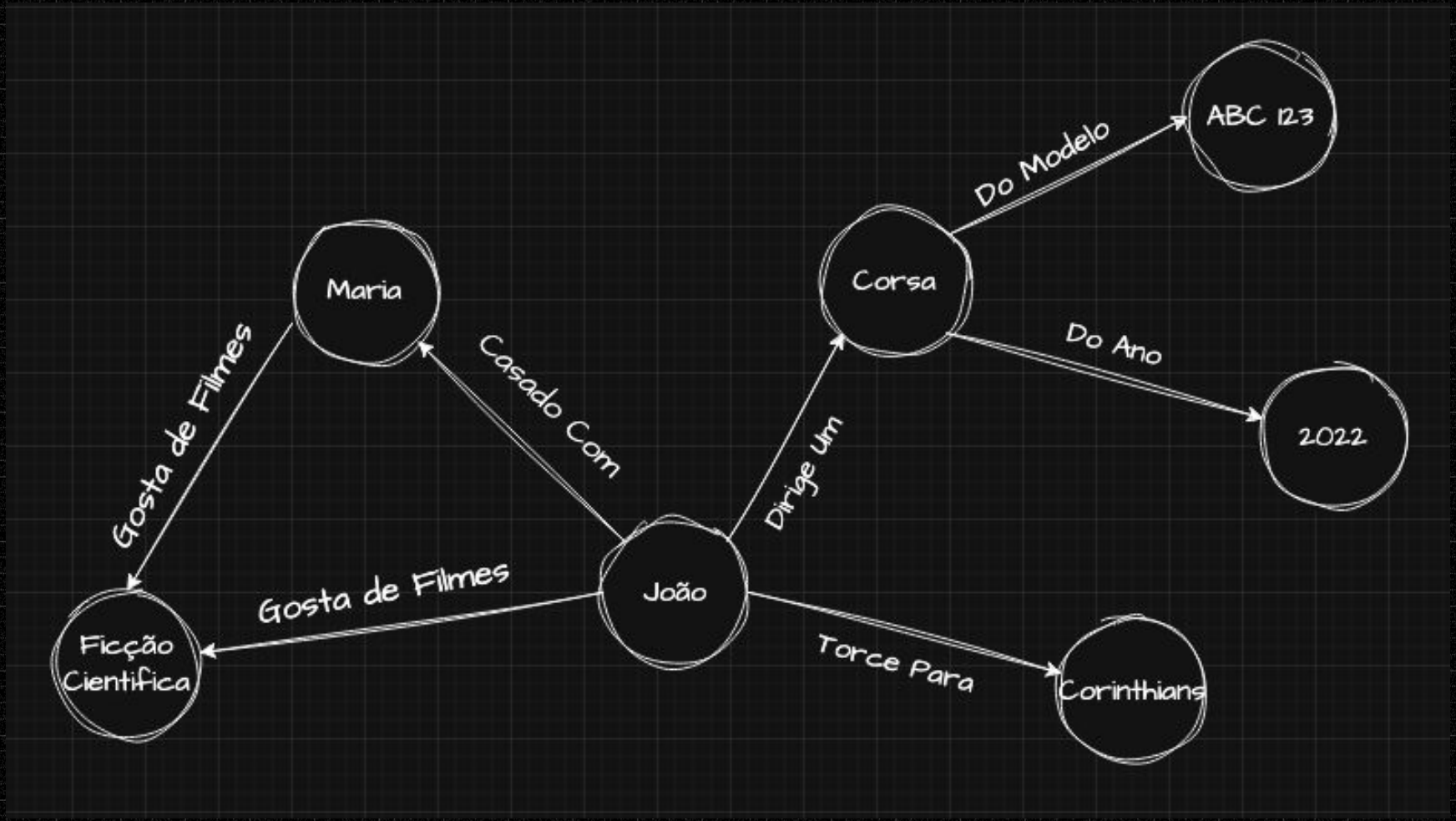

Key
Value
12312312123
{'foo': 'bar'}
8193218329
{'ping': 'pong'}
34627346128
{'leet': 'code'}
3121231231231
{'system': 'design'}

Key-Value (Chave-Valor)

- Redis
- Memcached
- Aerospike
- DynamoDB
- ETCD
- ZooKeeper

Grafos

- Relacionamento entre as entidades é tão importante quanto o próprio dado
- Nodes (entidades)
- Arestas (relacionamentos)
- Propriedades chave-valor chamadas de vértices
- Features de recomendação
- Modelos desestruturados


ABC 123
Do Modelo
Corsa
Do Ano
2022
Maria
Casado Com
Gosta de Filmes
Dirige Um
Gosta de Filmes
João
Ficção Cientifica
Torce Para
Corinthians

Grafos

- Neo4J
- OrientDB
- ArangoDB
- Neptune
- CosmosDB (API Gremlin)

5. Armazenamento e Indexação

Definições e Conceitos

Armazenamento e Indexação

- Forma de Armazenamento e Busca
- Impacto Direto em Desempenho e Flexibilidade
- O mecanismo de armazenamento e indexação é o coração do desempenho de queries
- Page Size, Columnar Index, LSM-Tree, B-Tree, Hashing, Inverted Index

Page Size

- Blocos Organizados e Armazenados em Tamanho Fixo (4 KB, 8 KB…)
- Páginas grandes: reduzir operações de I/O em leituras grandes de objetos fisicamente próximos
  - Aumentam o Custo de Transferência
- Páginas pequenas: menor leitura de dados irrelevantes em consultas pontuais
  - Aumentam as operações de I/O

Page Size

- RDBMS's Padrão
- PosgreSQL
- MySQL/InnoDB
- Orable Databases
- Apache Cassandra

Indexação Colunar

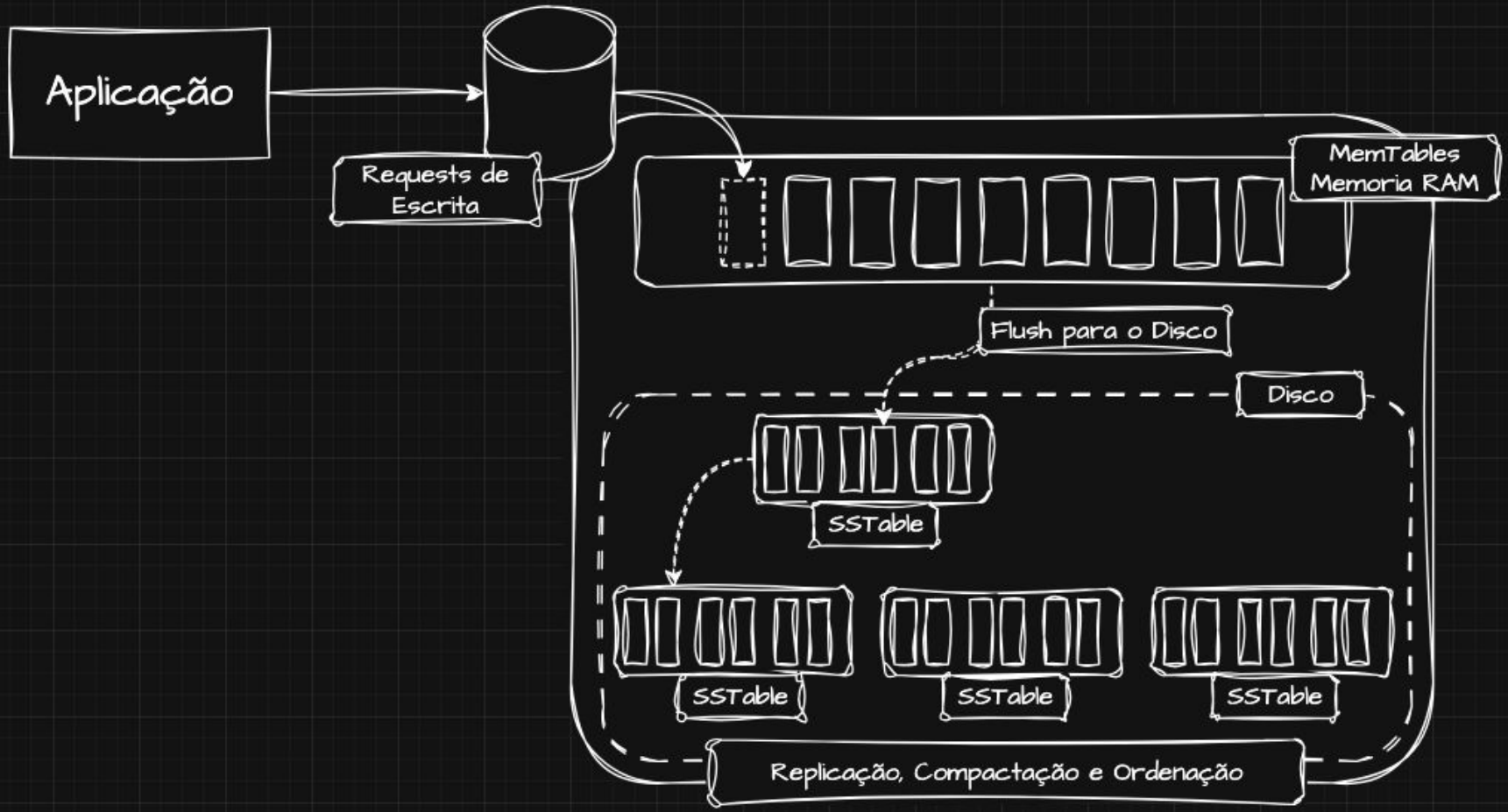
- Segmentos de Colunas Contíguas
- Cada Coluna é escrita em um segmento no sistema de arquivos
- Compressão por Dicionário
- Alta Performance em Agregação
- Reduz I/O com Compressão
- Busca por atributos específicos
- Workloads Analiticos

Indexação Colunar

- Amazon Redshift
- BigQuery
- Snowflake
- DuckDB

LSM-Trees (Log-Structured Merge-Tree)

- Otimizado para Escrita Intensiva
- Baixa latência de confirmação
- Memtable (RAM) -> SSTables (Disco)
- Append-Only, Tombstones e Compactação
- Escrita sequencial ultrarrápida
- Não realiza atualizações in-place


Aplicação
Requests de Escrita
MemTables Memoria RAM
Flush para o Disco
Disco
SSTable
SSTable
SSTable
SSTable
Replicação, Compactação e Ordenação

LSM-Trees (Log-Structured Merge-Tree)

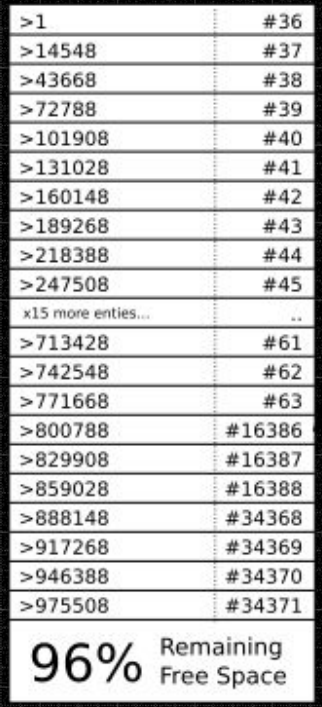
- Cassandra
- ScyllaDB
- Apache HBase
- InfluxDB
- Victoria Metrics
- Apache Lucene
- Elasticsearch

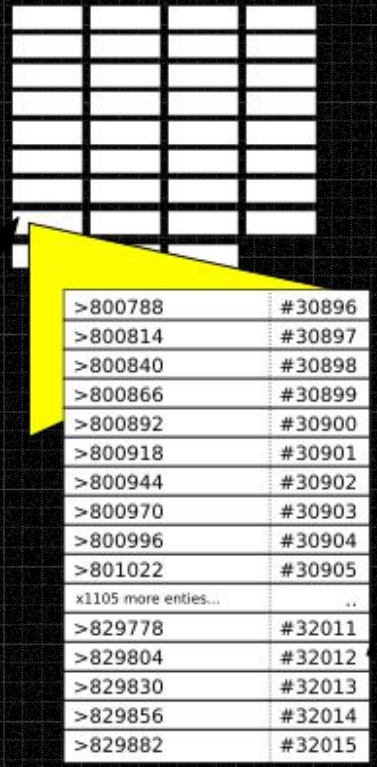
Indexação B-Tree (Árvores B)

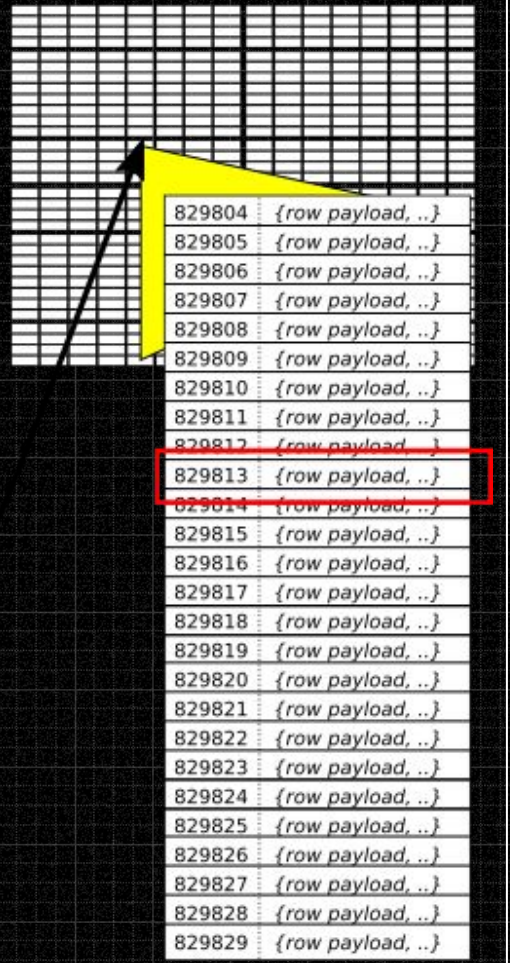
- Arvore Multi-Way Balanceada
- Nós armazenados em blocos de disco
- Busca, Inserção e Remoção O(log n)
- Padrões dos bancos SQL
- Acesso a disco otimizado por profundade mínima
- Buscas de intervalo (range queries)
- Sistema carrega apenas os poucos blocos de disco necessários para percorrer o caminho do nó raiz até o nó onde a chave

```
CREATE TABLE `users` (
  `id` int(11) NOT NULL AUTO_INCREMENT,
  `username` varchar(32) NOT NULL,
  `payload` text,
  PRIMARY KEY (`id`)
) ENGINE=InnoDB
```

```
SELECT * FROM users
WHERE id = 829813
```





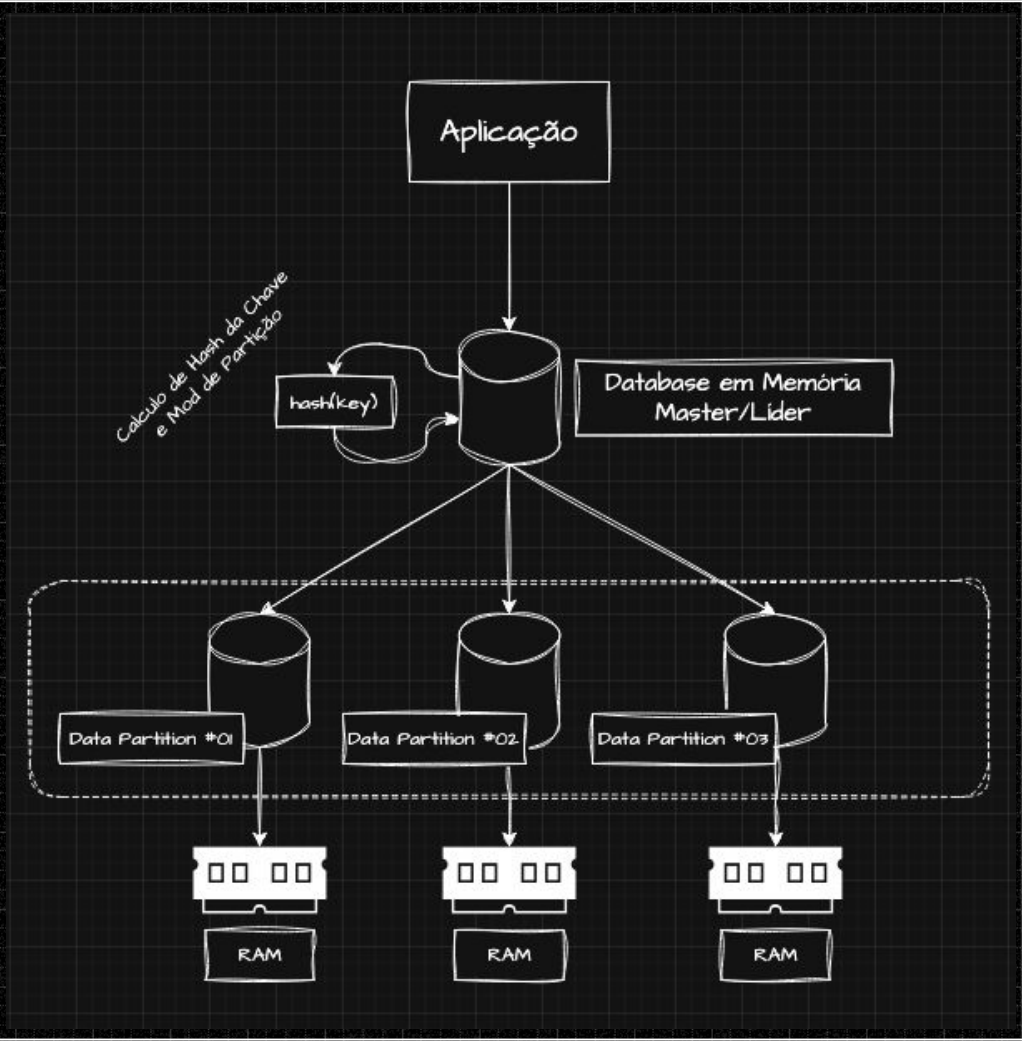


Indexação B-Tree (Árvores B)

- PostgreSQL
- SQL Server
- MySQL
- SQLite
- Lucene

Indexação por Hashing

- Lookup Exato do valor (exact-matches)
- Buscas diretas e instantâneas
- Colisões tratadas por Chaining
- Bucket Hash
- Ideal para chaves exatas

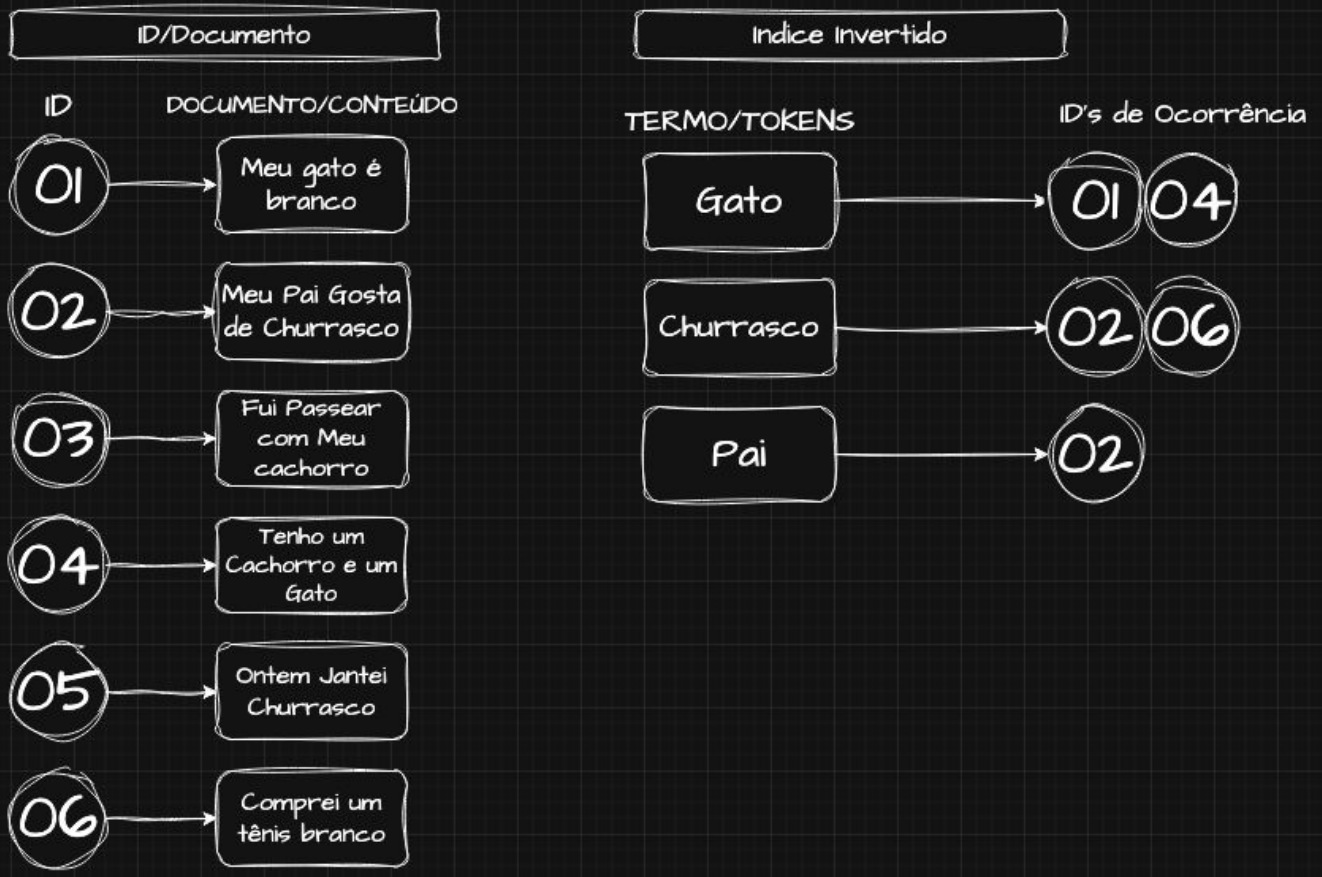

Aplicação
Calculo de Hash da Chave e Mod de Partição
hash(key)
Database em Memória Master/Lider
Data Partition #01
Data Partition #02
Data Partition #03
RAM
RAM
RAM

Indexação por Hashing

- PostgreSQL (Indice Hash)
- MySQL (Memory Engine)
- Oracle (Hash Cluster)
- Redis
- Memcached
- Valkey
- DynamoDB
- Cassandra

Indices Invertidos

- Tokens de Padrões
- Full-text search
- Papeia o "Termo" para os documentos em que o mesmo é encontrado
- Buscas de padrões de textos em larga escala
- Permitindo buscas em textos e valores longos por meio de termos simples


ID/Documento
ID
DOCUMENTO/CONTEÚDO
01
Meu gato é branco
02
Meu Pai Gosta de Churrasco
03
Fui Passear com Meu cachorro
04
Tenho um Cachorro e um Gato
05
Ontem Jantei Churrasco
06
Comprei um tênis branco
Indice Invertido
TERMO/TOKENS
ID's de Ocorrência
Gato
01
04
Churrasco
02
06
Pai
02

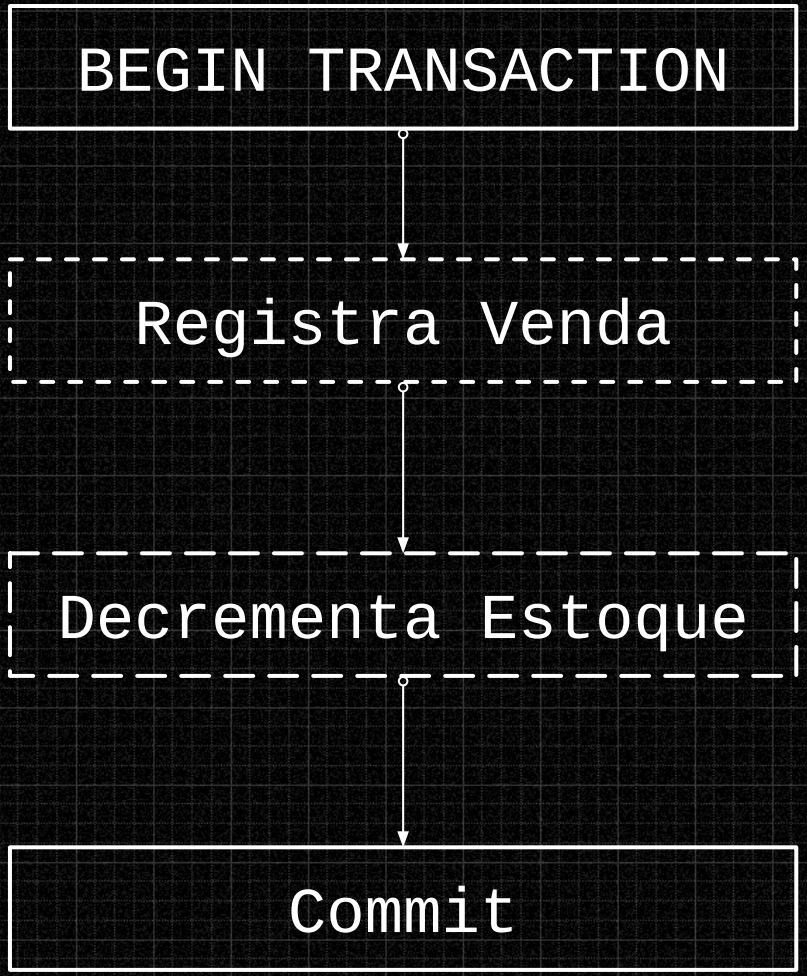
Indices Invertidos

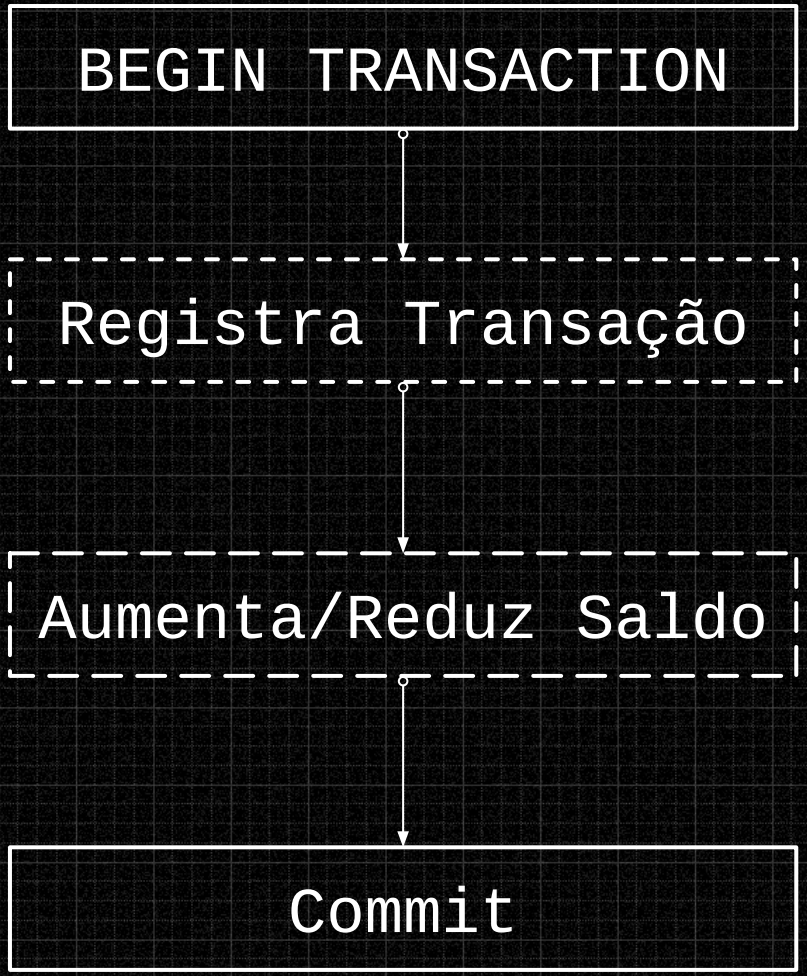
- Elasticsearch
- Apache Lucene
- MongoDB

6. Arquiteturas e Cases

Cenário Transacional

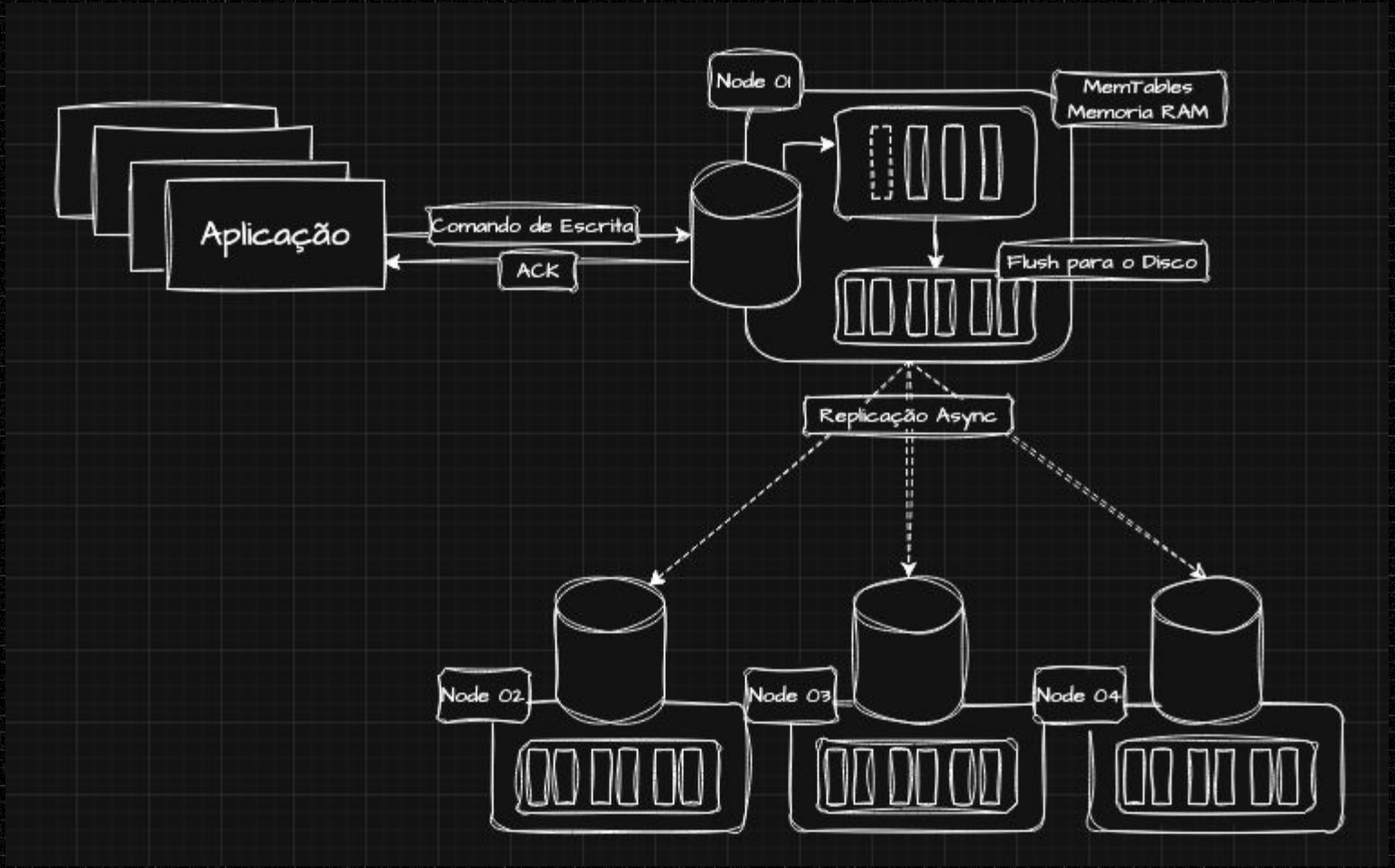
- ACID
- Atomicidade, integridade e isolamento
- Escala vertical vs NewSQL horizontal
- Consistência e isolamentos fortes
- Complemento de cache para leitura


BEGIN TRANSACTION
Registra Venda
Decrementa Estoque
Commit


BEGIN TRANSACTION
Registra Transação
Aumenta/Reduz Saldo
Commit

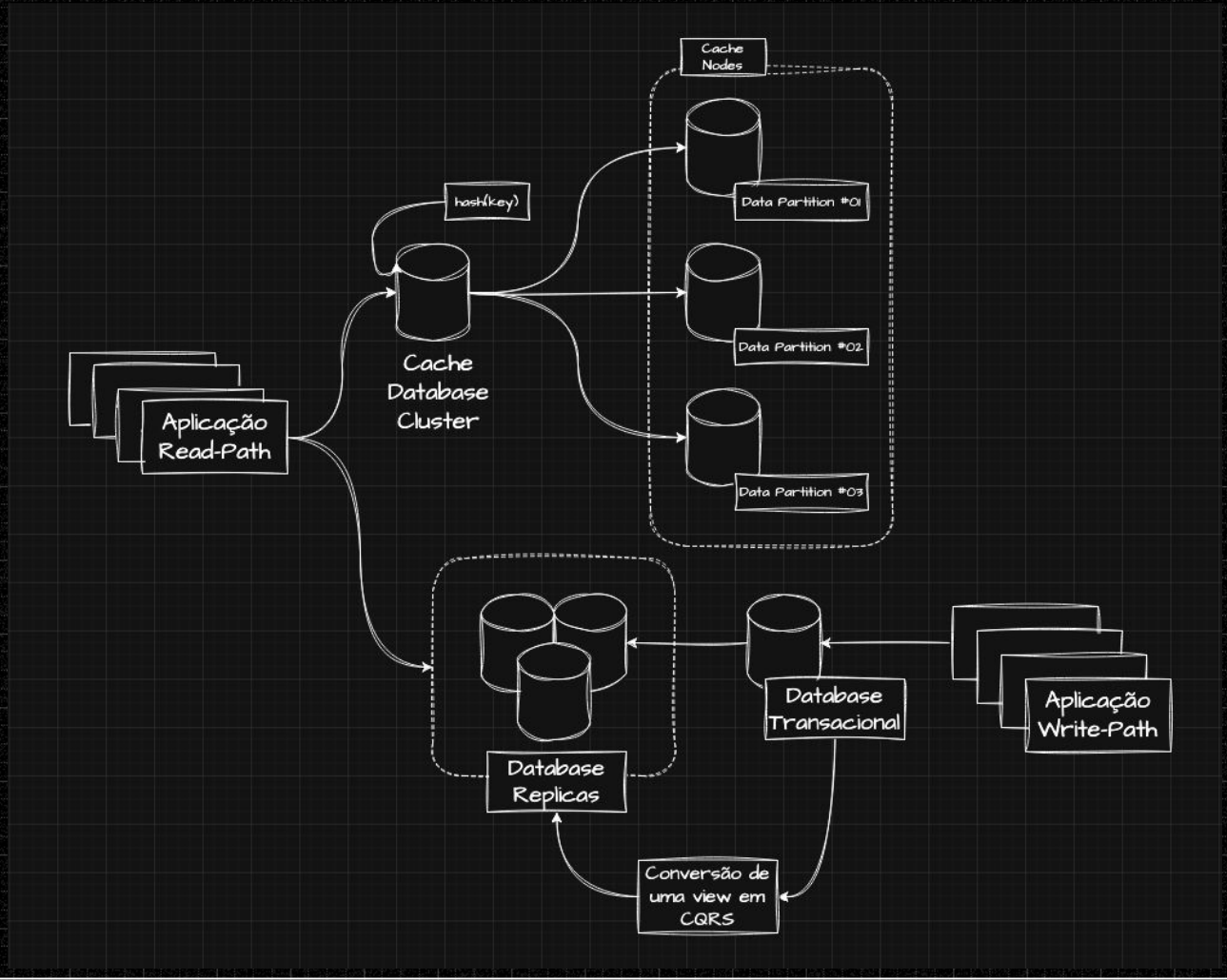
Cenário Write-Intensive

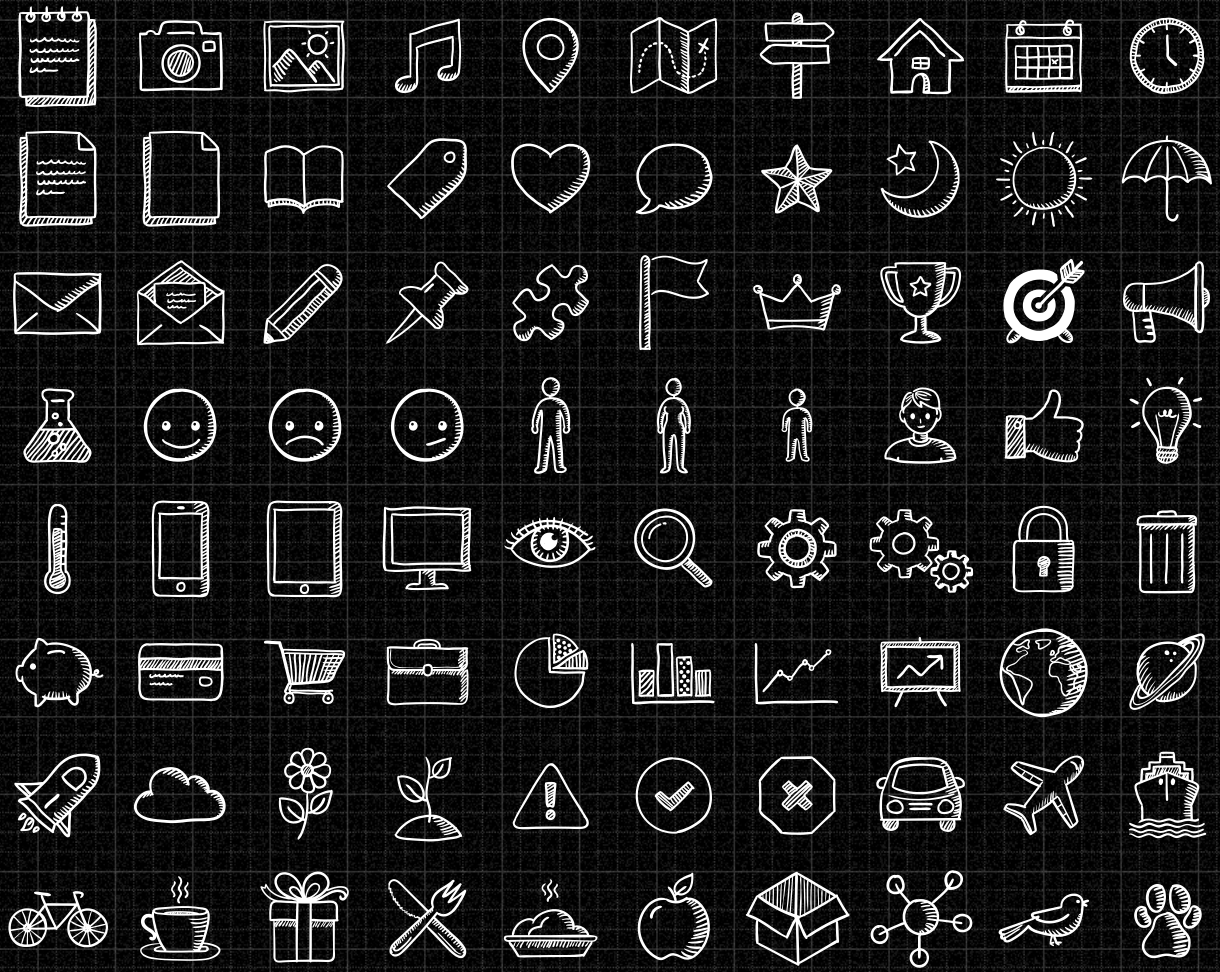
- Escrita > Leitura
- Append Only
- Replicação Assíncrona
- Consistencia Eventual
- LSM-Trees
- Caminho de escrita em memória
- Disco sem bloqueio


Node 01
MemTables
Memoria RAM
Aplicação
Comando de Escrita
ACK
Flush para o Disco
Replicação Async
Node 02
Node 03
Node 04

Cenário Read-Intensive

- Replicas de Leitura
- Camadas de Cache
- CQRS para otimização
- Leitura distribuída em várias replicas


Cache Nodes
Data Partition #01
Data Partition #02
Data Partition #03
hash(key)
Cache Database Cluster
Aplicação Read-Path
Database Replicas
Database Transacional
Aplicação Write-Path
Conversão de uma view em CQRS



SlidesCarnival icons are editable shapes.

This means that you can:

- Resize them without losing quality.
- Change fill color and opacity.

Isn’t that nice? :)

Examples: