Database World

Big 3 Cloud → Big 3 Cloud

Prop. → Proprietry

SaaS → SaaS

OSS → Open Source
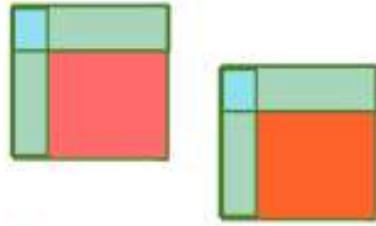
EXAMPES	CONCEPTS
oss Postgres	• WAL & Journaling
Prop Oracle	• Buffer Pool
oss MySQL	• Mem. $\rightleftarrows$ Disk
☁ AWS RDS	• ACID, BASE
SaaS Yugabyte	• CAP & PACLEC
SaaS TiDB	• Distributed databases
	• Txn & Isolation levels
	• MVCC
	• Query execution
	• Master-Slave
	• Master-Master (Active-Active)
	• 2PC vs Quorum → Paxos, RAFT
	• Storage engines
	⤷ MySQL (Innodb, MyISAM...)
	⤷ RocksDB
	• Clock Synchronization
	• SQL flavour
	• Materialized Views, Views
	• Window fn.
	• Connection Pool
	• Btree vs Log Structured Merge Tree
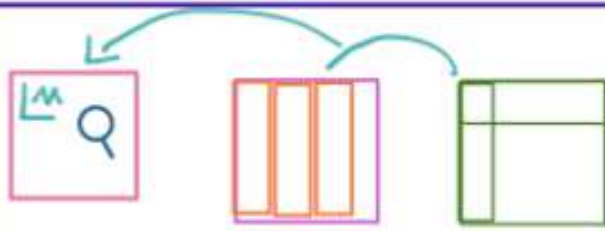
Columnar Storage

(for OLAP & Data Warehouse workloads)
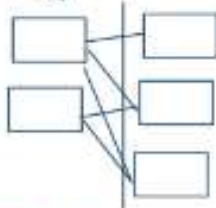
EXAMPLES

- OSS Apache Cassandra (datastax)
- OSS Apache HBase
- ☁ Google BigTable
- SaaS ☁ Snowflake
- OSS Clickhouse
- OSS Apache Hive
- OSS Apache Impala
- OSS Presto (Query engine)
- ☁ Amazon Redshift
- ☁ Google Bigquery
- ☁ Azure Synapse
- OSS Apache Pinot
- OSS Apache Druid

CONCEPTS

- Tablets
- Row vs Coloumn store
- Vectorized execution
- ○ Costs
- ○ OLAP vs OLTP
- ○ Vectorize Query execution
- ○ SIMD H/w optimization
- ○ complex joins
  - ↳ ‖ SORT-merge
  - ↳ ‖ hash
  - ↳ multiway / worstcase
- ○ User defined fns.

Object Store

EXAMPES	CONCEPTS
☁ AWS S3	• S3 API
☁ Google Cloud Storage	• Metadata Service
☁ Azure Blob Storage	• Consistency & Redundancy
SaaS Pure Storage	
OSS Ceph (Openstack)	
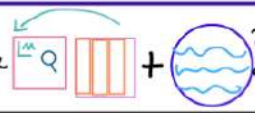
Key-Value DB

EXAMPES	CONCEPTS
OSS Redis	• Caching Strategies (Aside cash, write through etc)
OSS Memcached	• Disk persistance
☁ AWS ...	• SQL vs NoSQL Query (memcached) (redis)
SaaS TiKV	• Redis Internals
	• Copy on write
	• Cache vs In-memory datagrid

Document DB

EXAMPES	CONCEPTS
SaaS MongoDB	o ACID, Txn, JOINS later support
☁ AWS Document DB	o mmap → Wired tiger (mongo)

Graph Database

EXAMPES	CONCEPTS
oss Neo4j Amazon Neptune	• Native Graph Processing & Storage • Transaction Handling

Vector DB

EXAMPES	CONCEPTS

SaaS Pine cone
SaaS Weaviate
oss Chroma
oss Traditional "DB with Vector
   ↳ Postgres
   ↳ Cassandra
   ↳ Clickhouse
	• Vectors
	○ Embeddings
	• Distance metrics
	○ KNN vs ANN
	○ Index
	   ↳ Flat Index
	   ↳ Inverted File Index

Time-Series DB

EXAMPES	CONCEPTS
oss Graphite oss OpenTSDB SaaS InfluxDB (Push) oss Prometheus (Pull)	• Sampling Rate • Querying

Data Lakehouse

EXAMPES	CONCEPTS
oss dremio oss Deltalake	• File formats: Parquet, Arrow etc. • Integration with OLAP, BI etc. • Data governance