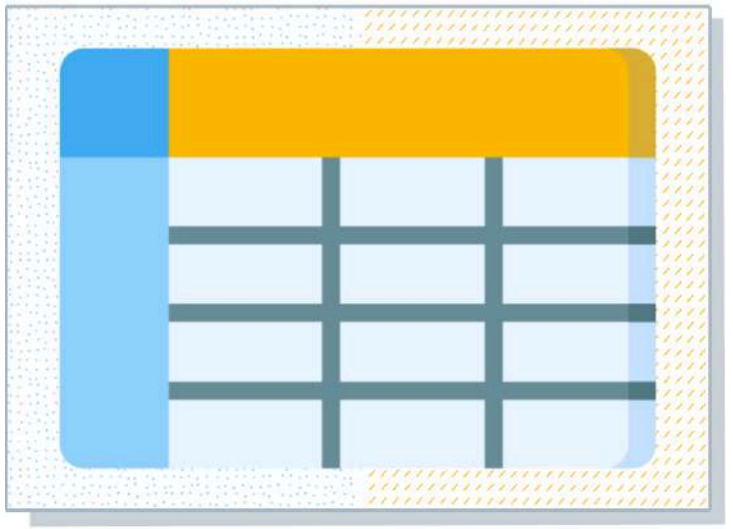
Evolution of Structured OLAP databases

Structured Data: Well organized data (with proper structure)

- **Database**: Stored in relational databases



Two Major Workloads:

- **OLTP**: Fast, real-time transactions
- **OLAP**: Large-scale analytics, reports and insights

Journey: Let's understand the evolution of OLAP databases from on-premises to cloud

OLAP Evolution – The Journey

OLAP: Applications analyzing massive volumes of data aggregated from different sources

Phase 0 - Analytics on OLTP (The beginning)

Phase 1 – Analytics on Read Replicas

Phase 2 – On-Prem Data Warehouses (RDW)

Phase 3 – Cloud Data Warehouses

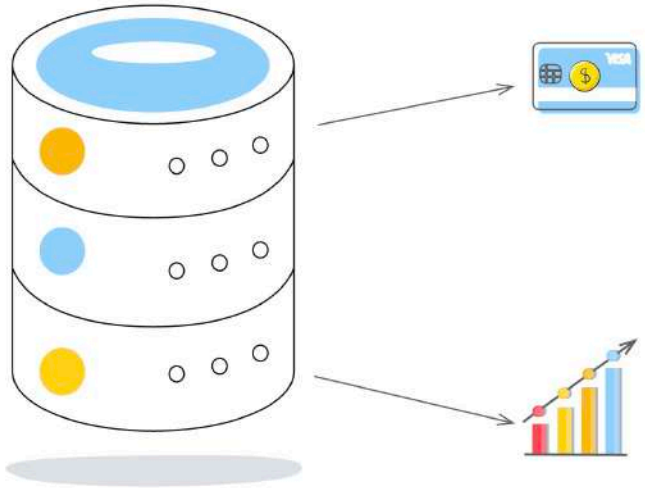
Phase 4 – Cloud Serverless Analytics

OLAP – Phase 0: Analytics on OLTP

OLAP on OLTP: Run queries directly on live production databases

Challenges: A couple of highlights

- **Performance Impact**: Analytics queries slowed down real-time customer transactions
- **Difficulty in Combining Data** - How can you combine data from multiple OLTP database for analytics?
  - *For example*: Risk Analysis needs Savings, Loans, Credit Cards and other data



Result: Realization that analytics needs a dedicated system

OLAP – Phase 1: Analytics on Read Replicas

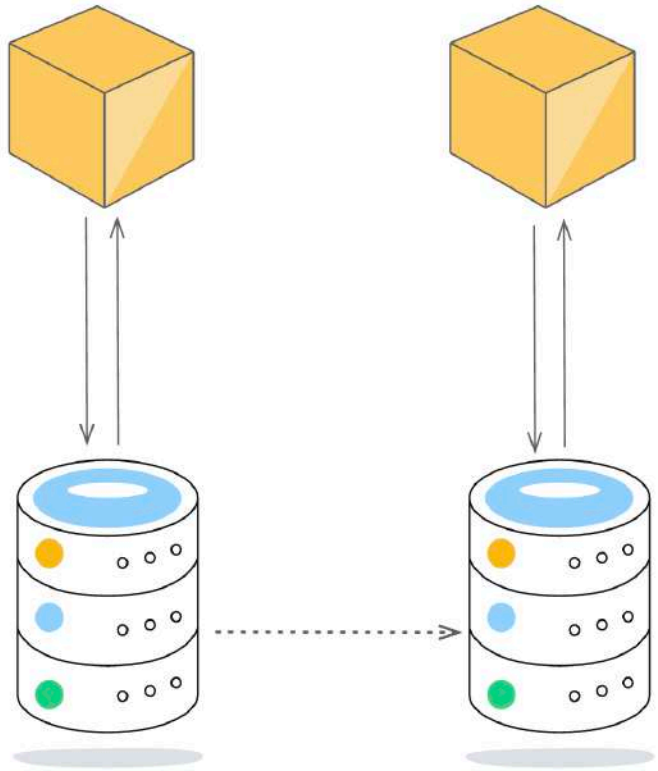
Concept: Create read-only copies of the primary database

- **Primary server**: Handles all **Writes**
- **Read Replicas**: Handle all **Reads** (SELECT queries)

Benefit: Primary can focus solely on writes, providing better performance

Limitation: Doesn't solve write bottlenecks or overall database size limitations

Use Cases: Heavy reporting, analytics



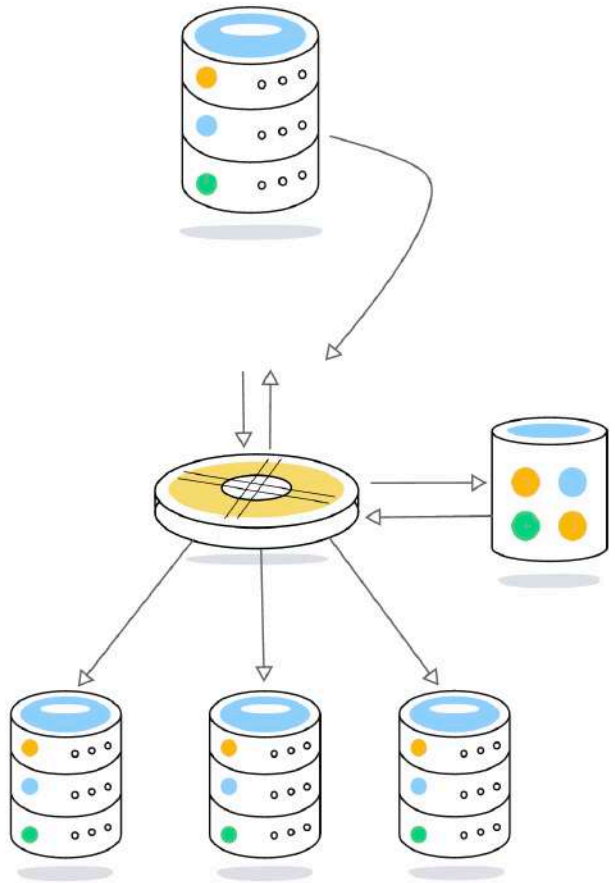
OLAP – Phase 2: On-Prem Data Warehouses (RDW)

The Pioneer: Teradata was among first popular Relational Data Warehouses

- **Capability**: Store huge volumes of data and run complex analytical queries
- **Value**: Helped businesses transform raw data into high-value insights

Complexity: Required expert management

- **Deployment**: Required massive physical servers in on-premises data centers
- **Slow Scaling**: Installing new servers took weeks
- **High CapEx**: Huge upfront investment



OLAP – Phase 3: Cloud RDW

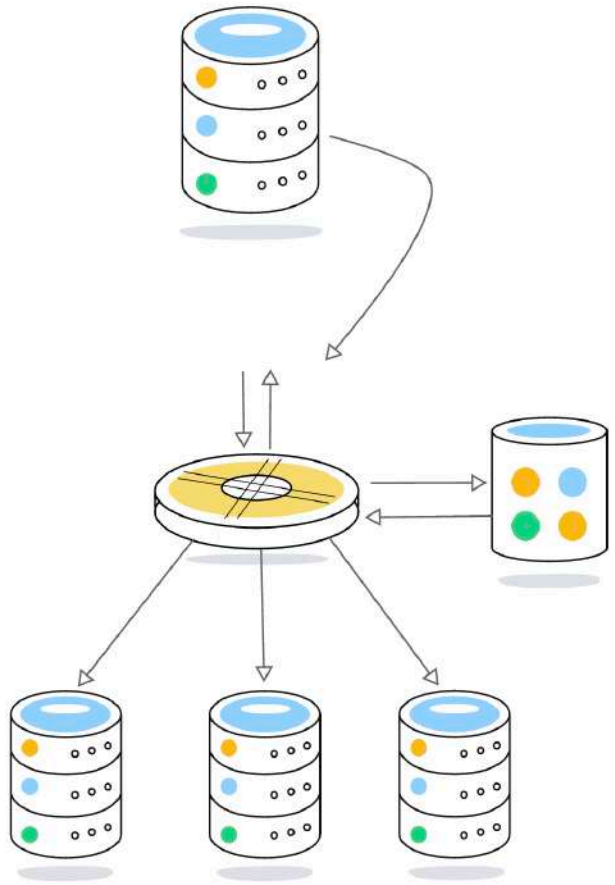
Cloud RDW: RDW moves to the cloud

- **Managed**: Cloud handles infrastructure
- **Easy Provisioning**: Choose cluster size
- **Manage Cluster**: Add nodes as needed (Manual configuration)

Limitation: Compute & Storage tied together

- If you want to store more data -> Add nodes (Manual)
- If you want to execute lots of complex queries -> Add nodes (Manual)

Managed Services: Amazon Redshift, Azure Synapse, ..



OLAP – Phase 4: Cloud Serverless

Serverless: No clusters or servers to manage

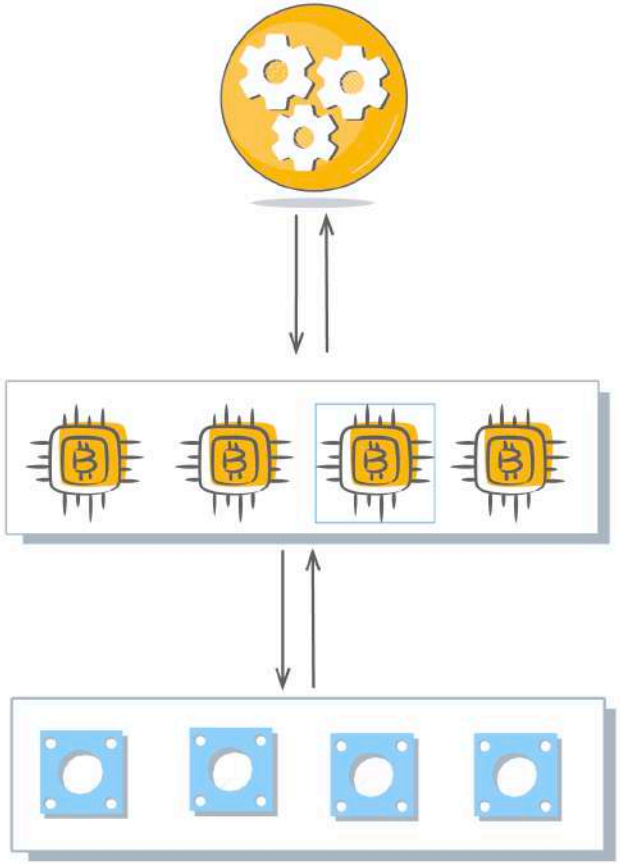
- **Independent Scaling**: Storage and compute scale independently and automatically

Billing: Pay only for Data you store and Queries you run

Focus: Pure Analytics - No Infra Management

- Store as much data as you want and query as much as you want without managing servers!

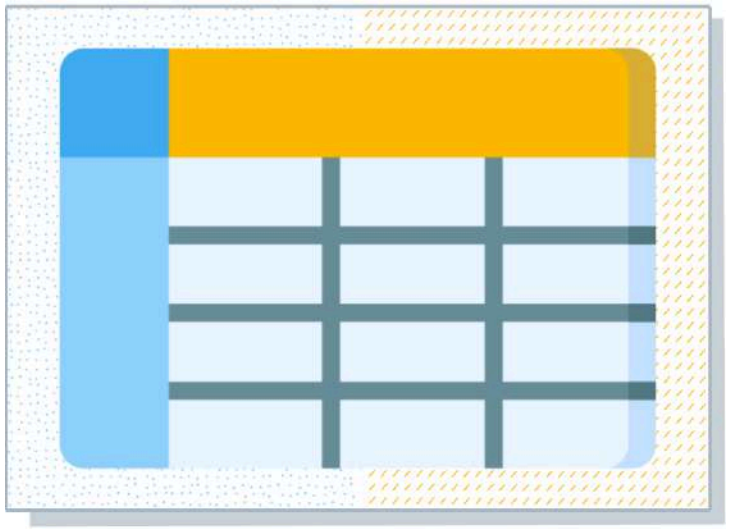
Cloud Services: Google BigQuery, Amazon Redshift Serverless, Azure Synapse Serverless



Evolution of Structured OLAP databases - Summary

Structured Data: Well organized data

- **Database**: Stored in relational databases
- **OLAP**: Large-scale analytics, reports and insights



Journey: We understood the evolution of OLAP databases from on-premises to cloud

- **Phase 0** - Analytics on OLTP (The beginning)
- **Phase 1** – Analytics on Read Replicas
- **Phase 2** – On-Prem Data Warehouses (RDW)
- **Phase 3** – Cloud Data Warehouses
- **Phase 4** – Cloud Serverless Analytics