DATA ARCHITECTURE CHEAT SHEET

by Sergey Gromov

Your Data Guy

DATA ARCHITECTURE TYPES

LAKEHOUSE ARCHITECTURE

- Combines the features of a data lake and a data warehouse.
- KPIs: Query performance, data freshness, data quality.
- Products: Delta Lake, Apache Iceberg, Databricks

DATA LAKE ARCHITECTURE

- Data is stored in its raw, untransformed form in a centralized repository.
- KPIs: Data ingestion rate, data retrieval speed, data discovery capabilities.
- Products: Apache Hadoop, Amazon S3, Apache Spark

DISTRIBUTED ARCHITECTURE

- Data is stored and processed across multiple distributed nodes.
- KPIs: Scalability, fault tolerance, latency.
- Products: Distributed databases (e.g., Apache Cassandra, Amazon DynamoDB)
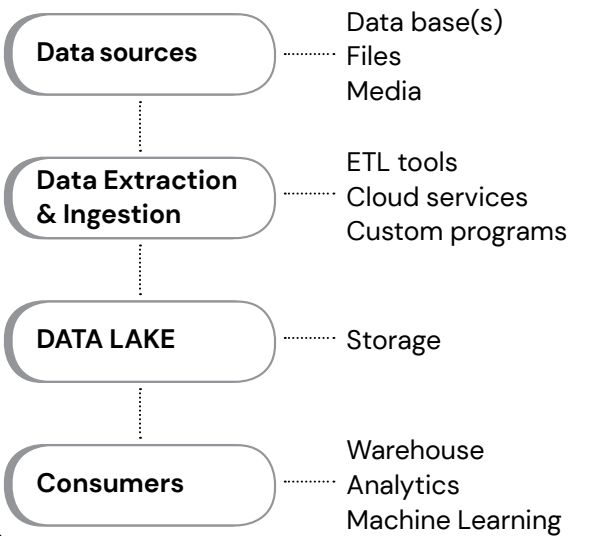
CENTRALIZED DATA ARCHITECTURE

- Data is stored and processed in a single central location.
- KPIs: Latency, throughput, data availability.
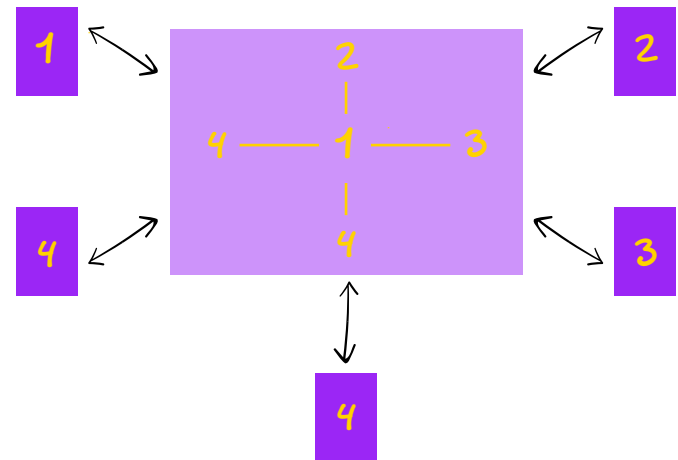- Products: Traditional relational databases (e.g., Oracle, Microsoft SQL Server)

OTHER ARCHITECTURE TYPES

- Examples: Data fabric architecture, federated data architecture, hybrid cloud architecture, event-driven architecture.
- Each has its own unique characteristics, pros, and cons.
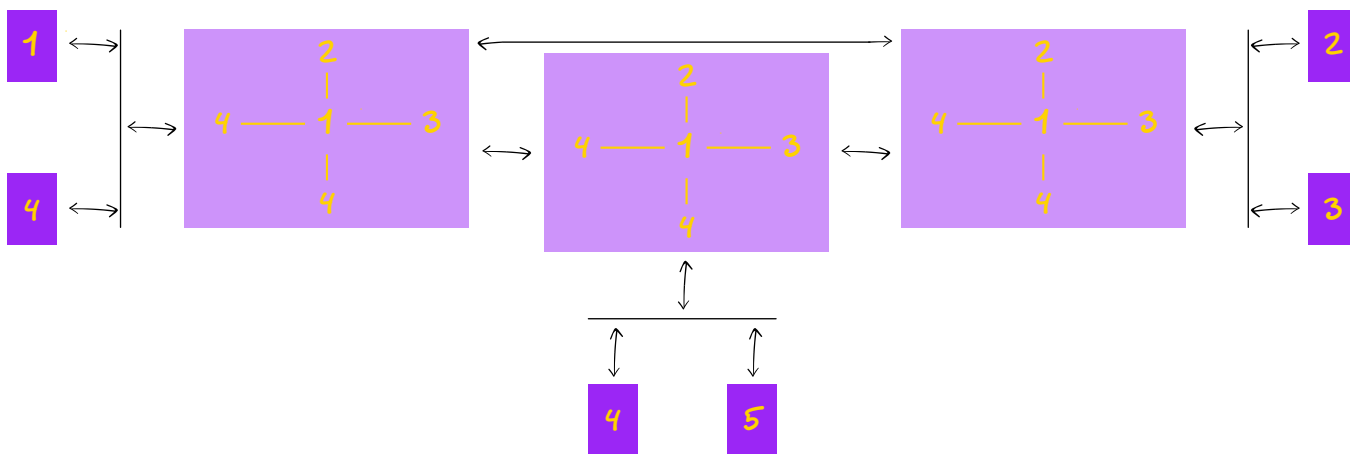- KPIs and products depend on the specific architecture type chosen

DATA LAKE ARCHITECTURE

- Data sources ········ Data base(s) / Files / Media
- Data Extraction & Ingestion ········ ETL tools / Cloud services / Custom programs
- DATA LAKE ········ Storage
- Consumers ········ Warehouse / Analytics / Machine Learning
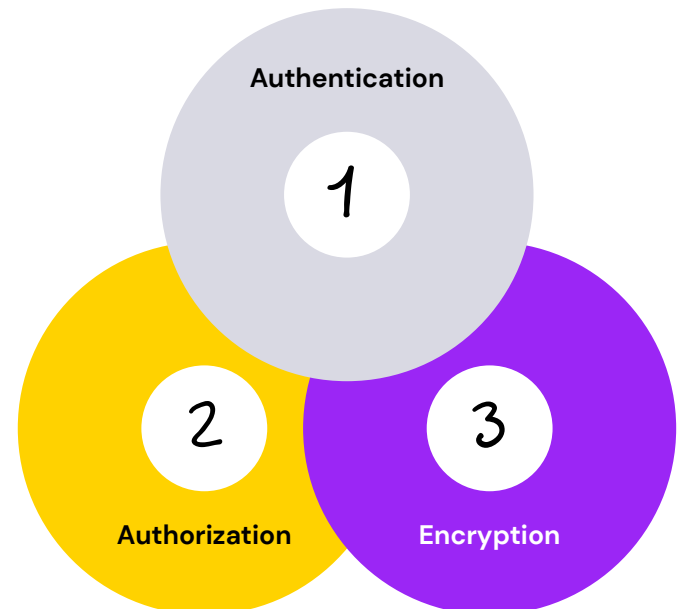
CENTRALIZED DATA ARCHITECTURE

DISTRIBUTED DATA ARCHITECTURE

DATA PRIVACY AND SECURITY

Verifying the identity of users accessing the data system.
KPI: Authentication success rate.
Products: LDAP (Lightweight Directory Access Protocol), Active Directory

- Authentication — 1
- Authorization — 2
- Encryption — 3

Controlling access to data based on user roles and permissions.
KPI: Authorization accuracy.
Products: Role-based access control (RBAC), attribute-based access control (ABAC)

Protecting data by transforming it into an unreadable form.
KPI: Encryption strength (e.g., AES-256).
Products: SSL/TLS, data encryption at rest (e.g., AWS KMS, Azure Key Vault)

DATA TRANSFORMATION

Batch Integration

Data integration through periodic batch processing

ETL Integration

Extracting data from multiple sources, transforming it, and loading it into a target system

ELT Integration

Extracting data from multiple sources, loading it into a target system, and performing transformations within the target system

TEAM AND PERFORMANCE EVALUATION

Team Structure

- ✓ Data Architects
- ✓ Data Engineers
- ✓ Data Analysts
- ✓ Data Scientists
- ✓ Database Administrators

! *Team roles may vary based on the organization's size and complexity*

Performance Evaluation

- Evaluate the team based on project delivery timelines, quality of data solutions, adherence to data governance standards, and feedback from stakeholders.
- Measure individual performance using metrics like code quality, documentation, collaboration, and technical skills