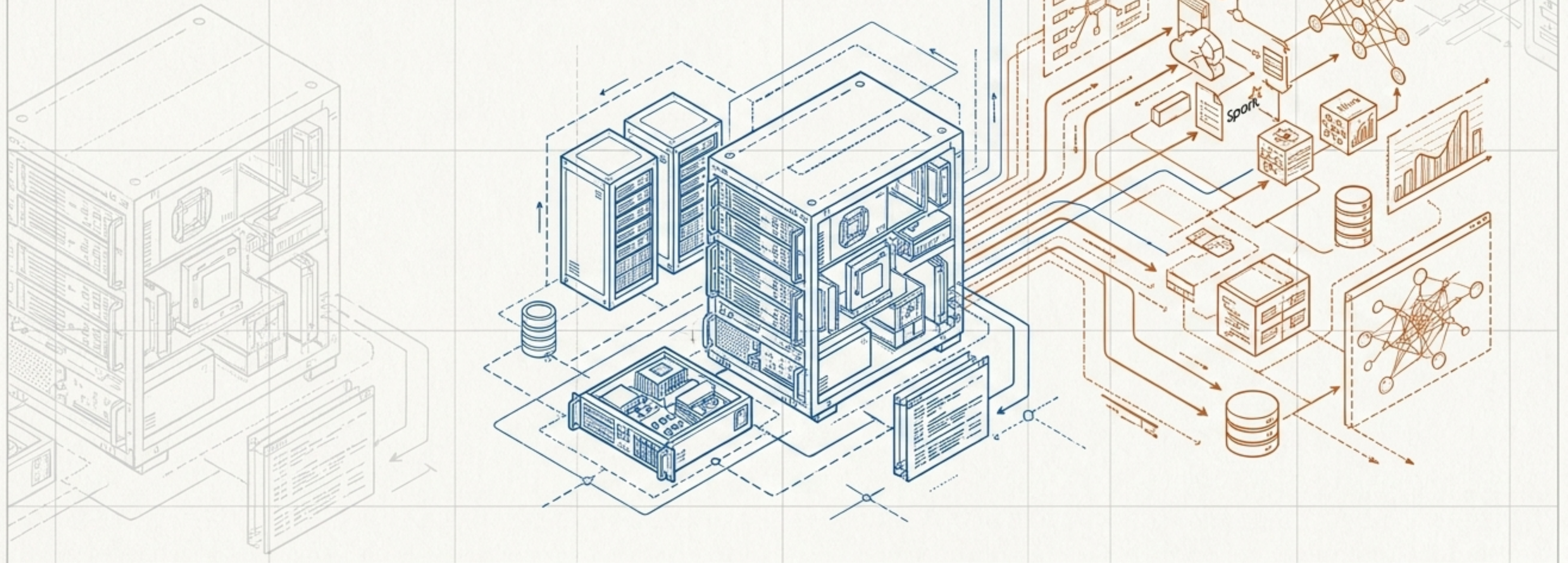


Architecting Big Data

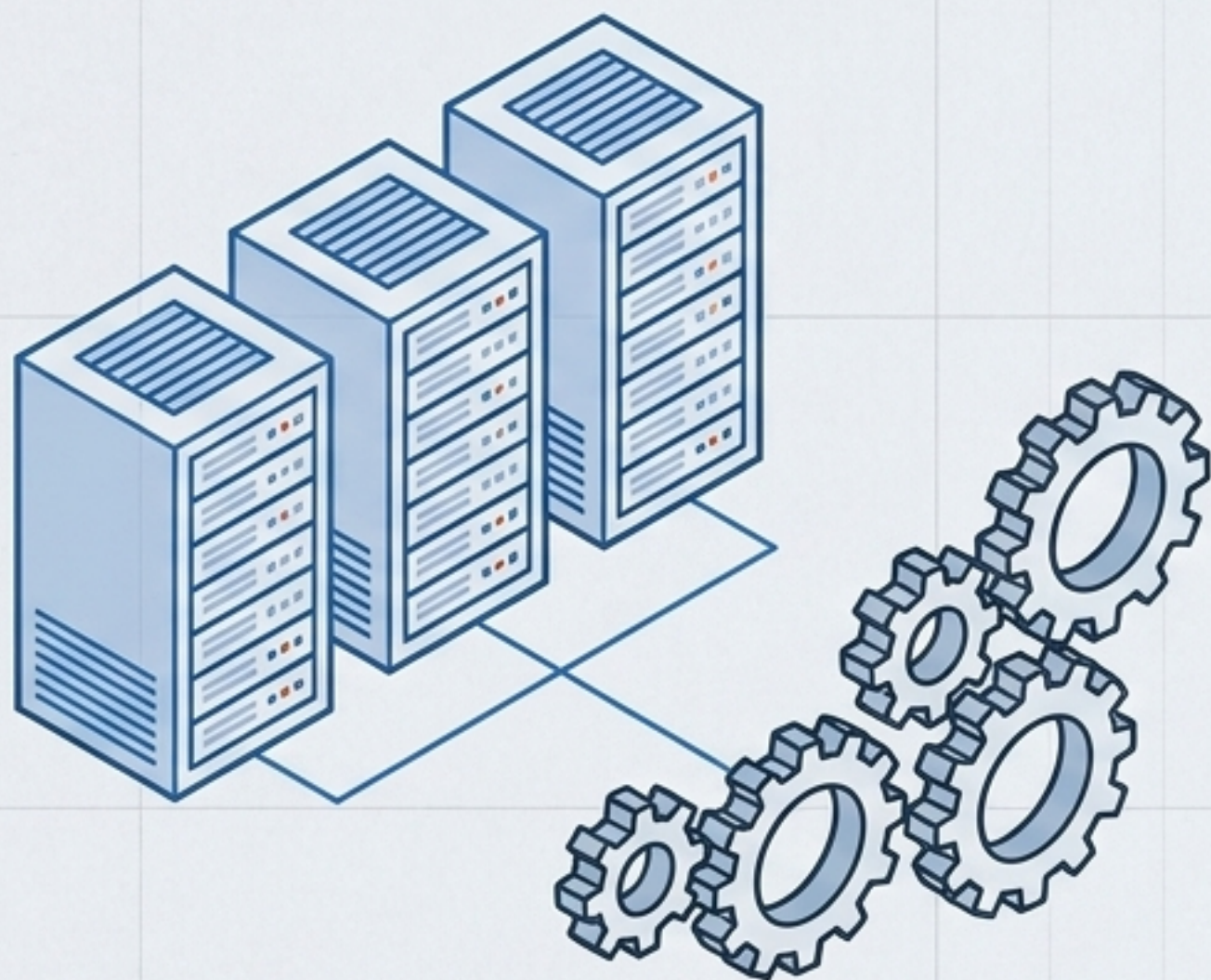
A Structural Blueprint of Management, Processing, and Modelling Frameworks



DATA REFINERY BLUEPRINT: MACRO-STRUCTURE OF BIG DATA ECOSYSTEM

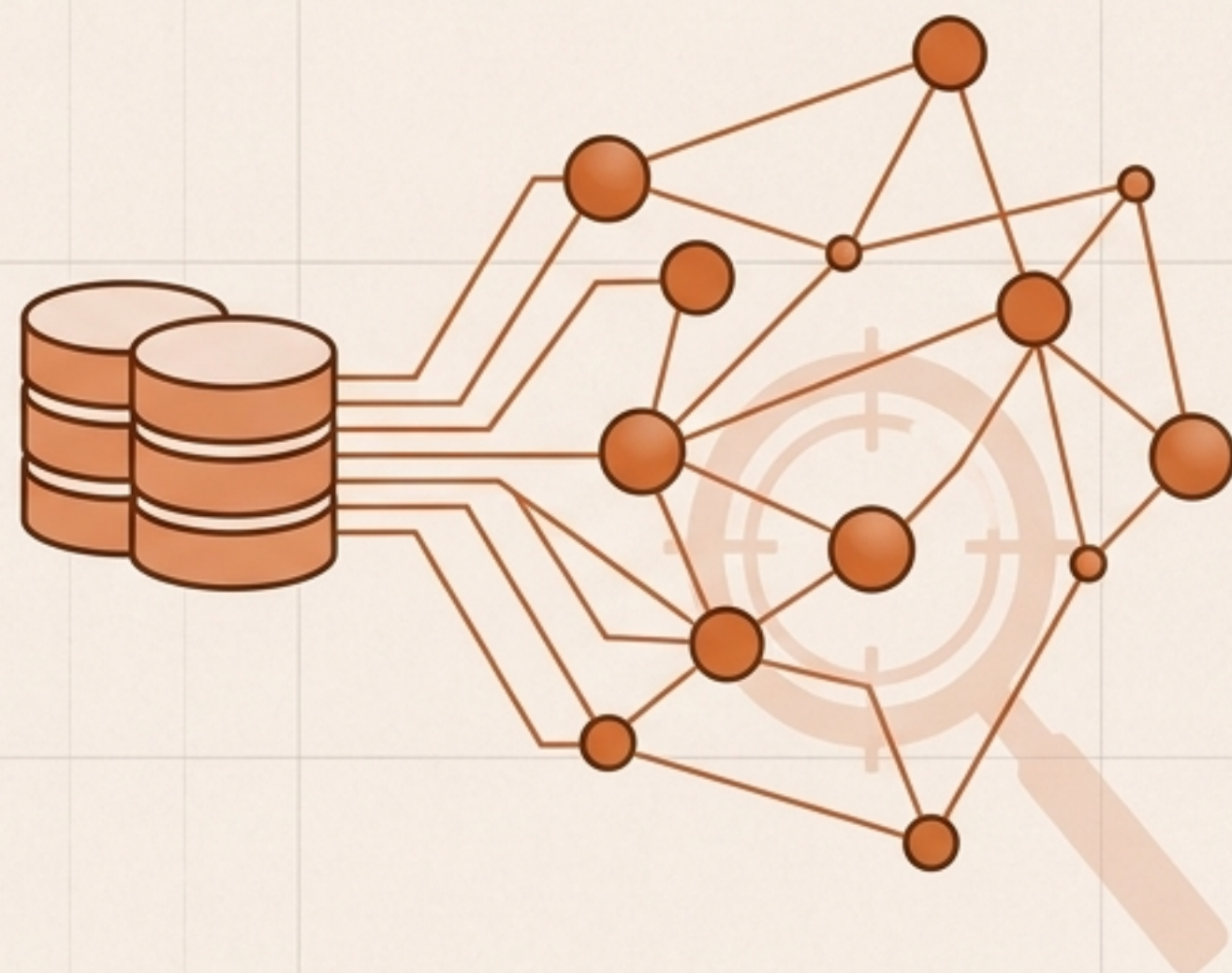
MANAGEMENT

The infrastructural foundation. Focuses on physical storage, resource allocation, and raw programmatic data processing.

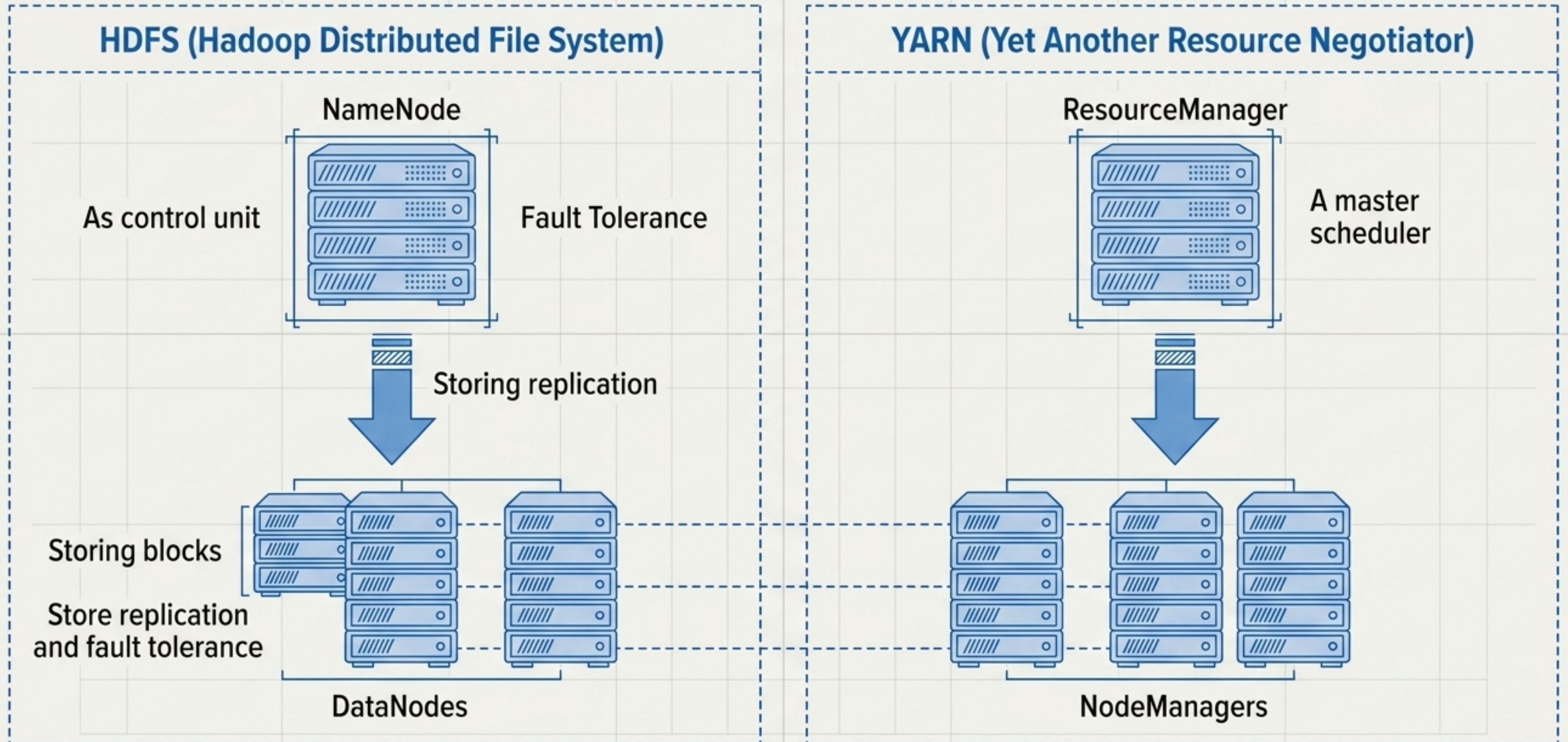


MODELLING

The access and structuring layer. Focuses on query-based processing, database architectures, and extracting insights.

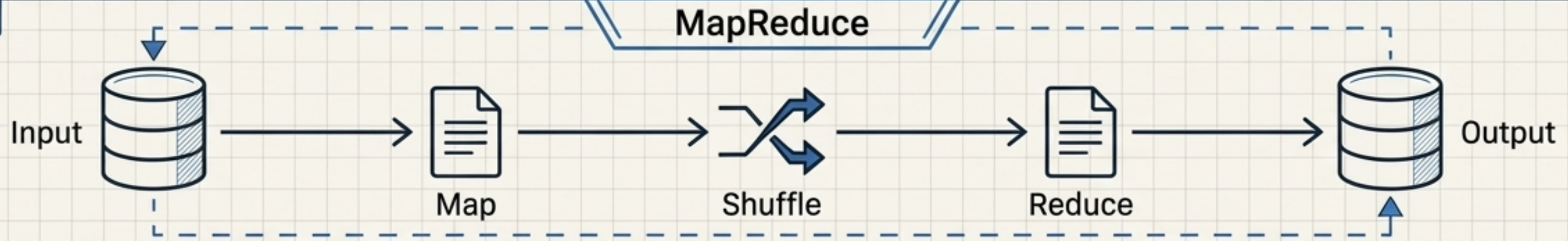


HADOOP ARCHITECTURE: MANAGEMENT LAYER SCHEMATIC

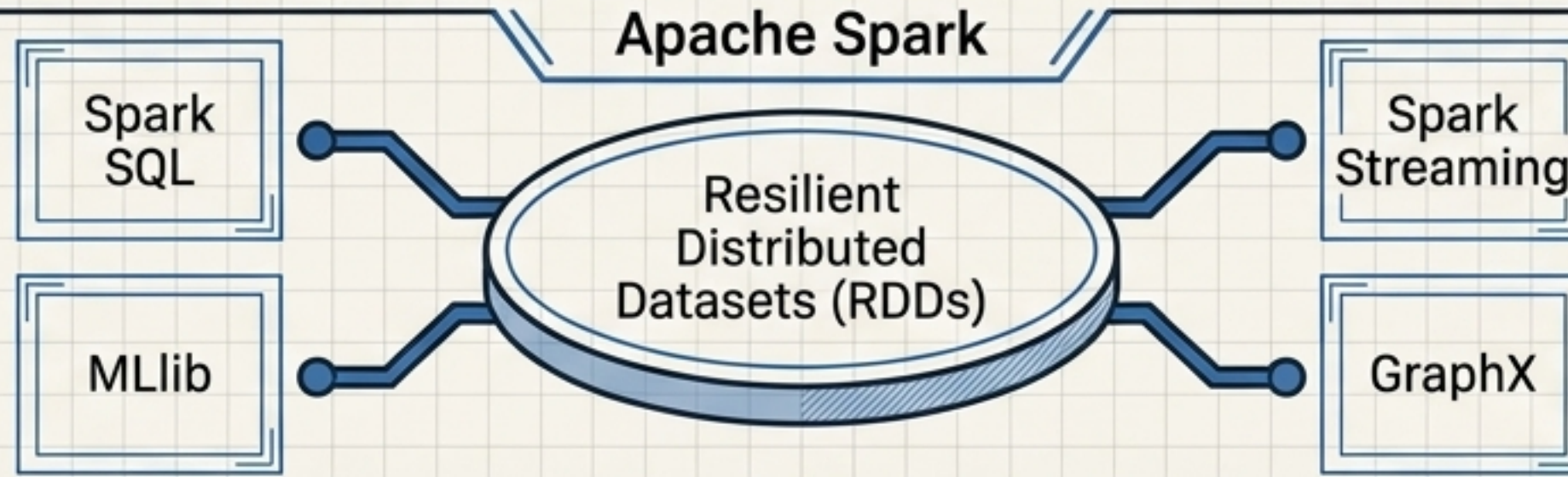


DATA REFINERY BLUEPRINT: PROCESSING SERVICES EVOLUTION

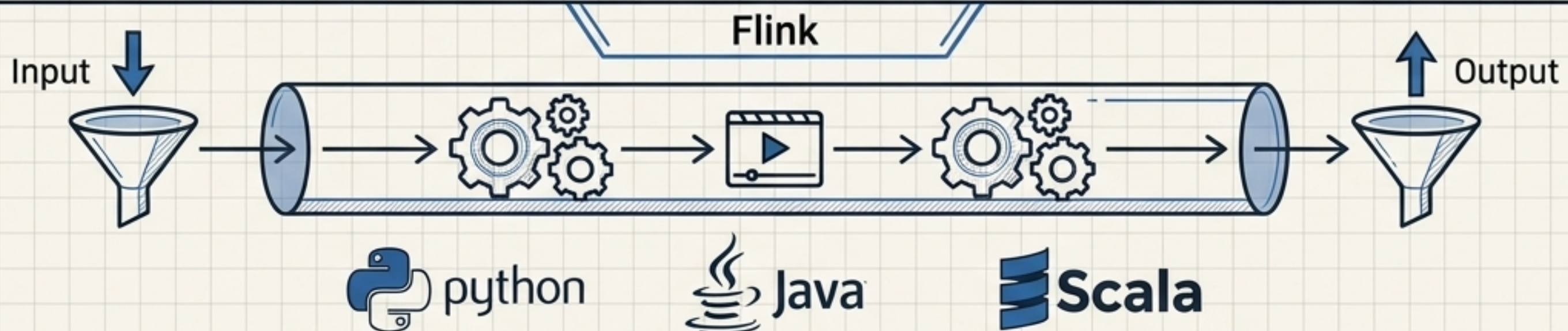
Tier 1



Tier 2



Tier 3



DATA REFINERY BLUEPRINT: SQL ENGINE COMPARISON MATRIX

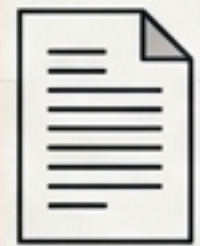
Hive

Impala

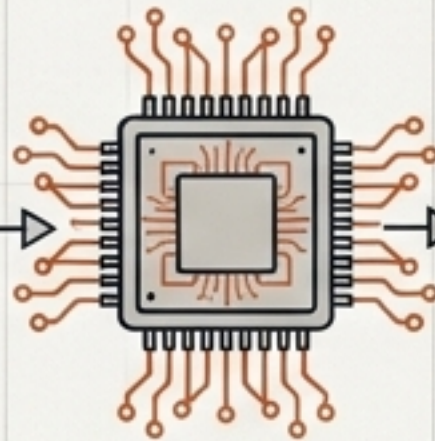
Presto



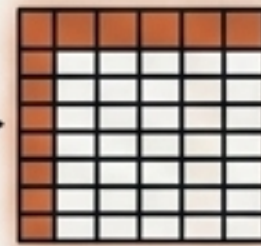
Metastore



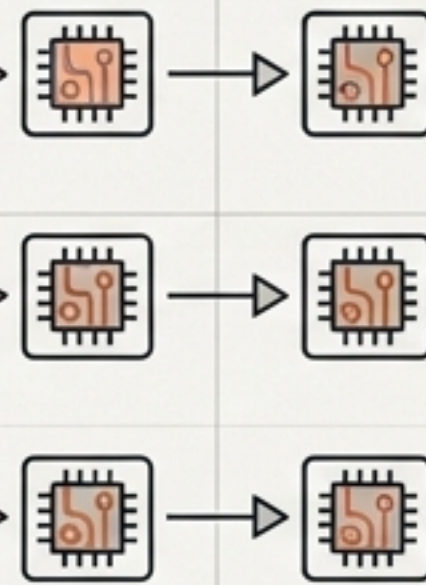
HiveQL



Execution Engine



Realistic table



Low latency



Presto



Database



Database



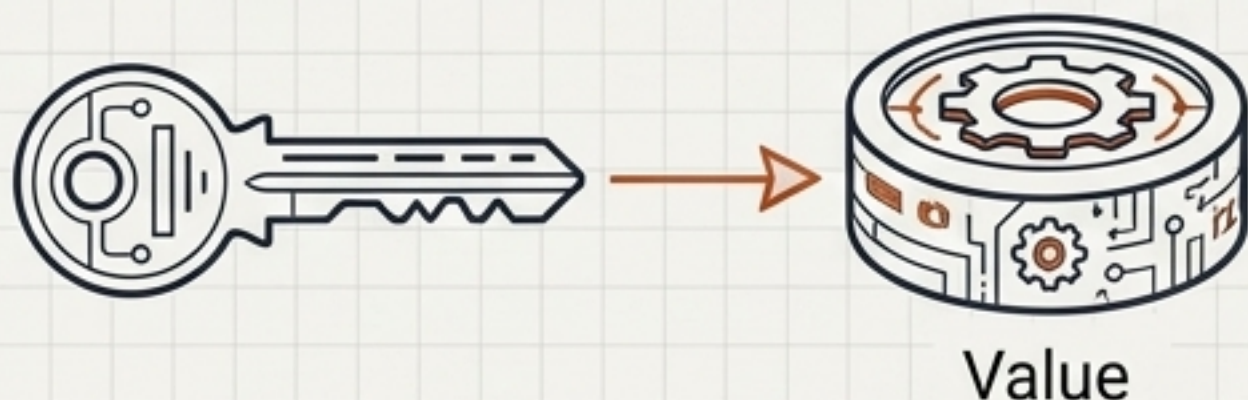
Database

A Distributed SQL query engine for ad-hoc queries

DATA REFINERY BLUEPRINT: NOSQL DATABASE LANDSCAPE

Key-Value

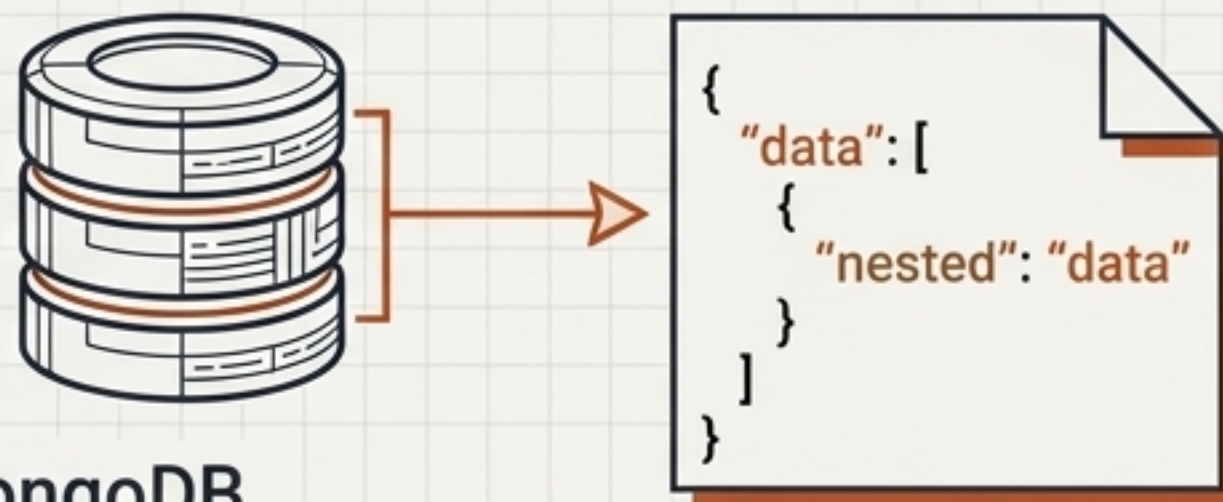
Key and value pair




 Redis  Couchbase

Document

JSON/XML documents are nested data



 MongoDB

Column-Family

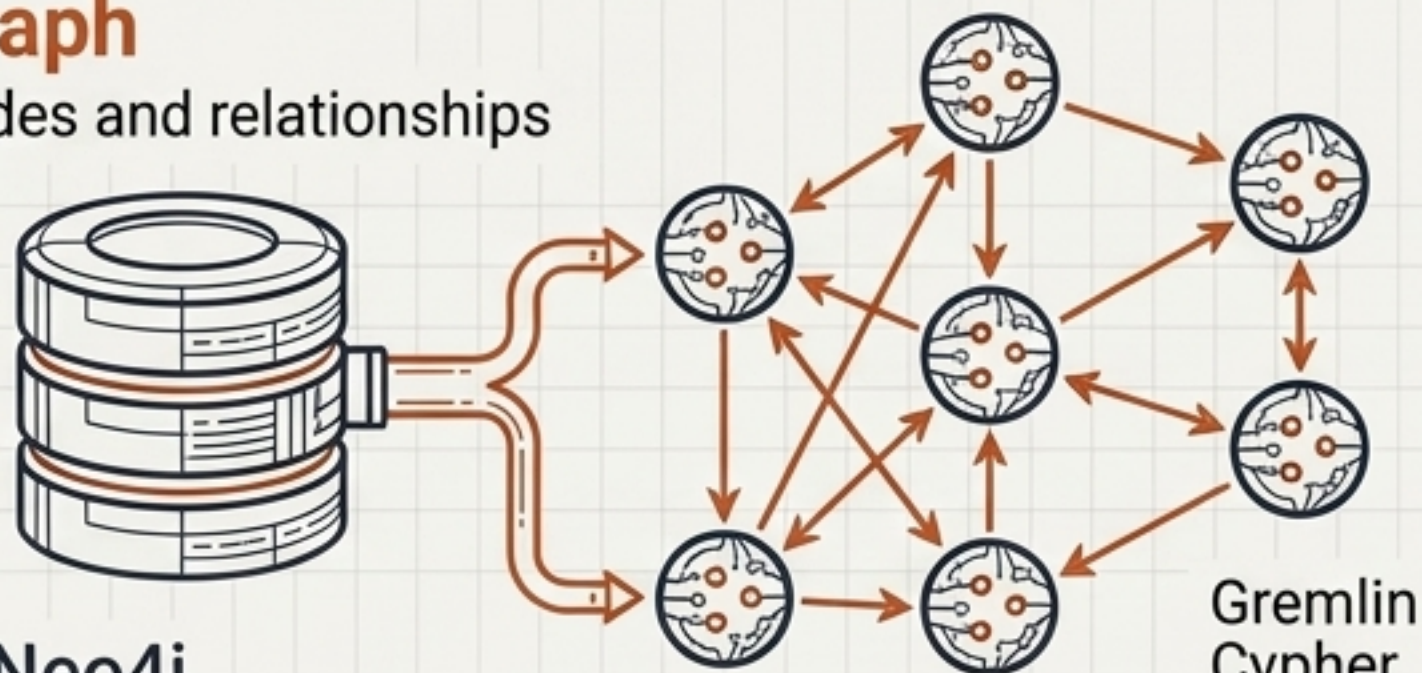
Routing a wide-column store (realistic sparse data)




Cassandra 

Graph

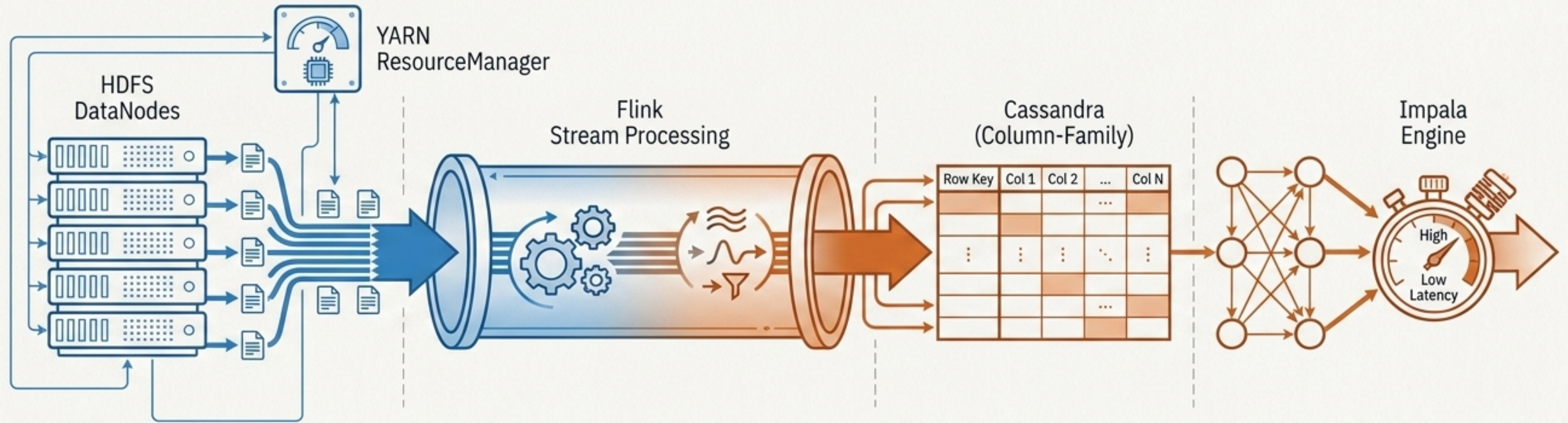
Nodes and relationships



 Neo4j

Gremlin
Cypher

DATA REFINERY BLUEPRINT: ECOSYSTEM MASTER PIPELINE



The Big Data ecosystem is not a collection of competing tools, but a highly specialized supply chain. Management frameworks handle the heavy lifting of storage and compute, while Modelling engines provide the agile, structured access required for human insight.