

Gartner Zone: Selecting the Optimal Data for Big Datastore Architecture With Gartner's Technical Assessments

Sanjeev Mohan

Why did the Oracle DBA leave the NoSQL party?



Long Live Bigger Data

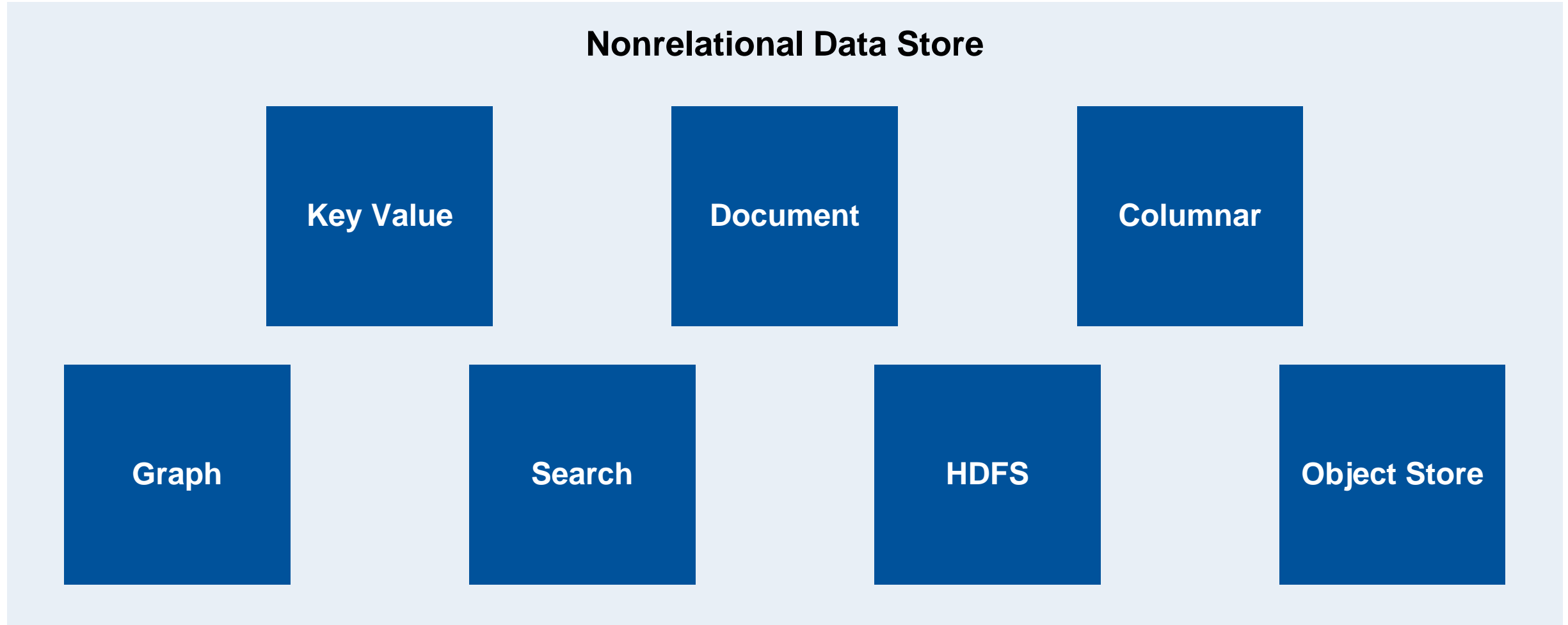
Comparison of relational and non-relational data stores

Relational

Non-relational

Consistent	Eventual consistency but configurable
Schema-on-write (normalized or dimensional)	Schemaless / schema-on-read
Structured data	Structured, semi-structured & unstructured
Scaling up (scaling vertically)	Scaling out (scaling horizontally)
Data Access: SQL / ODBC / JDBC or native APIs.	Data access: APIs and SQL
Complex SQL (joins & aggregations: focus on data)	Defined query patterns (focus on query access)
Performance: partitions and indexes	Performance: partitions, sharding and indexes
Security: tightly coupled	Security: fine-grained access control

Types of Nonrelational Data Stores



Key Relational Database Vendors

Established

ORACLE® TERADATA



Pivotal™



Open Source



Scale-Out

Google Cloud Spanner



Key Value Data Stores



redis



Document Data Stores



Azure Cosmos DB



Google Cloud Datastore

Column-Oriented and Columnar Data Store

Wide Columnar



Column-Oriented Relational / Analytical

- Oracle
- SQL Server
- MariaDB
- MemSQL
- IBM DB2
- SAP HANA
- Teradata



Cloud Data Warehouse

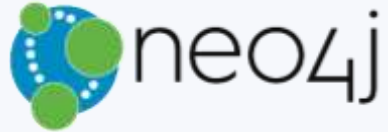


Azure SQL Data Warehouse

Gartner®

Graph Data Store

Property Graph



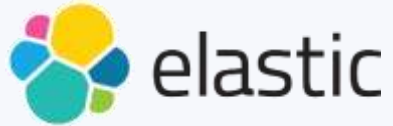
Semantic / RDF / Triples

- Allegro
- Cambridge Semantics
- BlazeGraph
- OpenLink Virtuoso
- GraphDB
- ArangoDB
- MarkLogic



Search Data Store / Time Series / Object Stores

Search



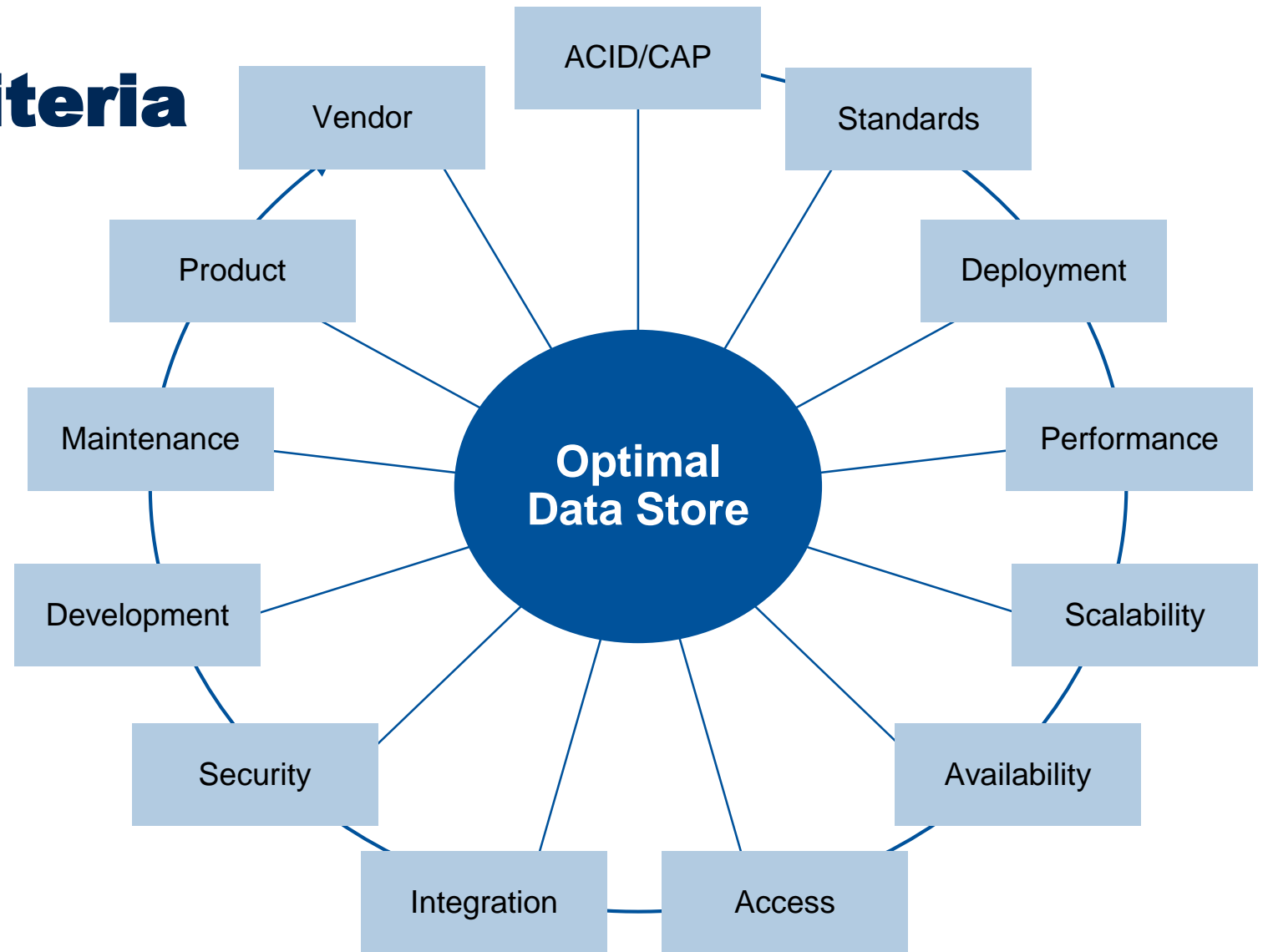
Time Series



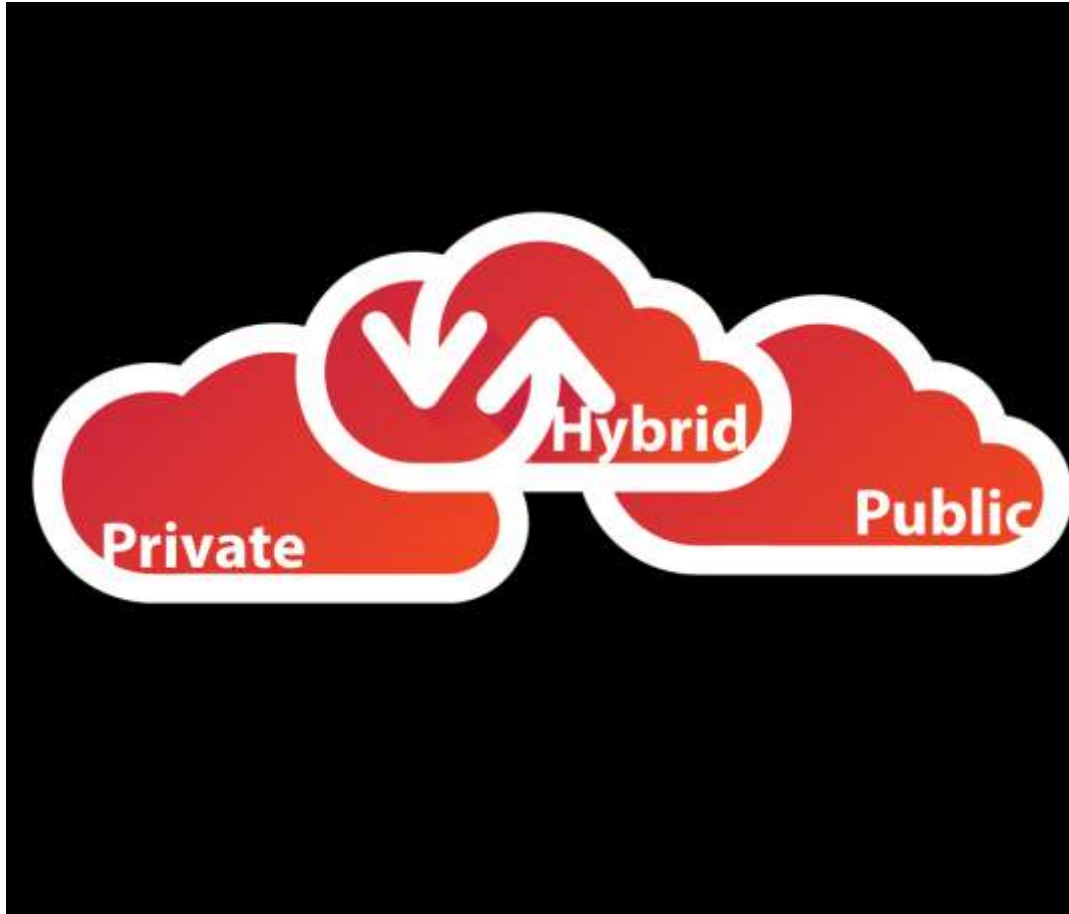
Object Stores



Optimal Persistent Store Selection Criteria



Trends in databases



- ***Goodbye OLAP cubes*** - Hybrid transactional and analytical processing (HTAP)
 - ***Multi-model databases***
 - ***Edge Computing***
 - ***Bring down the walls*** - find relations and intelligence across internal and external data sources
 - ***Non-volatile memory*** – performance of RAM, price of hard disk
-
- ***Serverless computing*** – fewer DBAs, focus on business SLAs

Recommendations

- ✓ Identify pain points and gaps with current the data infrastructure.
- ✓ Develop a structured evaluation criteria with weights.
- ✓ Shortlist a selected list of potential vendors.
- ✓ Rate the vendors and send an RFI template.
- ✓ Rank the vendors by evaluating the vendors.
- ✓ Migrate a use case and perform a proof of concept (PoC).

Action Plan for Effective Data Governance

Monday Morning:

- Assess current state of data infrastructure
- *Identify* gaps and remediation opportunities

Next 90 Days:

- *Develop* evaluation criteria and send out RFIs
- *Rank* vendors and select the optimal vendor.
- *Perform* proof of concept (POC) and determine costs/benefits.

Next 12 Months:

- *Productize* the solution.
- *Enhance* and improve solution as new workloads are added.

Recommended Gartner Research

- ▶ [Identifying and Selecting the Optimal Persistent Data Store for Big Data Initiatives](#)
Sanjeev Mohan (G00322578)
- ▶ [Enabling Essential Data Governance for Successful Big Data Architecture Deployment](#)
Sanjeev Mohan (G00327532)
- ▶ [EIM 1.0: Setting Up Enterprise Information Management and Governance](#)
Thornton Craig (G00342309)