#### **Gartner Data & Analytics Summit**

18 - 21 March 2019 / Orlando, Florida

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

Sanjeev Mohan

© 2019 Gartner, Inc. and/or its affiliates. All rights reserved. Gartner is a registered trademark of Gartner, Inc. and its affiliates. This publication may not be reproduced or distributed in any form without Gartner's prior written permission. It consists of the opinions of Gartner's research organization, which should not be construed as statements of fact. While the information contained in this publication has been obtained from sources believed to be reliable, Gartner disclaims all warranties as to the accuracy, completeness or adequacy of such information. Although Gartner research may address legal and financial issues, Gartner does not provide legal or investment advice and its research should not be construed or used as such. Your access and use of this publication are governed by Gartner's Usage Policy. Gartner prides itself on its reputation for independence and objectivity. Its research is produced independently by its research organization without input or influence from any third party. For further information, see "Guiding Principles on Independence and Objectivity."



# Why did the Oracle DBA leave the NoSQL party?





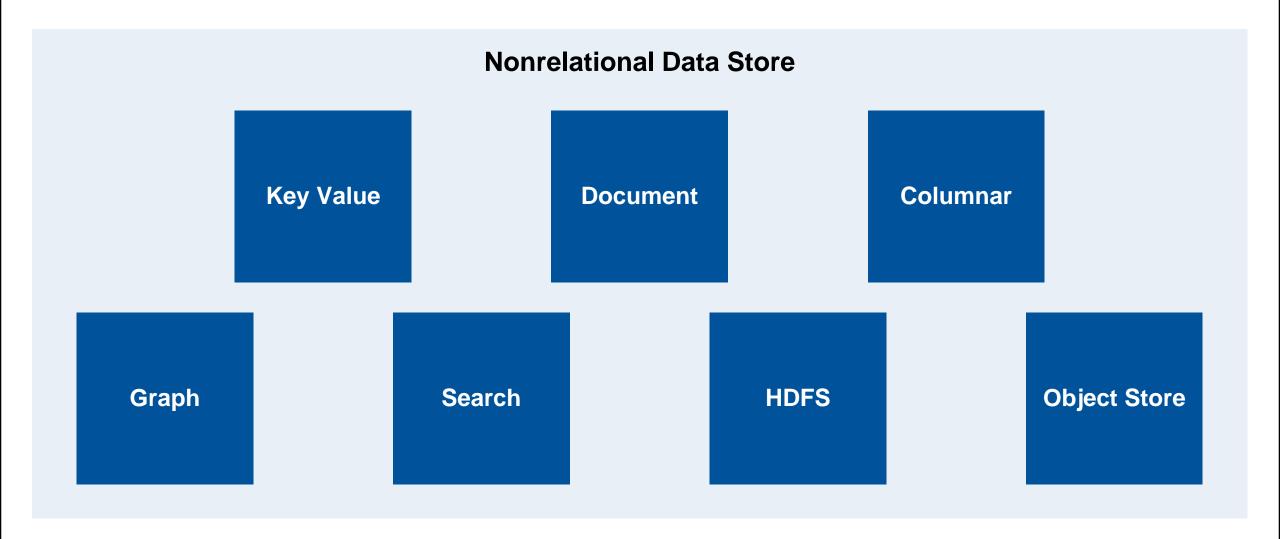
# Comparison of relational and non-relational data stores

# 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: <b>SQL / ODBC / JDBC</b> 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**















# **Open Source**







#### **Scale-Out**

Google Cloud Spanner









# **Key Value Data Stores**









∢EROSPIKE-



## **Document Data Stores**





















# **Column-Oriented and Columnar Data Store**

#### **Wide Columnar**









#### **Cloud Data Warehouse**









# Column-Oriented Relational / Analytical

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









# **Graph Data Store**

### **Property Graph**















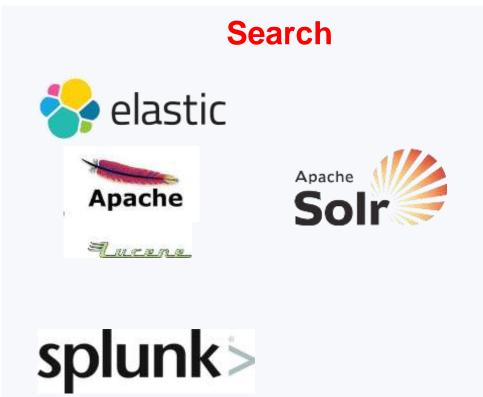
# Semantic / RDF / **Triples**

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





# Search Data Store / Time Series / Object Stores









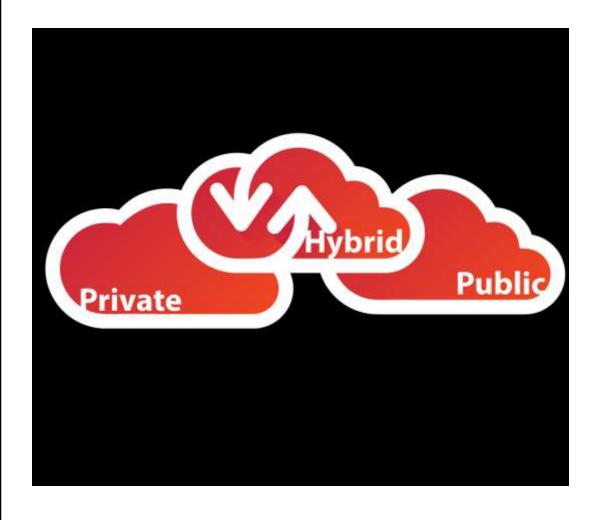
# **Optimal Persistent** ACID/CAP **Store Selection Criteria** Vendor **Standards** Deployment **Product** Maintenance Performance **Optimal Data Store** Development Scalability Security Availability

Integration

Access



# **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)

