Scalable SQL and nosql Data stores

Author: Rick Cattell

Presentation by : Gurpreet Bajwa

?Database Model?

It tells you about the logical structure of a database and ways to store and manipulate the data

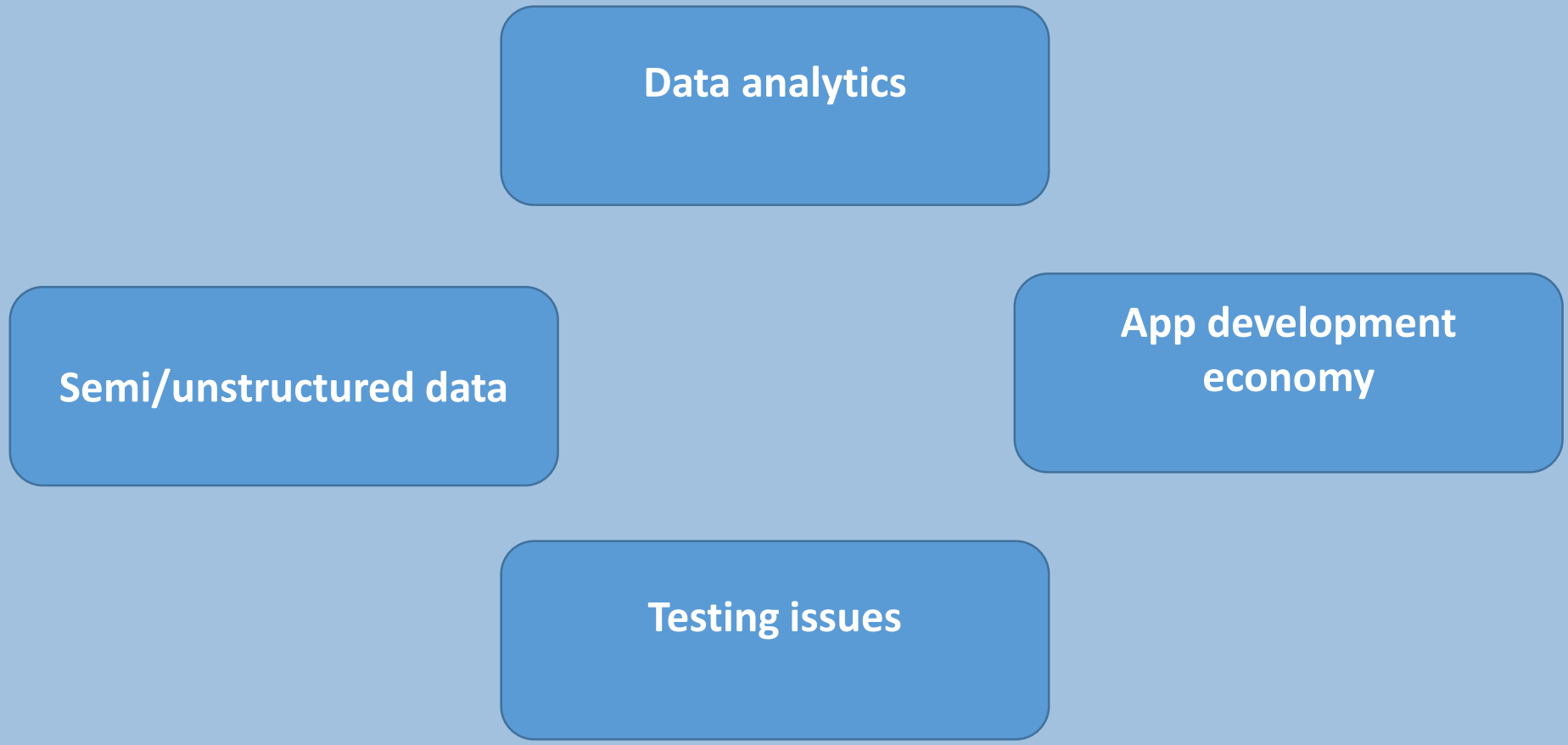
The RIGHT Data Model

Picking up the right model helps everyone:

- Company (profits)
- Client (trust)
- Developer (time)

Studies have shown that about 70% of software development efforts fail in wake of premature development.

Top factors impacting global NoSQL market



A few Data Stores..

- *Key-value Stores*
- *Document Stores*
- *Extensible Record Stores*
- *Scalable RDBMS*

Key-value stores

- Simple
- one object per database
- Efficient
- Easily scalable*

Key	Value
K1	AAA,BBB,CCC
K2	AAA,BBB
K3	AAA,DDD
K4	AAA,2,01/01/2015
K5	3,ZZZ,5623

When?

For persisting objects obtained by redundant CPU and memory intensive computations. E.g. dictionary implementations

Redis is one of the fastest in-memory data stores, stores all data in RAM, high speed and performance, best used when time is an issue

e.g. Job management, queuing, real-time analytics, messaging, in-app social capabilities and geo search.

Document stores

- Multiple objects per database
- Multiple indexes and object nesting
- Nesting of objects affects overall performance (update, retrieve)
- Simple interface

When to use?

For applications agreeable with ‘eventual consistency’, involving multiple objects; look-up dependent on multiple constraints

Extensible Record Stores

- Columnar data storage
- Rows and columns (like RDBMS)
- Dynamic columns
- Rows of a table may subscribe to different columns

When to use?

Can be chosen over Document stores when your application requires multiple horizontal/vertical partitioning

Cassandra is most effective when used for managing huge volumes of data (the kind that don’t fit in a single server), such as Web/click analytics and measurements from the Internet of Things.
Writing to Cassandara is extremely fast.

Both HBase and Cassandra follow the BigTable model. As such, HBase can be scaled linearly simply by adding more nodes to the setup.

Relational Databases

- Trade-off between scalability and performance
- RDBMS with horizontal scaling
- Better option over NoSQL in absence of cross-node operations

How to pick the right Data Store

- Requirements gathering & analysis
- Data store feedback and analysis
- Code maturity

What Industry needs..

Database scalability

Handling unstructured data

Robustness

Failure Handling

Distributed computing

Market trends

Lets have a look at Gartner's Magic quadrant for years 2013 & 2016

Figure 1. Magic Quadrant for Operational Database Management Systems

Source: Gartner (October 2013)

A few pointers from 2013

- Substantial interest in NoSQL offerings
- Confusion in picking NoSQL vendors

"NoSQL deployments have been overwhelmingly supplemental to traditional relational DBMS deployments, not destructive."

Nick Heudecker

Research Director, Gartner

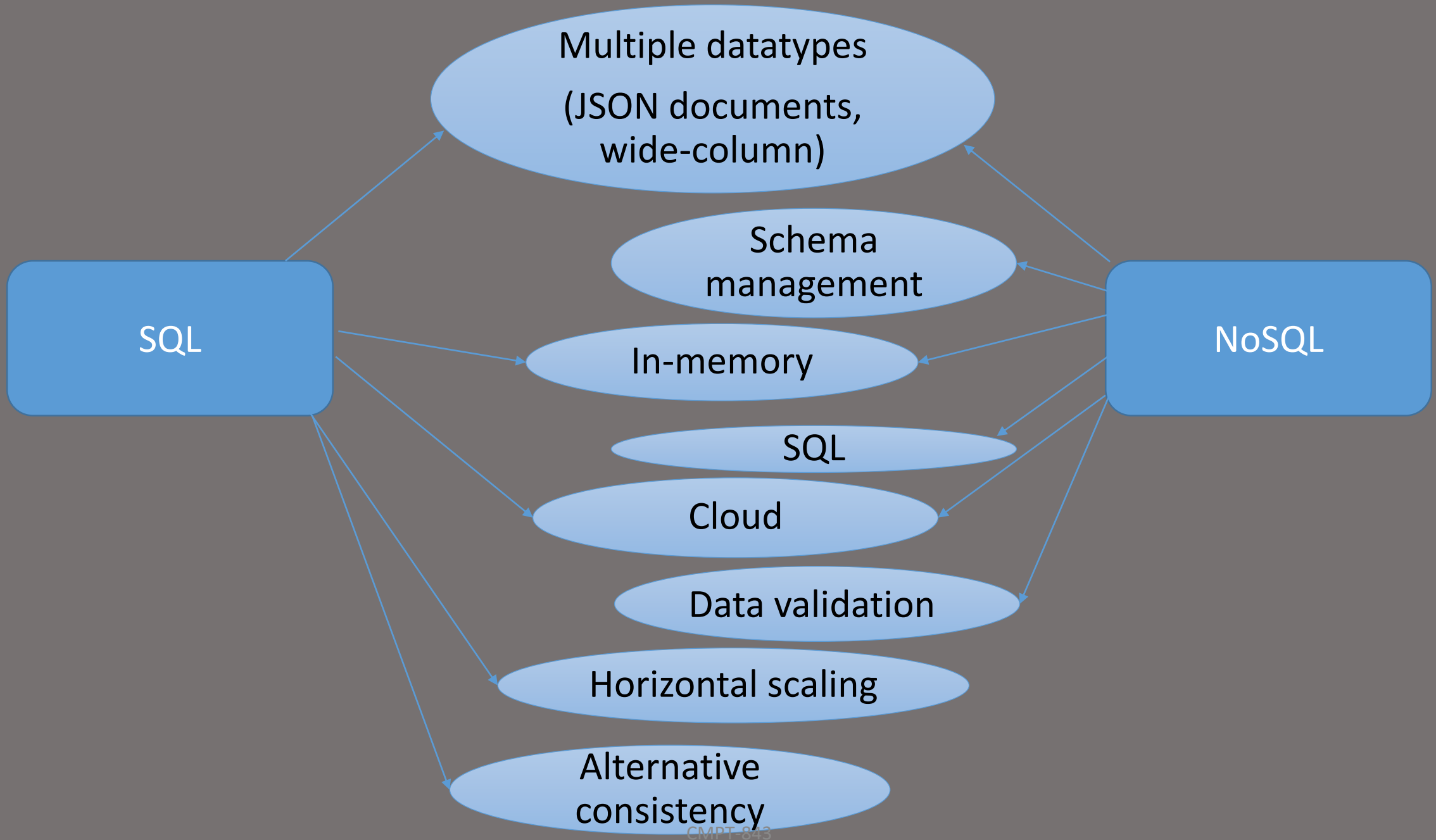
Magic Quadrant (as of Oct 2016)

*********NO VISIONARIES *********

Figure 1. Magic Quadrant for Operational Database Management Systems

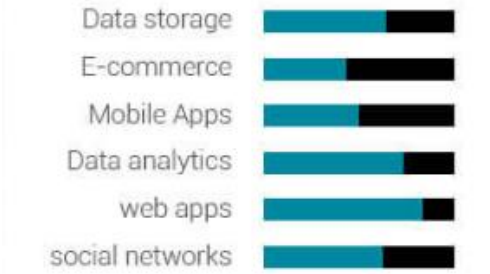
Source: Gartner (October 2016)

- **Lack of overall vision in the market**
- convergence between **NoSQL** vendors
- convergence between **NoSQL** and traditionally relational vendors


Multiple datatypes
(JSON documents, wide-column)
Schema management
SQL
In-memory
NoSQL
SQL
Cloud
Data validation
Horizontal scaling
Alternative consistency

nosql

Global NoSQL Market By Application



Web Apps
to be highest revenue generating segment throughout 2013 - 2020

Global NoSQL Market By Type

Key-value store to be highest revenue generating segment throughout 2013 - 2020

Global NoSQL Market By Geography

North America to be the highest revenue generating region throughout 2013 – 2020

Global NoSQL Market Dynamics

Drivers:
- Rise in unstructured data
- Attractive business models
- Demand for data analytics
- Growing App development business

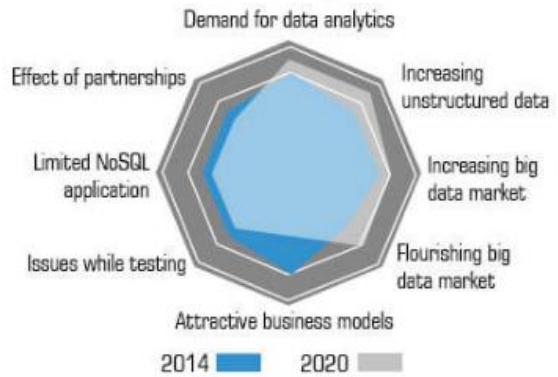
Restraints:
- Difficulties while testing of NoSQL applications
- Problems while adoption in structured data applications

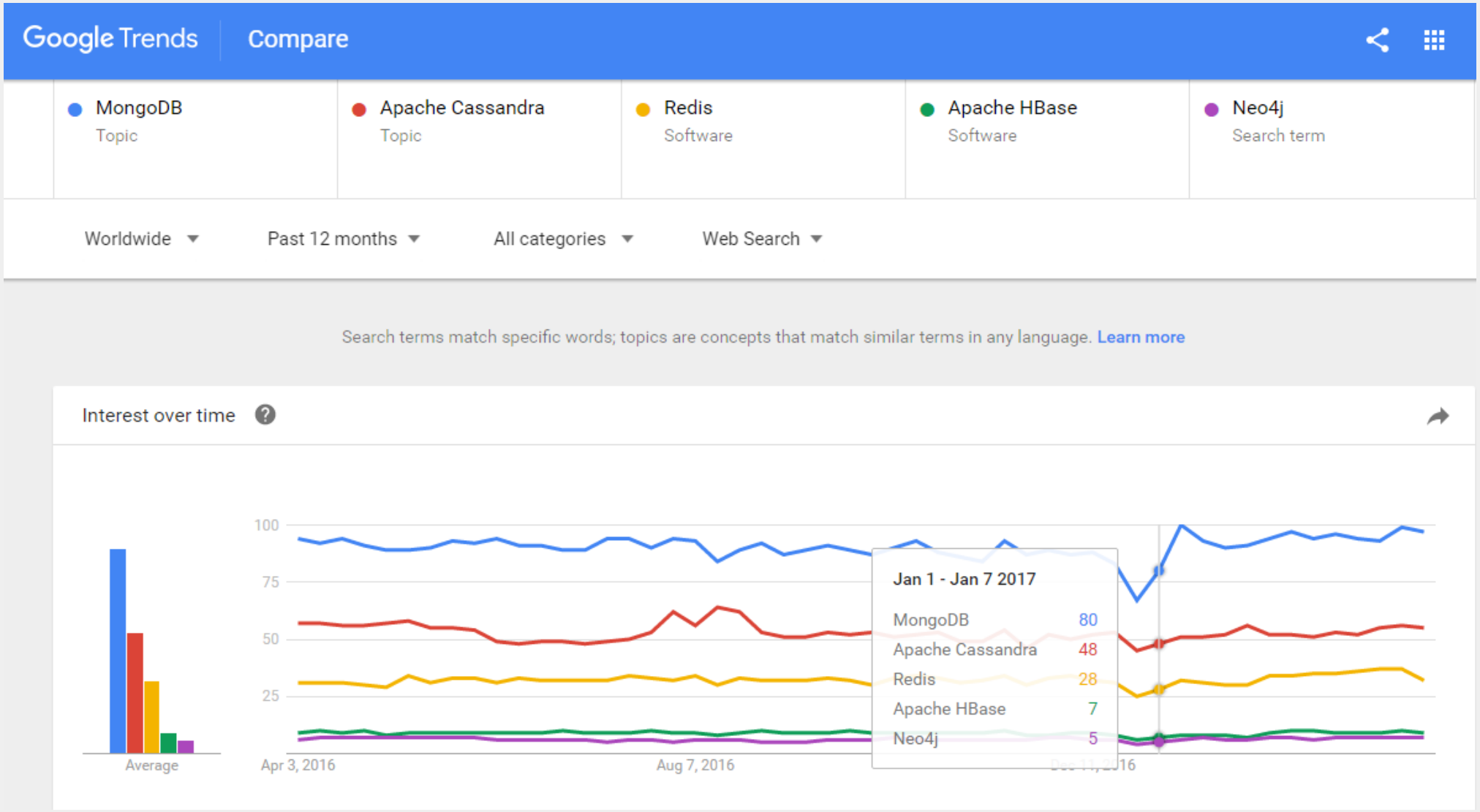
Global NoSQL Market

Global NoSQL Market is expected to reach at $4.2 billion by 2020

Growing at a CAGR of 35.1% (2014-2020)

Global NoSQL Market Top Impacting Factors




Google Trends
Compare
MongoDB
Topic
Apache Cassandra
Topic
Redis
Software
Apache HBase
Software
Neo4j
Search term
Worldwide
Past 12 months
All categories
Web Search
Search terms match specific words; topics are concepts that match similar terms in any language. Learn more
Interest over time
100
75
50
25
Average
Apr 3, 2016
Aug 7, 2016
Jan 1 - Jan 7 2017
MongoDB 80
Apache Cassandra 48
Redis 28
Apache HBase 7
Neo4j 5

?Internet of Things?

***"The Internet of Things (IoT)** is the network of physical objects that contain embedded technology to communicate and sense or interact with their internal states or the external environment."*

IoT & NoSQL

- Traditional Databases cannot manage IoT scale
- Time-series data, critical to IoT analytics

Handling Time series data
Flexible data modeling
Data compaction
Fast and easy range queries

Conclusion

- No tailor-made answer to Data store selection
- Prioritize requirements and determine tradeoffs
- A detailed market analysis suggested before finalizing a Data store

References

- http://www.odbms.org/blog/2013/12/operational-database-management-systems-interview-with-nick-heudecker/
- http://www.odbms.org/blog/2016/11/new-gartner-magic-quadrant-for-operational-database-management-systems-interview-with-nick-heudecker/
- https://www.gartner.com/doc/3467318/magic-quadrant-operational-database-management
- http://cattell.net/datastores/
- https://thestack.com/iot/2016/02/15/why-nosql-is-the-future-of-iot/
- https://www.alliedmarketresearch.com/NoSQL-market
- http://www.itbusinessedge.com/slideshows/top-five-nosql-databases-and-when-to-use-them-05.html

fin

Questions?