AEROSPIKE IN FINANCIAL SERVICES APPLICATIONS – A COMPETITIVE EDGE

Abstract

Data is the hinge point of today's Financial Services' organizations business operations and processes. With the onslaught of web, digital commerce and social computing, there are large volumes of unstructured and information-oriented data being generated every minute. Often NoSQL databases are used to store and retrieve the structured, unstructured or semi-structured data. The data explosion has given great impetus to adoption of NoSQL Databases.

Even amongst NoSQL databases there is a differentiation on various factors. One of the very important being – the cost-to-latency ratio. Often, it is difficult to predict the response time when the volumes of data increase exponentially. Among the NoSQL database providers, Aerospike was first to build and introduce Flash Optimized NoSQL database, which can provide a linear predictable response for real time applications, even if the data footprint expands to billions of transactions per day and processing terabytes of data. Aerospike's Hybrid Memory Architecture offers a fundamentally different approach by persisting data on fast SSD devices and leveraging primary key indexes. This paper discusses how NoSQL databases help address the challenges faced by financial services organizations and how Aerospike can be the preferred choice amongst the various NoSQL databases. The paper also discusses the comparative analysis of various NoSQL databases and how the Aerospike's unique architecture provides a competitive edge to the financial services organizations.

Infosys®

Navigate your next

Transforming Financial Services – Using Aerospike NoSQL Database

Common data processing challenges in FS

Financial services organizations have needs which can often be solved easily by adoption of NoSQL family of databases - specifically Aerospike. Here are some of such challenges as faced by the financial organizations in processing data and how Aerospike helps address the challenge.

Fragmented data with multiple sources	Data in financial service institutions is complex and has multiple sources within the organization. Data driven decisions require a consolidated view of data. Hence this might cause hindrance for data driven decisions capabilities. Also, it is uneconomical for financial organizations to spend on scaling maintaining these multiple systems

Aerospike can store all kind of data in the same cluster. This is because it does not have rigid schema; it is non-relational and cluster friendly. This helps solve the problem of silos and allowing the financial organization to use this data centrally to take data driven decisions. Aerospike is economical and can be easily scaled out.

Realtime processing	In traditional data processing techniques real time data processing is not possible for because of complex relational data structures

Aerospike enables in-memory data storage. The fast storage and retrieval combined with the non-relational nature enables highly scalable low latency operations. This offers financial organizations perform real time processing in marketing, personalization applications

Availability and Performance	Mission-critical consumer-facing applications demand availability and have performance challenges. Increase in volumes coupled with the transition to distributed architectures need to have modern ways of managing clustering and load balancing

Due to its non-relational nature, Aerospike can handle data smartly across clusters. These offer balance between highly consistent and highly available models. Smarter cluster management and load balancing options combined with tunable consistency available in NoSQL databases makes handling increased load with improved performance lot simpler

Data integrity	Traditional RDBMS mandates multiple sources of data for different types of consumption. Combination of OLAP, OLTP and custom reporting databases are deployed for different use cases. In such scenarios data integrity becomes a critical aspect to consider

Aerospike, enables single source of data for all kinds of use cases. Different consumers can access this single source of data for different operation on account of non-relational and low latency. This does not require to maintain consistency between multiple sources

Data Intensive Operations	For Legacy systems like Mainframe were designed for stability, but modern applications require data intensive operations like machine learning and real-time operations

Aerospike offers non-rigid schema and stability of the kind that mainframes offer. On account of their smart storage techniques, it is more suited to Machine Learning and real-time processing. This makes these databases more suited for transition from mainframe

Machine Learning	Traditional databases have difficulties in processing both structured and unstructured data which is a trait of Machine Learning systems

Aerospike is schema-free and highly scalable. This makes them highly suited to machine learning systems.

Pre-generated data	Most Financial organizations are working on pre-generated dashboards for better customer experience. Traditional RDBMS databases lag the speed required to pre-generate the data required for dashboards in the stipulated time

Aerospike is highly performant and provides fast in-memory storage and retrieval. The real time dashboards and views can be quickly generated and stored into these – making fan out microservices a reality.

Among the various available NoSQL databases, Aerospike's Hybrid Memory Architecture offers a fundamentally different approach, by persisting data on fast SSD devices and leveraging primary key indexes. Many advantages are realized without compromising performance while adding persistence, correctness, and security. Few advantages of Aerospike usage for financial services are:

- Improved data accuracy
- Faster data processing
- Better customer experience, high reliability
- Low TCO
- Reduce dependency on legacy systems like Mainframes

Scenarios for FS organizations leveraging Aerospike are,

Caching	Aerospike Caching is an ideal choice for building interactive customer experiences such as real-time use cases like Leaderboards, Real-Time Analytics, Geospatial Services and Queue Management.
Fraud Prevention	Because of its predictabe high performance against large transaction volumes, industry leading availability and uptime, scaling with low latency and significantly low TCO Aerospike is a good choice for fraud prevention activiies
Real Time System of Records	Data in the SOR can be available for Realtime Analytics and ML systems. The models for these systems can be updated in real-time and the model updates can be synchronized frequently for real-time decision making.
User Profile Store	Business requirement for User Profile Store are, consistent availability, often 24x7x365, and integration with business intelligence systems that provide actionable information. This can be easily achieved by Aerospike
Predictable high performance	Modern day applications demand an instant response of the tune of few milliseconds. Aerospike can provide nearly constant low linear latency across all data accesses, from low data volumes to 100's of TB of data.
Choice of Availability vs Consistency	Aerospike allows choosing between High Availability and High Consistency. These along with replication factor and various options of cluster formation allowing customizations based on the business need
Parallel Architecture and Linear Scalability	Aerospike is a multi-master distributed database. It is comparatively extremely simple to add a node to an Aerospike cluster. Moreover, Aerospike automatically balances and rebalances data across multiple nodes.

Aerospike: In Tech Conversation

Aerospike is a distributed NoSQL database. It is very fast, highly scalable, and reliable for use in real time data applications. It has schema-less data model, a multi master configuration and a configurable consistency model.

With the evolution around DRAM technology and the need of extremely low latency in applications, in-memory databases began to flourish. Aerospike was the first to build Flash optimized NoSQL database that can provide linear predictable response for real-time applications dealing with billions of transactions per day and processing terabytes of data.

Ranking and Market Research

- **Forrester Consulting:** Aerospike customers can achieve a projected ROI of 446% to 574% over three years: Forrester-total-economic-impact.

- **G2 Grid:** Report for Key-Value store Summer 2019 shows Aerospike as definitive leader among key-value databases – g2-grid-report

- **Bloor Research:** Aerospike has top rankings for Architecture, performance and scalability - Bloor-research
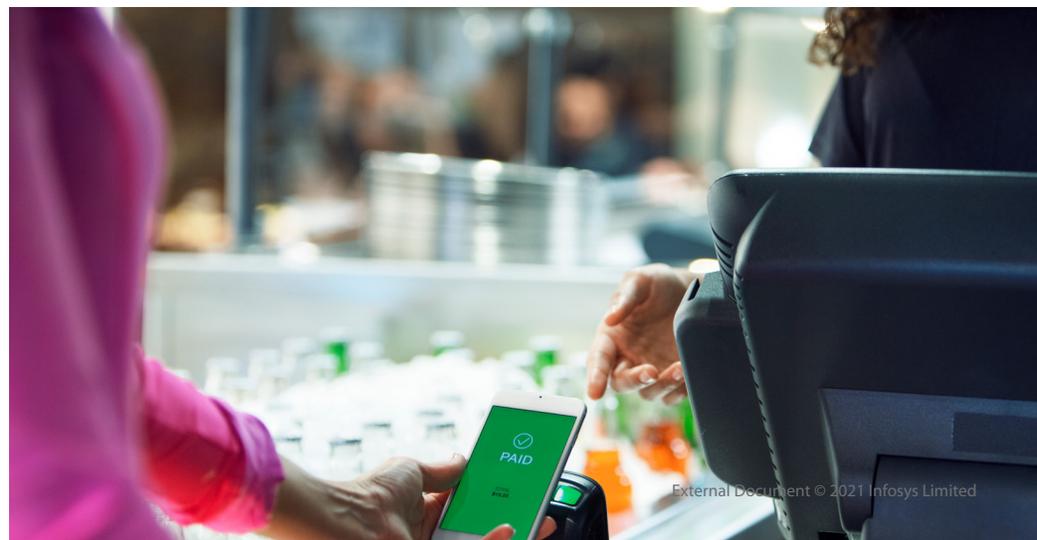
High Performance Architectures with Aerospike

High Level Architecture
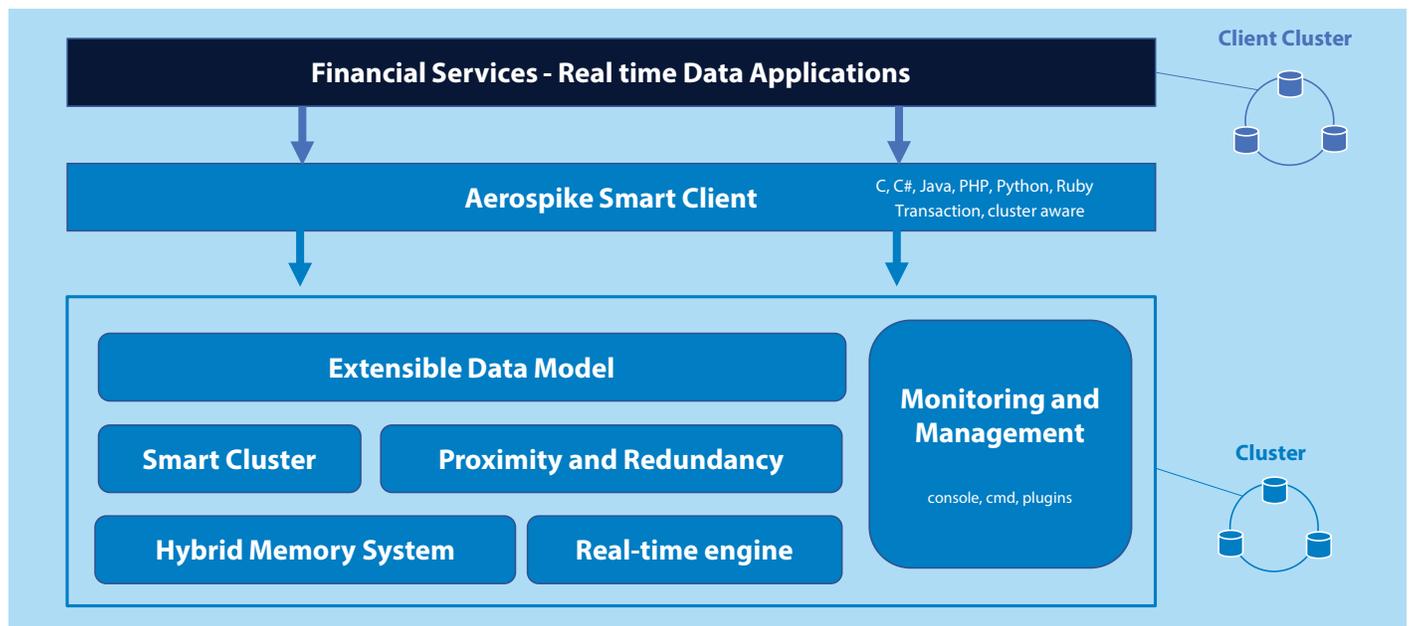
Key architecture features of Aerospike are

Smart Client

This is available for most of the commonly used programming languages and frameworks. Automatically distributes both data and traffic to all the nodes in cluster. Each node maintains a list of its neighboring nodes and this list is used for discovery of the cluster nodes

Each client process stores the partition map in its memory. Client periodically consults server nodes to check if there are any updates by checking the version stored locally against latest version of server

Clustering

Aerospike clustering plays an important role in reliability and high availability features as it quickly recovers and reforms cluster in case of failure. This recovery is done using the heartbeat feature where all nodes track each other and coordinate themselves

Monitoring and Management

Provides Management Console, command-line tools and Plugins like Nagios Graphite and Zabbix

Extensible Data Model

Aerospike's data model does not conform to rigid schema. New data types can be added on the fly. The data model consists of:

Namespace (Contains records, indexes, and policies),

Set (it is within the namespace which is a logical container for records)

Record (A row, unit storage in the database)

Bin (Columns within a row/record with name and value)

This helps retrieve and update precise amount of data.

Architectural Features

Hyper scale Data Architecture

- Aerospike provides great performance at scale for Systems of Record by utilizing an end-to-end platform
- Edge Database SOE (System of Engagement)
- Query and Reporting DB
- Cross Datacenter replication (XDR)
- Connect for existing infrastructure
- Connect for Kafka
- Aerospike Clients support for multiple development languages like C#, Java

Hybrid memory Architecture

- Aerospike implements a hybrid memory architecture wherein the index is purely in-memory (not persisted) and data is stored only on a persistent storage (SSD) and read directly from the disk
- High performance Key-Value store DB
- Strong consistency
- Performance is predictable
- Liner scalability
- Low Total Cost of Ownership (TCO)
- Self healing clusters gives high availability

CAP and ACID Support

- Aerospike prioritizes availability over consistency in each subsystem
- Typically supports ACID features except Multi-record isolation, consistency and Query isolation
- A multi-record query will be consistent in the face of writes

# Patterns for Integrating Aerospike

Greenfield development

- For greenfield development, it is best to start with cloud offerings and make use of SSD optimized VMs they offer to develop a cloud native application. Aerospike has been working with cloud providers to offer optimized performance with low latency.

- Cross Data Center replication feature with the help of Rack Aware Aerospike Enterprise solution helps to provide data at the nearest node further reducing latency.
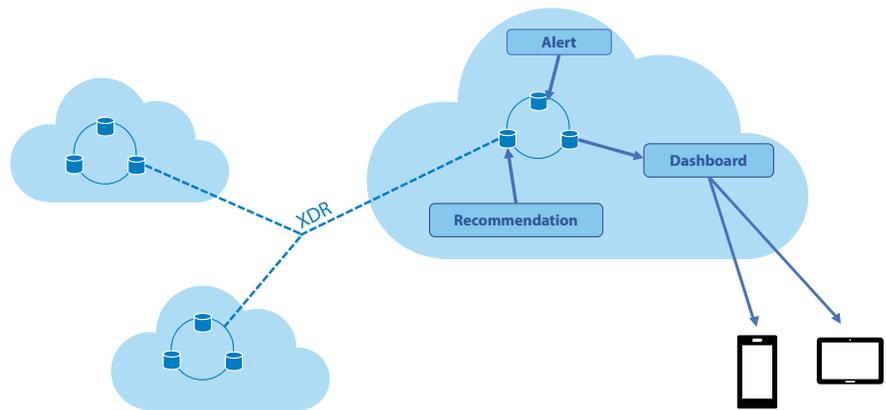
- Many Financial Organizations use a micro-service-based approach to pre-generate dashboard, recommendations and alerts

as in the depiction above. These are readily provided to different channels of customers like mobile devices or browsers.

Brown Field Development
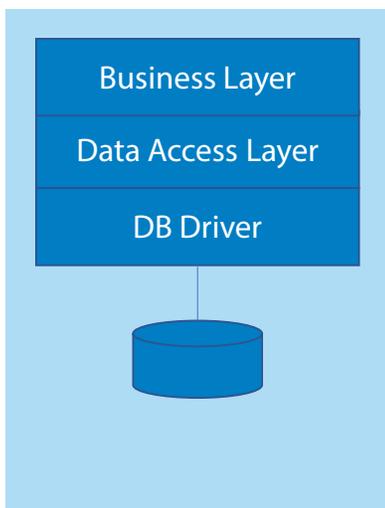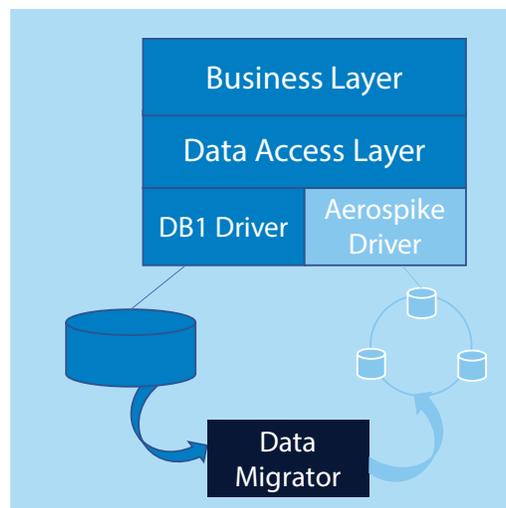
| Before | Intermediate | Post Migration |
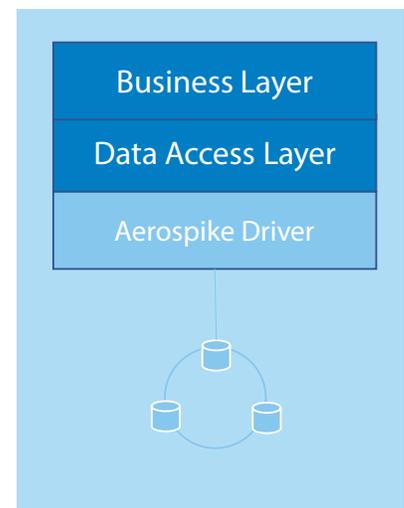
1. Financial services which need to achieve super low latency response decide to replace underlying relational database with Aerospike. The biggest challenge in utilizing Aerospike in an existing Financial Services application is design the most suitable data model to achieve the latency and reduce the data size footprint of the resulting application.

2. Next, the Data Access Layer would need to support both the underlying databases till the final transition.

3. Before making the final switch, we need a fast utility to transform, load data and verify the migrated data into Aerospike. It is important for Financial services organizations to minimize this down time and strategizing so that end customers do not experience the glitch. Custom migration utilities are usually developed as the currently available tools do not support Aerospike and are not cost effective for one-time use.

Motivation to adopt Aerospike across the FS Industry

With public clouds offering SSDs, the SSD optimized performance can be utilized in these cloud native applications

Aerospike is being considered by multiple industries to replace the existing NoSQL and some SQL databases.

Following is summary of the benefits gained by using Aerospike:

Reduced number of instances: Case studies indicate that (reference: Playtika case study) in some scenarios the number of instances required is half the number required with other NoSQL databases

Improved Performance: With the usage of in Memory/SSD Storage, some industries achieved up to 50% improvement in performance

Simplified clustering: Aerospike is comparatively easier to configure. This causes significantly lower effort in managing the cluster

Auto Balancing: Adding and removing nodes is easier and due to auto balancing, there is no manual intervention required

TCO: All these factors lead to a reduced Total Cost of Ownership.

Aerospike Decision Tree

Here is a simple representative decision diagram to help you decide if Aerospike is the right choice for you. Practically there might be more factors specific to the case.
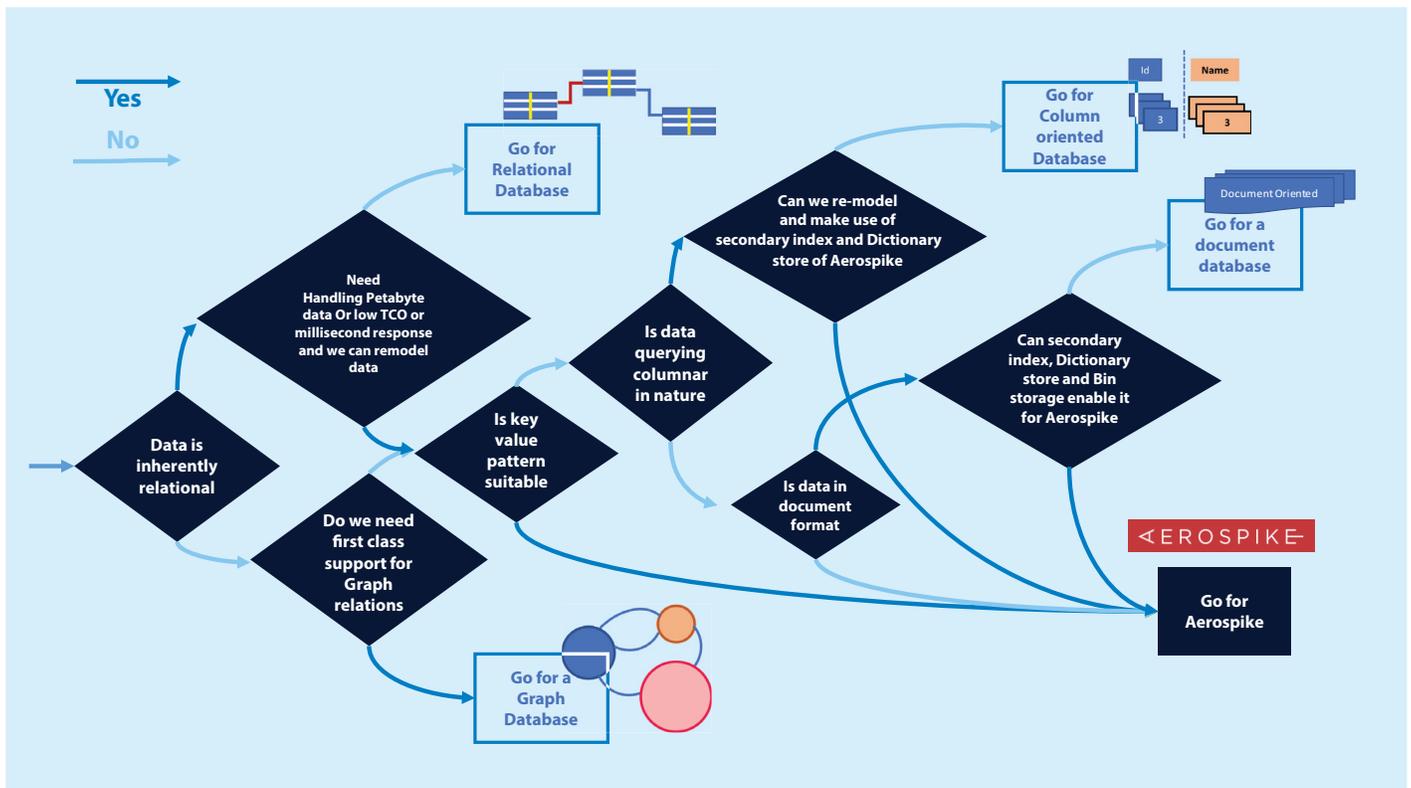
Author acknowledges that Aerospike image s taken from https://en.wikipedia.org/wiki/Aerospike_(database) *and is licensed under* CC BY-NC-ND

How Aerospike Stands Out

	Aerospike	Redis	MongoDB	Couchbase	Cassandra
Short Description	Flash Optimized, Multi-Master In-Memory Datastore	In-Memory Datastore mainly used for cache and message broker	Widely used Document Database	JSON-based document store derived from CouchDB with a Memcached-compatible interface	Distributed wide column store, No SQL database designed to handle large amount of data.
Primary DB Model	Key-Value	Key-Value Store	Document Store	Both Document and Key-Value	Key-Value with Column Family
Hardware/ OS	Supported only on Linux	Linux, OS X, Solaris and Windows	Linux, OS X, Solaris and Windows	Linux, OS X, Windows	Runs on Java. Hence supported on all Major platforms like Linux, Windows, Solaris and OS X.
Secondary Index Support	Yes	Yes	Yes	Yes	Yes
Ability to fetch/ Update Partial records	Has a concept of Bin. Can selectively update only the required Bin using all available drivers.	No inherent support. It needs to be done using LUA script.	Due to Document Oriented Nature, can selectively update a subset of document.	Has concept of Sub-Document which allows subset document update	Possible due to key-value nature and efficient due to column family type of storage.
Server-Side Scripts	User Defined Functions	JavaScript	LUA	Functions and timers in JavaScript	UDF and UDA in Python
MapReduce	Yes	Yes	Yes	Yes	Yes
Clustering	Multi Master	Master-Slave	Master-Slave	Multi Master	Multi Master
Responsiveness	Due to multi master and optimized in-memory, offers the best performance among all.	Offers good responsiveness on pure RAM in-memory, but responsiveness is lower compared to Aerospike.	Offers more features but lower in responsiveness.	Memory-centric system keeps frequently accessed documents, metadata, and indexes in RAM. Speed is comparable to Aerospike.	Cassandra provides better response for write-heavy loads. Read response is slower compared to Aerospike.
Ability to scale out	Comparatively Easiest as data rebalances automatically	Needs manual intervention and re-distribution of data.	Needs manual interventions in rebalancing data.	Rebalancing is easy and comparable to Aerospike.	Rebalancing is easy, but Cassandra is difficult to configure compared to Aerospike.
SSD Support	Yes	No	No	No	Yes
Gartner peer insights score	4.7	4.8	4.5	4.6	
Development ease	Client Libraries available in all major languages Since server available only for X64 Linux, cannot be easily evaluated. Does not have good GUI query utilities.	Client Libraries available in all major languages Can be installed on a dev machine. Like Aerospike, does not have great GUI tools for visualization.	Client Libraries available in all major languages Can be easily installed and tried. Has excellent GUI and third-party tools for visualization.	Client Libraries available in all major languages Can be installed on a dev machine. Has Web interface GUI and third-party tools for visualization.	Client Libraries available in all major languages Server based on Java Hence, easy to install on many platforms. Has rich GUI like DevCenter, DataStax Studio.
Summary	Offers the best responsiveness especially with SSD and low TCO.	Best use for Cache, rebalancing difficult.	Lower in responsiveness compared to Aerospike. Preferred where low response is acceptable but higher feature set is desirable.	Used where document db features required with multi master. Compared to Aerospike, low responsiveness.	Cassandra – like Hadoop can use low-cost commodity hardware, offers better response for high write volume and can handle larger records. However, is not as responsive (~1ms response) as Aerospike.

Conclusion

Aerospike is a leader in next generation financial services and real time NoSQL data solutions for any scale. It is an excellent database for a recommendation engine. Aerospike's Hybrid Memory Architecture unlocks full potential of modern hardware, delivering value from vast amount of data at edge. Aerospike helps financial organizations who are looking to move operational and delivery models from physical to digital platforms. Aerospike can be deployed in areas like digital payment gateways, shopping carts, fraud detections, personalized experience to customers and many more.

Aerospike is the only key value, NoSQL database that delivers:

- More opex cost reduction

- Predictable performance from TBs to PBs

- Server footprint reduction

Key features like large lists (for efficiently recording behavior), optimized Flash support to handle datasets from terabytes to petabytes, queries and aggregations for real-time reporting, and strong support for languages such as Python and Go is contributing to make Aerospike industry leader.

About the Authors

Dipali Khire
Senior Technology Architect, FSSTAR, Infosys Limited

Dipali works as a Technology Architect for USA based capital markets domain client. Her total 18 years of experience in .NET technology includes architecting, designing, and delivering highly available, scalable, and resilient solutions for the client. She can be reached at dipali_khire@infosys.com

Bhaskar Shanbhag
Technology Architect, FSSTAR, Infosys Limited

Bhaskar works for Application Modernization in finance domain over DotNet stack. He has over 17 years of experience in Technology Architecture, Solution Design & Delivery. He can be reached at bhaskar_ds@infosys.com

Infosys®
Navigate your next

For more information, contact askus@infosys.com

Infosys.com | NYSE: INFY   Stay Connected