

NOSQL OPTIMIZED FOR TIME SERIES AND IoT DATA

Internet of Things (IoT) is changing the way companies interact with their customers and manage their data. According to research conducted by Accenture, by 2030, IoT will add \$14.2 trillion to the global economy. Most industries will transform themselves with companies like The Weather Company already leading the way. As our homes, cars, fitness, and even health monitoring are getting “smarter” with Internet-connected devices, enterprise applications are collecting the resulting time series data.

What is time series data? It is data generated across time that requires some level of aggregation and analysis to be useful. Time series data is any data that has a timestamp, like IoT device data, stocks, commodity prices, tide measurements, solar flare tracking, and health information.

While storing time series data is not new, the need to collect and analyze massive amounts of sequenced, often unstructured data from thousands or more devices is a new and growing requirement.

Basho Riak® TS is a distributed NoSQL database. Its architecture is optimized for fast reads and writes of time series data. Riak TS provides resiliency, massive scalability, and can be operationalized at lower costs than traditional relational databases while being easy to manage at scale. Riak TS includes support for Apache™ Spark for in-memory analytics.

TIME SERIES DATA IS A SEQUENCE OF DATA POINTS COLLECTED AT REGULAR INTERVALS OVER A PERIOD OF TIME.

RIAK® TS BENEFITS

OPTIMIZED FOR TIME SERIES DATA

Read and write performance is optimized specifically for time series data. Easily analyze time and location data.

RESILIENCY

Ensures your IoT or time series application is always available for both read and write operations.

MASSIVE SCALABILITY

Easily scale using commodity hardware as devices or users increase. Near-linear performance increase as your data grows.

OPERATIONAL SIMPLICITY

Simple setup for faster ROI. Easily add nodes to the cluster to scale as your needs grow.

FAULT TOLERANCE

A masterless, multi-node architecture ensures no data loss in the event of network or hardware failures.

FAST QUERIES

Optimized range queries, built-in aggregations, and time series data co-location make queries faster and easier to run.

SIMPLIFIED DATA MODEL

When storing structured or semi-structured data, the data can be typed and have a defined schema.

MULTI-CLUSTER REPLICATION

Replicate data across the datacenter or across the globe to ensure time series applications are always available.

LOWER TOTAL COST OF OWNERSHIP

Lower cost to operationalize than traditional relational databases.

TIME SERIES AND IoT APPLICATION CHALLENGES

Companies rely on data to drive strategic decisions related to cost reduction, business process optimization, customer analytics, profitability, and risk analysis. Increasingly, this data is sequenced data collected from sensors located across the globe but it also comes in the form of financial and economic indicators and even from scientific observations. It is a massive amount of data.

Collecting, storing, accessing, and analyzing this data with traditional databases is often not possible. Applications need to scale out, up, and down predictably and linearly as your data grows. The challenge of time series data is that reads and writes to the database must be fast, reliable, and scalable. This requires a NoSQL database optimized for times series data.

TIME SERIES USE CASES

INTERNET OF THINGS (IOT) / SENSOR / DEVICE DATA

Connecting smart devices in our homes to provide better service and save money is helping drive the growth of the Internet of Things (IoT). Examples include: Utilities that have meters creating billions of data points a year and companies like The Weather Company managing 20 terabytes of new data per day. With such high volumes of data, it can be challenging to find a simple, scalable solution to store and access this data easily.

Riak TS is a key/value store that easily scales using commodity hardware. It supports rapid ingestion of time series data from connected devices through extremely fast reads and writes. Riak TS enables application processing of this data to generate actionable information. It is designed to scale horizontally with commodity hardware, making it easy for administrators to add capacity without creating complex sharding.

METRICS / LOG ANALYTICS

Across industries, companies must store and analyze vital metric and log data. Whether your data is system logs, gaming user logs, weather sensor logs, or health device alerts, you need fast storage, retrieval, and analysis of your data. Riak TS ingests and stores your most vital metrics and log data, and makes it easy to retrieve and analyze your data over time.

EDGE DEVICE ANALYTICS

Traditionally, data is analyzed at the core of your network, but with the growth of IoT sensors and devices, data must be analyzed closer to its source and aggregated for core analysis. From cruise ships to health monitoring to system utilization, edge analytics create a better user experience and faster response times. Riak TS requires fewer hardware resources for the same computational power, making it an ideal choice for edge analytics, plus Riak TS makes it easy to do analysis using SQL range queries.

TIME-STAMPED DATA FEEDS

Time-stamped data feeds are made up of any data that contains a time stamp, and are used across industries. For example, an eCommerce application would want to know the total value and delivery location of an order over time; a gaming application will track user actions in a video game; financial apps monitor stock sale prices and trading volumes; and IT team need to know IT metrics like SLAs, API performance, and system metrics.

Data co-location makes it fast to store time series data. This, along with the ability to create tables and perform SQL queries, makes it fast to analyze your time series data. Riak TS is operationally easy to use and allows you to add capacity on demand using commodity hardware. There is no need for complex data sharding.

HOW IS TIME SERIES DATA DIFFERENT?

- Data location matters, it impacts performance
- Data needs to be easy to retrieve with range queries
- Data often has very high write volumes
- Data must eventually be rolled up and compressed

RIAK TS OPTIMIZATIONS FOR TIME SERIES DATA

OPTIMIZED DEPLOYMENT

SIMPLIFIED DATA MODELING

FAST QUERIES AND ANALYTICS

RIAK TS FOR TIME SERIES DATA

Riak TS is a distributed NoSQL database architected to meet your application time series data storage and retrieval needs.

RIAK TS FEATURES

Riak TS has the same distributed systems functionality as Riak KV plus optimizations for high performance reads and writes of time series data.

RESILIENCY

Riak TS has a masterless architecture and automatically replicates data to ensure that your data is always available. This is especially important when ingesting potentially millions of time series data points.

SCALABILITY

Riak TS groups and stores time series together and automatically distributes replica data around the cluster. The scale-out architecture lets you add capacity seamlessly using commodity hardware for near-linear performance improvement.

OPERATIONAL SIMPLICITY

Riak TS allows you to add machines to the cluster easily, without a large operational burden. Data is automatically and uniformly distributed across the cluster with time series data stored optimally on disk; no need to shard your data.

DATA CO-LOCATION

Riak TS uniquely ensures related time series data is co-located on the same physical storage on the same vnode. This allows you to easily analyze temporal or geolocated data.

SQL COMMANDS

Using Riak TS, semi-structured data can be stored in a schema with defined fields making data easy to read using SQL queries. Data co-location and range queries makes reading your data fast.

SQL RANGE QUERIES

Leverage your existing knowledge by writing SQL queries to analyze your time series data.

AGGREGATIONS

Over time, individual data points are less valuable than aggregate data. When querying large data sets, built-in aggregation dramatically reduces the time it takes to build and run queries.

ROBUST APIS AND CLIENT LIBRARIES

PBC and HTTP APIs provide developer flexibility to meet your application needs. Supported languages include: Java, Ruby, Python, Erlang, and Node.js.

MULTI-CLUSTER REPLICATION

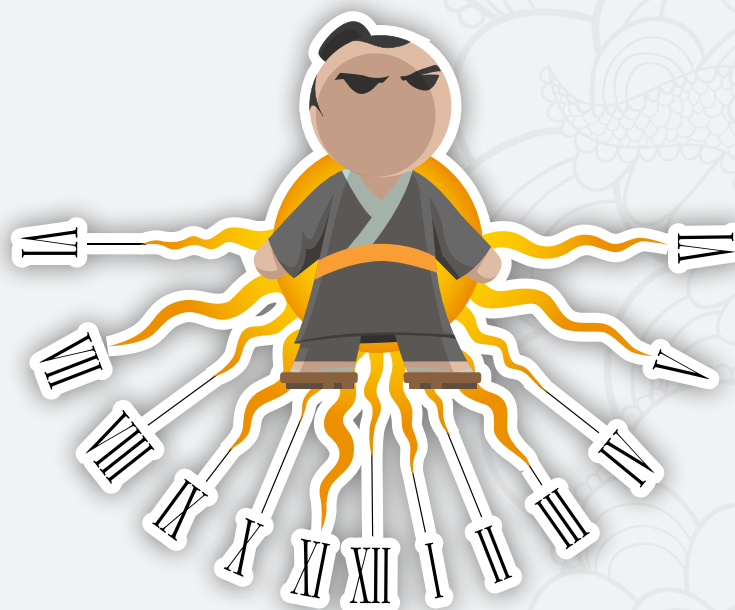
Scale your time series installation by deploying multiple clusters, and prepare your site for disaster recovery, data geolocation, and more by replicating clusters across your datacenter or around the world.

APACHE SPARK CONNECTOR

Seamlessly integrate with Apache Spark to ensure easier and faster operational analysis of time series data.

24 / 7 CUSTOMER SUPPORT

Riak TS Enterprise includes 24 / 7 access to Basho's Client Services team, including 1-hour response time for emergency production help. Basho's support team has extensive experience with Riak TS installations and has worked on some of the largest Riak clusters in the world. Enterprise licensees have unlimited access to that experience and knowledge. Basho provides SLAs based upon the severity of the issue with 24x7 coverage.



“ At The Weather Company, we manage 20 terabytes of new data a day, including real-time forecasting data from over 130,000 sources. The sheer volume of time series data requires a database that can efficiently and reliably store and query time series data. Riak TS delivers on this need and allows us to perform the associated queries and transactions on time series data, while maintaining high availability and scale.

– Bryson Koehler, Executive Vice President and CIO, The Weather Company

GET STARTED

If you are interested in more information and would like to discuss your possible use case, please contact us at techtalk@basho.com. For more information on Riak TS, visit www.basho.com/riak-ts. To learn more about Riak KV, our support and other services we offer visit www.basho.com or follow us on Twitter at www.twitter.com/basho.

ABOUT BASHO TECHNOLOGIES

Basho, the creator of the world's most resilient databases, is dedicated to developing disruptive technology that simplifies enterprises' most critical distributed systems data management challenges. Basho has attracted one of the most talented groups of engineers and technical experts ever assembled devoted exclusively to solving some of the most complex issues presented by Big Data and IoT. Basho's distributed database, Riak® KV, the industry leading distributed NoSQL database, is used by fast growing Web businesses and by one-third of the Fortune 50 to power their critical Web, mobile and social applications. Built on the same foundation, Basho introduced Riak TS, which is the first enterprise-ready NoSQL database specifically optimized to store, query and analyze time series data. The Basho Data Platform helps enterprises reduce the complexity of supporting Big Data applications by integrating Riak with Apache Spark, Redis, and Apache Solr.



SEATTLE - HEAD OFFICE

10900 NE 8th Street
Suite 1580
Bellevue, WA 98004
617.714.1700

WASHINGTON, D.C.

12930 Worldgate Drive
Suite 120
Herndon, VA 20170
617.714.1700

LONDON

Fourth Floor
South Warwick House
65/66 Queen Street
London, EC4R 1EB
+44 020 3201 0032

PARIS

6th Floor
105 rue Anatole France
Levallois-Perret, Paris
92300 France
+33 1 73 44 66 71

TOKYO

Basho Japan KK
NK7 Building 3rd Floor 2-9
Yotsuya Shinjuku-ku
Tokyo, Japan 160-0004
03-5953-1780