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 key/value architecture is optimized for fast reads and writes of time series data. Riak TS provides high availability, massive scalability, and can be operationalized at lower costs than traditional relational databases while being easy to manage at scale. Riak TS is integrated with Basho Data Platform (BDP), which includes support for Apache™ Spark (in-memory analytics), Redis (caching), and Apache™ Solr (distributed search queries).

**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.

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

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

**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 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.

**LOWER TOTAL COST OF OWNERSHIP**
Lower cost to operationalize than traditional relational databases.

**TIME SERIES DATA IS A SEQUENCE OF DATA POINTS COLLECTED AT REGULAR INTERVALS OVER A PERIOD OF TIME.**
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) AND CONNECTED 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.

FINANCIAL TIME SERIES DATA

Time series data not only comes from devices but is also generated in our financial systems in the form of stock market indices, commodity prices, unemployment numbers, and many other financial and economic indicators. Time series data can be used to see how a given asset, security, or economic variable changes over time or how it changes compared to other variables over the same time period.

Riak TS is uniquely architected to handle financial time series data. It enables related data to be stored, queried, and analyzed together to optimize the performance of reads and writes.

SCIENTIFIC TIME SERIES DATA

Scientific fields are also increasing their collection and analysis of large-scale time series data. For example, understanding the relationship between El Niño weather patterns and fish populations requires measuring and comparing air pressure changes in the Pacific Ocean against the number of new fish reported by fisheries and research expeditions.

To perform this type of analysis, you must be able to run queries of time series data at scale, as well as validate your data while it’s being collected to assist with data integrity and compliance. You also need a strong and familiar query language to quickly analyze your data without resorting to complex ETL tricks to do ‘off-box’ analysis with custom tools.

Riak TS provides all of this critical functionality and more so that you can focus on understanding and analyzing your data instead of being overwhelmed by the data tsunami.

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 DATASHEET

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.

HIGH AVAILABILITY
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.

LOW LATENCY
Riak TS is designed to store data and serve requests predictably and quickly, even during peak loads.

FAULT TOLERANCE
Riak TS is fault tolerant so you can lose access to nodes due to network partition or hardware failure and never lose 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.

COMPOSITE KEYS
Composite keys (time, geohash, and data family) can be used to define sort order on disk for faster read performance.

SIMPLIFIED DATA MODELING
Key/value NoSQL data model provides flexibility. The DDL (Data Definition Language) for table and field definitions supports both structured and semi-structured data. Since time series data can be typed, developers can define their own schema.

RANGE QUERIES (SQL LIKE)
Leverage your existing knowledge by writing SQL like queries to analyze your time series data.

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

ROBUST APIS AND CLIENT LIBRARIES
PBC and HTTP APIs provide developer flexibility to meet your application needs. Supported languages include: Java, Ruby, Python, Go, Erlang, Node.js, and .NET.

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 Basho Data Platform and Riak KV, visit www.basho.com or follow us on Twitter at www.twitter.com/basho.

ABOUT BASHO TECHNOLOGIES

Basho Technologies, Inc. is a distributed systems company dedicated to developing disruptive technologies that simplify enterprises’ most critical 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 scaling distributed systems and enjoys a large and growing following among influential programmers, architects and academics.

Basho’s distributed database, Riak KV, the industry leading distributed NoSQL database, and Basho’s cloud storage software, Riak S2, are used by fast-growing Web businesses and by one-third of the Fortune 50 to power their critical Web, mobile and social applications, and their public and private cloud platforms. The Basho Data Platform was recently introduced to help enterprises control and simplify distributed Big Data. Basho is the organizer of RICON – a distributed systems conference.

Riak is the registered trademark of Basho Technologies, Inc. The trademarks and names of other companies and products mentioned herein are the property of their respective owners.