Techstrong Research

PulseMeter

Sponsored by influxdata®

Drivers of Time Series Database Adoption

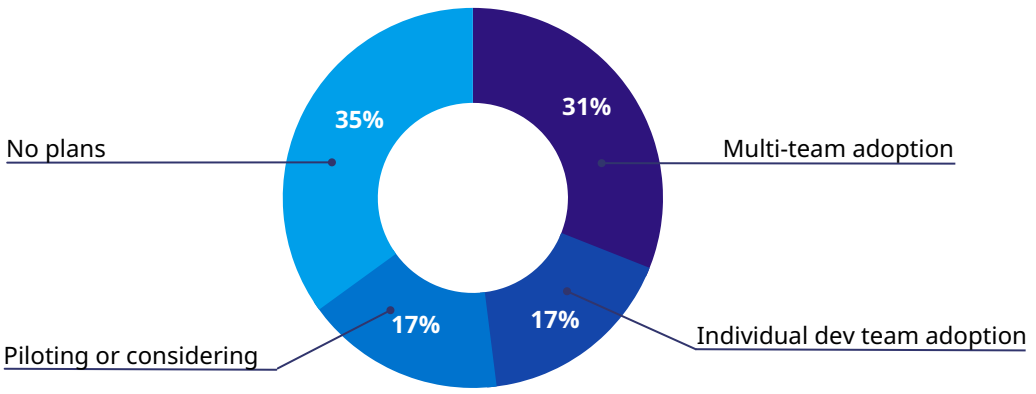
With the unprecedented growth in time-based data generated by applications, systems, networks, security, sensors or virtually any piece of software or device, putting these vast amounts of data to use quickly overwhelms the capabilities of most general-purpose SQL and NoSQL database technologies. Time series databases are particularly well suited to handling large volumes of linear time-based data in use cases such as sensor data acquisition, tracking and analyzing data across multiple periods, and analyzing systems behaviors and situations requiring near-real-time data insertion and retrieval.

Time series databases and data platforms are optimized for high-speed data ingestion, highly efficient queries, analytics, data visualization and temporal comparison due to its high-speed columnar data ingestion, high-performance queries with low latency, faster analytics and high data compression storage resulting in cost reductions on real-time data. Best-of-breed solutions frequently rely on open-source technologies such as Parquet due to its combined benefits of I/O and compute reduction, per-column encoding and storage schemes and bulk data retrieval. In addition, time series databases provide a wide range of integrations with open source, cloud services and commercial software data sources and output data to dashboards and analytics tools.

In late 2022, Techstrong Research polled our community of DevOps, cloud-native, cybersecurity and digital transformation readers and viewers to take their pulse on using time series database technologies. Respondents' use of time series technologies was significant, with 48% indicating that time series data technology is actively used in their data platform strategy. But we also found that respondents encountered some challenges. Managing large volumes of data (20%), scaling data infrastructure (18%) and performing analytics were among the top three challenges of working with time series data. Unsurprisingly, technical expertise (28%) was the most significant roadblock in deploying time series data technologies, as innovative solutions typically require specialized knowledge.

Where are you on your time series journey?

48% of respondents indicated they actively use time series data technology in their data platform strategy.

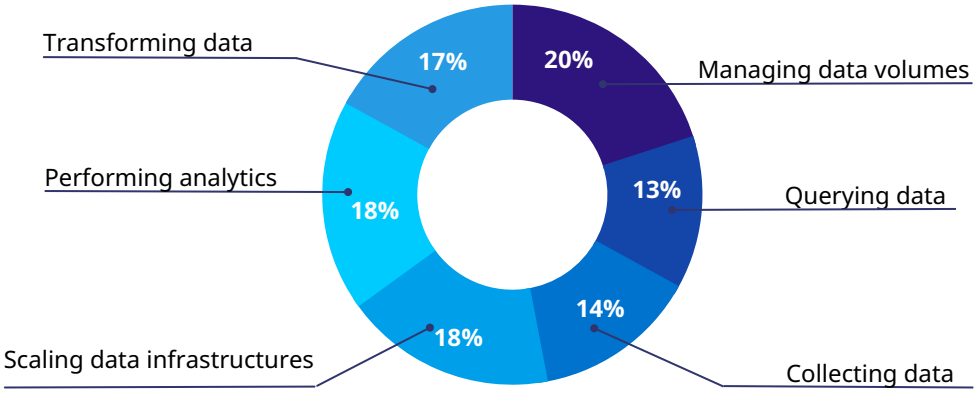


What do you consider the top benefits of collecting, storing and analyzing time series data?

Real-time data (33%), predictive intelligence (23%) and deeper observability (26%) represent the top benefits of time series data use.

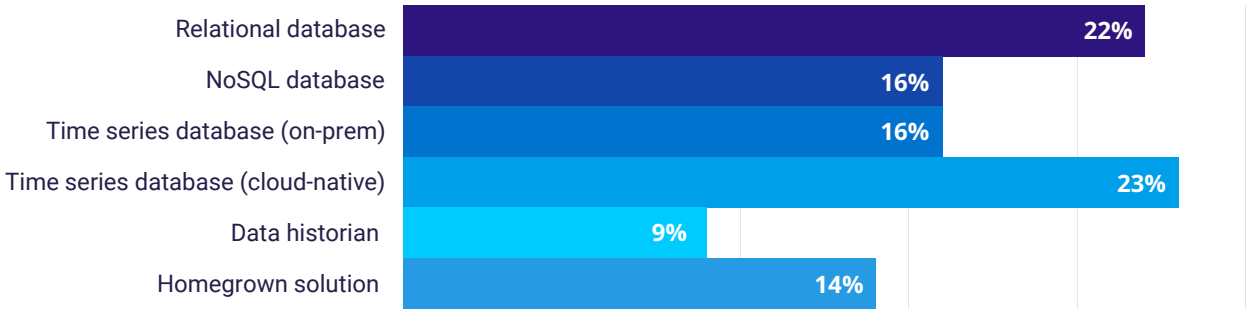
What are your primary challenges when working with time series/timestamped data?

Given the increased use of cloud and distributed computing, it's unsurprising that managing data volumes and scaling infrastructure are top challenges.



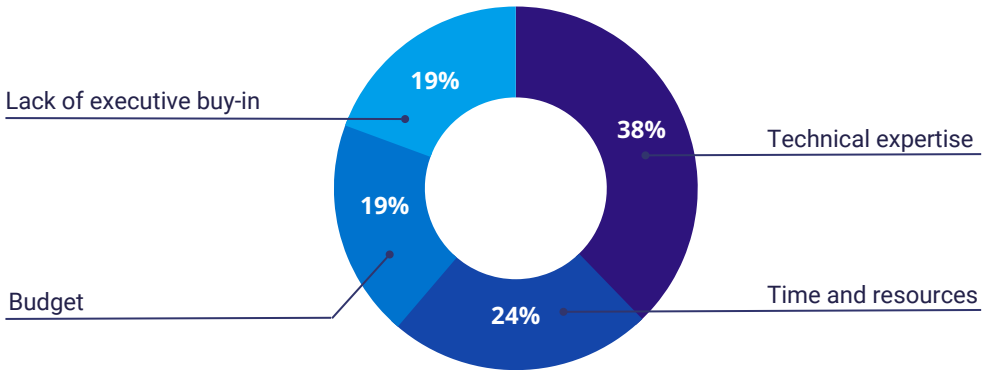
What data stores do you use (or plan to use) for your time series data?

39% indicated they use purpose-built time series technology in the cloud or on-premises. Others are force-fitting time series data into general-purpose data stores.



What is your biggest roadblock to deploying a time series database or platform?

Emerging technologies need an ecosystem to become established. Lack of expertise and time/resources present the top challenges to widespread adoption.



Techstrong Research Analyst View

Time series databases are showing their strengths by carving out new use cases, including real-time measurement of SLAs and customer experience metrics, performance and utilization of cloud services and cloud-native infrastructure, greater visibility into applications and performance, real-time monitoring of large service provider networks and system behavior data captured via Kubernetes in distributed cloud-native applications.

These use cases generate millions—sometimes billions—of time series data points each minute, requiring a scalable, high performance solution. Log aggregation and analysis may provide for the centralized collection, correlation and use of time series data, but the analysis and performance capabilities of these technologies with time series data present clear limitations. Distinct from traditional database technologies, time series databases are purpose-built for rapid, high-volume data ingestion, high data storage compression, low-latency query performance and transmission of large amounts of data between systems. Modern-day solutions build upon a robust ecosystem of proven open-source technologies such as Apache Parquet for fast processing and definition of columnar data coupled with high storage compression, Apache Arrow for high performance, in-memory computing of columnar data, Arrow Flight for bulk dataset transfer and DataFusion's extensible query execution framework.

Time series databases and platforms are available in many offerings, from the open-source Apache Cassandra to commercial offerings and can be deployed on-premises or as cloud- and SaaS-based offerings. Commercial time series offerings typically provide increased performance, ease of adoption through SaaS deployment and integration with a wide variety of programming languages and other third-party services and platforms. They also offer training and support to overcome skill gaps with time series technologies.

Looking beyond data management, performance, scalability and real-time analytics, we see three significant trends for the future of time series databases:

1. Expanded integration with non-traditional data sources, including graph databases, unstructured data sources and API access to SaaS and external online data services.
2. Embedded and third-party artificial intelligence and machine learning (AI/ML) technologies and AIOps platforms for advanced predictive analytics, contextual insights and automation.
3. Integration with edge systems, including cloud-native architectures, to quickly and efficiently work with time series data created at the edges of multiple networks.