Advanced Database IA-1

Time-Series Databases

Group 15 : SY-IT (B3)

16010423076 - RITESH SUDHIR JHA

16010423077 - RITWIK RANJAN MOHANTY

16010423078 - SANJANA SUPRIYA DAS

Introduction of Time-Series Databases

Time-series databases are designed to efficiently manage and query time-series data, which is critical in various domains such as scientific research, industrial monitoring, and financial markets. These databases optimize storage, retrieval, and analysis of time-dependent data, ensuring high performance and scalability.

This report summarizes three research papers that explore different aspects of time-based databases: optimizing large-scale time-series data storage in industrial research, developing a time-series database solution for astronomy, and evaluating modern time-series database technologies for performance improvements.

Research Paper 1 : EAST Plant Data Storage System Based on IoTDB Time-Series Database

This paper explores how the EAST (Experimental Advanced Superconducting Tokamak) plant optimized its data storage system by transitioning from MySQL to IoTDB, a specialized time-series database. The key contributions include:

- **Challenges of Traditional Relational Databases:** The existing MySQL-based system struggled with handling billions of records, leading to slow queries and inefficiencies.
- **Introduction of a MySQL-IoTDB Hierarchical Mechanism (MIHM):** This system seamlessly integrates MySQL with IoTDB, transferring performance burdens to the TSDB.
- **Performance Improvements:** The new system increased data writing speed by **20 times** and large-scale query execution by **100 times**, enhancing EAST's ability to handle massive experimental data.
- **Scalability and Data Management:** The transition ensures a scalable and flexible data storage architecture that supports future experimental needs.

By adopting IoTDB, the EAST plant achieved significant improvements in data retrieval efficiency, making it a strong case study for industrial-scale time-series data management.

Research Paper 2 : LCGCT: A Light Curve Generator in Customizable Time-Bin Based on Time-Series Database

This research focuses on LCGCT, a time-series database solution for astronomy, aimed at optimizing light-curve generation in time-domain astronomy. Key highlights include:

- **Challenges in Astronomy Data Storage:** Traditional relational databases fail to handle the high cardinality and large data volumes required for time-series astronomical observations.
- **Development of LCGCT:** A tool that allows customizable time-bin light-curve generation using a time-series database, significantly improving data retrieval speed.
- **Storage and Query Performance Gains:** Compared to a PostgreSQL-based approach, LCGCT reduces storage requirements by **75%** and improves query speed by **3 times**.
- **Impact on Time-Domain Astronomy:** The tool enhances data processing efficiency for transient object detection, variable stars, and other astronomical studies.

This paper demonstrates how time-series databases can transform astronomical research, ensuring efficient storage and rapid access to vast observational datasets.

Research Paper 3 : Performance Evaluation of Modern Time-Series Database Technologies for ATLAS Monitoring

This paper evaluates modern time-series database solutions for handling massive amounts of operational monitoring data generated at CERN's ATLAS experiment. The major findings include:

- **Limitations of Traditional Systems:** The existing P-BEAST system, a custom-built time-series database, was limited by vertical scaling and outdated technology.
- **Evaluation of New TSDBs:** The study compared **InfluxDB** and **ClickHouse** against P-BEAST to determine their suitability for high-frequency operational data storage.
- **Performance Benchmarks:** The tested databases demonstrated:
    - Faster write speeds, achieving up to **500,000 metrics per second**.
    - Improved query efficiency, critical for real-time monitoring.
- **Scalability and Reliability Improvements:** Horizontal scalability ensures better fault tolerance and more efficient resource utilization.

By adopting modern time-series databases, large-scale monitoring systems can improve data ingestion, analysis, and retrieval, making real-time insights more accessible.

Conclusion

These three research papers highlight the significance of time-series databases in handling vast amounts of time-dependent data across different domains:

- **Industrial Research:** IoTDB enhances data storage and retrieval efficiency in fusion experiments.
- **Astronomy:** LCGCT revolutionizes how astronomical time-series data is processed and analyzed.
- **Scientific Monitoring:** Modern TSDBs like InfluxDB and ClickHouse provide scalable and high-performance solutions for real-time operational monitoring.

Together, these studies demonstrate how time-based databases are transforming data management, ensuring faster, more scalable, and efficient data processing in critical scientific and industrial applications.

References

1. Yang, G., Wang, F., & Fan, D. (2024). *EAST Plant Data Storage System Based on IoTDB Time-Series Database*. Fusion Engineering and Design.
2. Zhang, Z., Xu, Y., & Cui, C. (2024). *LCGCT: A Light Curve Generator in Customizable Time-Bin Based on Time-Series Database*. Astronomy and Computing.
3. Vasile, M., Avolio, G., & Soloviev, I. (2023). *Performance Evaluation of Modern Time-Series Database Technologies for the ATLAS Operational Monitoring Data Archiving Service*. IEEE Transactions on Nuclear Science.

Individual Contributions

- **16010423076 - Ritesh Sudhir Jha**

  Poster summary of '*EAST Plant Data Storage System Based on IoTDB Time-Series Database'* & Group Report Compilation.

- **16010423077 - Ritwik Ranjan Mohanty**

  Poster summary of '*LCGCT: A Light Curve Generator in Customizable Time-Bin Based on Time-Series Database'*.

- **16010423078 - Sanjana Supriya Das**

  Poster summary of '*Performance Evaluation of Modern Time-Series Database Technologies for the ATLAS Operational Monitoring Data Archiving Service'*.