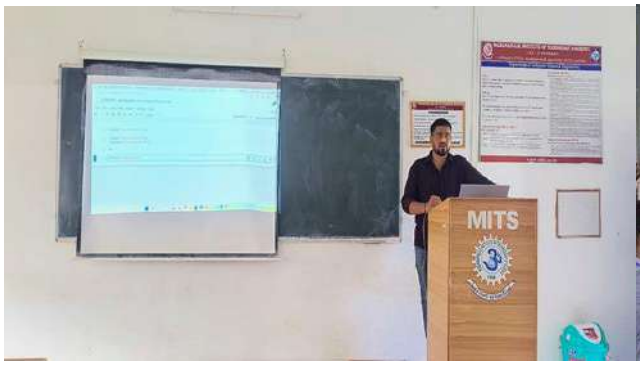
A Report on Alumni Guest Lecture on "Database Scalability: Handling Growing Data in Real-Time Systems" Organized by Department of Computer Science & Engineering on 21st November 2024



Report Submitted by: Mrs. V. Geetha, Assistant professor and Alumni Coordinator, Department of CSE.
Resource Person Details: Mr. Hanny, Working as Software Engineer, TCS, Bangalore.
Venue and Time: WB 308 and 2:00 PM to 3:00 PM.
Mode of Conduct: Offline.
Report Received on 30.11.2024.

The Department of Computer Science & Engineering organized a Guest Lecture on "**Database Scalability: Handling Growing Data in Real-Time Systems** " for II B. Tech students. The inauguration of the guest lecture started at 2:00 p.m. in the WB 308. The dignitaries were Dr. M. Sreedevi, HoD—CSE, Mr. Hanny, Working as Software Engineer, TCS, Bangalore; Dr. R. Kiran Kumar, Alumni Relationship Officer; and Mrs. V. Geetha, Department Alumni Coordinator.

The lecture was started with opening remarks by, Dr. M. Sreedevi who thanked Management for this great initiation of creating an opportunity to invite the Alumni members of the institute and enabling them to interact with the students and enlightening them with the current developments in the corporate world. Dr. R. Kiran Kumar has shown pleasure and promised to conduct many more lectures in future for the benefit of the students.

Mrs. V. Geetha introduced the speaker and invited him to share his valuable experiences to the students. The number of students who participated in the lecture were around 50.

Introduction:

- On 21-11-2024, a guest lecture was delivered on the topic **"Database Scalability: Handling Growing Data in Real-Time Systems"** by Mr. C. Hanny. The lecture focused on the critical issue of managing data growth in real-time systems and the importance of scalability in database architecture. As organizations continue to collect vast amounts of data, especially in industries like e-commerce, social media, and IoT, the ability to scale databases effectively becomes increasingly important.

Key Themes and Concepts

The lecture covered several key aspects of database scalability, including:

Understanding Database Scalability

- The speaker began by defining scalability as the capacity of a database to handle an increasing amount of work, or its potential to accommodate growth. Scalability can be vertical (adding more resources to a single server) or horizontal (adding more servers to distribute the load).
- The speaker emphasized that as data grows exponentially, especially in real-time systems, databases need to evolve from traditional relational models to more distributed and scalable systems.

Challenges in Real-Time Data Processing

- Real-time systems, such as financial transaction systems, social media platforms, and IoT devices, generate enormous amounts of data that must be processed without delay.
- The challenge lies not only in storing this data but also in ensuring that queries and operations can be performed quickly, without degrading system performance. The speaker highlighted key performance issues such as latency, throughput, and concurrency that arise when scaling databases in real-time environments.

Types of Scalability

- **Horizontal Scaling (Sharding):**
  The speaker explained sharding, a method used to distribute data across multiple machines or servers. Each shard stores a portion of the dataset, allowing the database to distribute query load and enhance performance. This method helps reduce bottlenecks caused by single-server reliance and enables handling growing data.
- **Vertical Scaling:**
  Unlike horizontal scaling, vertical scaling involves increasing the capacity of a single server (e.g., adding more CPU, memory, or storage). However, vertical scaling has limitations and might not be sufficient for handling massive data volumes in real-time.
- **Distributed Databases:**
  The lecture delved into the use of distributed databases, which split data across multiple physical machines, ensuring high availability and fault tolerance. Technologies like Apache Cassandra and Google Spanner were mentioned as examples of distributed databases that support real-time data scalability.

Real-Time Data Systems and Technologies

- The speaker discussed various NoSQL database technologies, such as Cassandra, MongoDB, and Redis, that are commonly used for real-time, high-performance applications.
- Event-Driven Architectures: Real-time databases often rely on event-driven architectures to manage data efficiently. This approach is useful in handling incoming streams of data and triggering real-time processing.
- Stream Processing: The lecture covered the growing importance of stream processing systems like Apache Kafka, Apache Flink, and Apache Storm, which enable real-time data ingestion, processing, and distribution.

Database Partitioning and Replication

- The speaker also addressed data partitioning and replication as strategies for improving database scalability. Partitioning allows for splitting a large dataset into smaller, more manageable parts, while replication ensures data redundancy and fault tolerance.
- The use of Master-Slave Replication and Multi-Region Replication in distributed databases ensures that real-time systems can provide continuous data availability across geographical regions.

Monitoring and Tuning

- As databases scale, monitoring becomes critical. The speaker explained how real-time systems require constant monitoring to detect and mitigate issues related to performance degradation, such as slow queries or resource exhaustion.
- Tools like Prometheus, Grafana, and ELK Stack were discussed for monitoring database health and performance in real-time systems.

Best Practices for Achieving Scalability

- The speaker outlined several best practices for achieving scalability in real-time systems:
- Design databases with flexible schemas that can handle varying data types and sizes.
- Use load balancing techniques to evenly distribute traffic across servers.
- Implement caching strategies to reduce the load on databases by storing frequently accessed data in memory (using tools like Redis or Memcached).
- Invest in automated scaling systems that can scale up or down based on demand.

Conclusion

The guest lecture concluded with a discussion on the future of scalable databases in real-time systems. The speaker emphasized that as data volumes continue to grow, both the **complexity** of database management and the need for **efficient data processing** will increase. Future advancements in AI, machine learning, and cloud technologies will likely drive further innovations in the scalability of real-time databases.
The speaker also encouraged attendees to explore emerging technologies in **distributed databases** and **event-streaming architectures** and to consider how they can be applied to solve scalability challenges in their respective fields.

The outcome of the programme:

The program outcomes for students attending a guest lecture on "Database Scalability: Handling Growing Data in Real-Time Systems" could include several educational and skill-based benefits.

- Enhanced Understanding of Database Scalability Concepts.
- Introduction to Advanced Database Technologies.
- Cloud Computing and Managed Database Services.
- Future Trends and Innovations in Database Scalability.

The session is completed at 3:00 PM, and he clarified the queries of enthusiastic young minds with great zeal during the interaction time. The resource person was honored by a token of respectable appreciation by Dr. M. Sreedevi CSE – HoD, Dr. R. Kiran Kumar, Alumni Relation Officer and all department faculty members.

Vote of Thanks:

Mrs. V. Geetha proposed a vote of thanks to the Resource person, HoD and Alumni Relations Officer for attending the function. She extended her thanks to the Principal and the Management for their support in conducting the training.