Databases II
Semester 2025-I
Workshop No. 3 — Concurrency, Parallelism, and Distributed Databases

Eng. Carlos Andrés Sierra, M.Sc.
Computer Engineering
Universidad Distrital Francisco José de Caldas

Welcome to the third workshop of the *Databases II* course! This session is the next step in your course project and focuses on analyzing concurrency, exploring parallel and distributed database solutions, and proposing strategies to increase system performance.

Workshop Scope and Objectives:

- **Concurrency Analysis:** Identify and analyze potential concurrency issues in your project (e.g., lost updates, dirty reads, deadlocks). Propose mechanisms to ensure data consistency and isolation.

- **Parallel and Distributed Databases:** Explore how parallel and distributed database architectures can be applied to your project to improve performance, scalability, and availability.

- **Performance Strategies:** Propose concrete strategies (e.g., sharding, replication, partitioning, parallel query execution) to increase throughput and reduce latency.

- **Continuous Improvement:** Review feedback from Workshop 2 and improve your previous deliverables (architecture, queries, information requirements, etc.) as needed.

Carlos Andrés Sierra, Computer Engineer, M.Sc. in Computer Engineering, Lecturer at Universidad Distrital Francisco José de Caldas.

Any comment or concern regarding this workshop can be sent to Carlos A. Sierra at: *cavirguezs@udistrital.edu.co*.

Methodology and Deliverables:

1. **Concurrency Analysis:**

   - Identify scenarios in your application where concurrent access to data may occur.
   - Describe possible problems (e.g., race conditions, deadlocks) and propose solutions (e.g., isolation levels, locking protocols, optimistic concurrency control).

2. **Parallel and Distributed Database Design:**

   - Propose a high-level design for integrating parallel or distributed database technologies into your system.
   - Illustrate with diagrams how data and queries will be distributed or parallelized.
   - Justify your choices based on your project's requirements (e.g., big data, multi-location, high availability).

3. **Performance Improvement Strategies:**

   - List and explain at least three strategies to improve system performance using parallelism or distribution (e.g., horizontal scaling, data partitioning, replication, distributed transactions).
   - Discuss trade-offs and potential challenges.

4. **Improvements to Workshop 2:**

   - Review and refine your architecture, queries, and information requirements based on feedback.
   - Clearly indicate what was improved or changed.

5. **Delivery Format:**

   - Compile your concurrency analysis, parallel/distributed design, performance strategies, and improved Workshop 2 deliverables into a single PDF.
   - Organize your files in a folder named `Workshop-3` in your course project repository, with a `README.md` referencing each section.

Project Requirements Reminder:

- Fast query execution in a big data context.
- Constant ingestion of data throughout the day.
- Business intelligence module for managerial insights.
- Multi-location data storage and access.
- Recommendation system for products or services.
- High availability and scalability.

Deadline: Thursday, June 19th, 2025, 10:00. Late submissions may affect your grading according to course policies.

Notes:

- All documents must be in **English**.

- Cite any references (articles, tutorials) that influenced your design choices.

- Focus on clarity, completeness, and how your concurrency and distribution strategies support the business and user needs.

Good luck! This step will help you ensure your system can scale, remain consistent, and perform efficiently as your project grows.