Advanced Databases - Project

Grading policy:

Your assignment is to specify and implement each phase of the project described below. You will work in teams of 3 Students. The whole project team is expected to be in class on due date of each phase. All assignments are expected to be done on time. A penalty of gaining maximum 50% of grade will be imposed for each week of lateness(!). The team organizes its own work. Final Grade depends on partial grades for each phase and is calculated as weighted average.

Project Schedule:

Due on meeting	Assignment
2	**1 - Data model (1 point)** You will work with an database application domain of your own choice. First assignment is to prepare the specification of database structure. For this assignment you are to choose a topic of your application domain and formulate its short description. Prepare an diagram describing your data model with attributes names and types, keys and associations between relations. Minimal Requirements: 6 Relations. Each Relation should have at least 4 attributes with different types (CHAR, NUMBER, DATE). Hint: Feal free to combine sample Oracle schemas from: https://www.oracle.com/pl/database/technologies/appdev/datamodeler-samples.html https://docs.oracle.com/database/121/COMSC/installation.htm#COMSC001 The sample schemas cannot be used directly without any custom adjustments of your own. Each group needs to have different database model.
3	**2 - Workload (1 point)** This assignment is to prepare the description of the workload of your database application. The workload consists of transactions that would usually be executed on your database. Transactions changing the data need to cover all elementary operations (insert, update, delete). Transactions querying the data need to read the data from several relations with joins, grouping and subqueries. Minimal Requirements: 6 Transactions (3 queries, 3 changing the data). Each operation should be complex enough to promise the observable execution times.
4	**3 - DBMS (1 point)** Chose a DBMS. Install DBMS. Create a database and its structure. Hint: It is recommended to use Oracle 21c Express edition Oracle XE-downloads
5	**4 - Data (1 point)** Populate the relations with sample data. Prepare the sample data so that it will reflect the real dependencies between the entities. The report from this phase should consist of: Documentation of the volumetry (table with number of rows per each table).

6	**5 - Database workload (1 point)** Prepare the SQL script to run and measure the execution time of the workload. Each transaction should be implemented as a single or multiple SQL expressions.
7	**6 - Query plans (1 point)** Collect and document the query plans for all transactions from the workload. Analyze the query plans and point the most costly operations. Analyze the queries and query plans to find candidates for performance improvements. In case of only trivial query plans review the transactions and rebuild them to get the more complex query plans. If needed populate additional data or apply changes to your database model to achieve more complex query plans. The report from this phase should consist of: Description of final query plans for your transactions
8	**7 - Testing application (1 point)** Prepare an application or scripts to run and measure the execution time of the workload. It is recommended to collect the query plans as well. The results of measuring of the running time of each transaction should be automatically saved in a file or logging database table. The state of the data should be resorted before each execution of the workload so that each execution would have the same initial state of the database. The memory buffers of the database should be flushed before execution of the workload. The amount of data and the complexity of the operations together with database configuration needs to be adjusted in such a way that the running time of the workload would be observable i.e. > second per each transaction. The report from this phase should consist of: Documentation of the execution times of each transaction and the whole workload (table with number of runs, min, max, avg, running times per each transaction). Conclusions. The running times and the volumetry will be verified on the classes (need to be repeatable!).
9	**8 - Indexes (1 point)** Prepare the proposition of indexes for your database. Specify at least 3 indexes. The different types of indexes (b-tree, bitmap, function) should be used for the better grade. Any experiments with e.g. comparing different types of indexes for the same purpose are highly expected. The report from this phase should consist of: Description of performance improvements to be applied with its type, purpose and link to operations expected to be improved.
10	**9 - Indexes (2 points)** Implement indexes. Document the execution times after optimization and compare the query plans of chosen queries before and after optimization. The report from this phase should consist of: Documentation of the execution times of each transaction without/with improvements and the whole workload (table with number of runs, min, max, avg, running times per each transaction). Conclusions including analysis of results of experiments with comparing different optimization tools (if done). The running times will be verified on the classes (need to be repeatable!).
11	**10 - Partitions (1 point)** Prepare the proposition of partitions for your database. Specify at least 3 partitions. The different types of partitions (hash, ranger, value list) should be used for the better grade. For the partitions think about using external memory storage to allocate part of

	the data so theta the reads could be executed by the DBMS in parallel. Any experiments with e.g. comparing different types of partitions for the same purpose are highly expected. The report from this phase should consist of: Description of performance improvements to be applied with its type, purpose and link to operations expected to be improved.
12	**11 - Partitions (2 points)** Implement performance improvements. Document the execution times after optimization and compare the query plans of chosen queries before and after optimization. The report from this phase should consist of: Documentation of the execution times of each transaction without/with improvements and the whole workload (table with number of runs, min, max, avg, running times per each transaction). Conclusions including analysis of results of experiments with comparing different optimization tools (if done). The running times will be verified on the classes (need to be repeatable!).
13	**12 - Columnar store (1 point)** Specify which tables/columns will be stored in columnar store. Declare which workload operations will be affected by this storage. Propose usage for in-memory expressions and join groups. Both types of transactions need to be included (queries, CRUD). Declare the scope of comparisons between different storage and compression methods.
14	**13 - Columnar store (2 points)** Implement columnar storage. Document the execution times with columnar storage and compare with row storage on query plans level. The report from this phase should consist of: Documentation of the execution times of each transaction and the whole workload (table with number of runs, min, max, avg, running times per each transaction). Conclusions including analysis of results of experiments with comparing different storage and comparison methods. The running times will be verified on the classes (need to be repeatable!).