

## Advanced Databases

Due Date	Assignment
2020-10-12(Mon), 2020-10-15(Thu)	<b>Assignment 1 - Database model (spec.)</b> 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 your application domain and formulate its description. Prepare an ERD or class diagram describing your data model. Minimal Requirements: 7 Relations. Each Relation should have at least 5 attributes with different types (CHAR, NUMBER, DATE). Hint: Feel free to combine sample Oracle schemas from: <a href="https://www.oracle.com/pl/database/technologies/appdev/datamodeler-samples.html">https://www.oracle.com/pl/database/technologies/appdev/datamodeler-samples.html</a> <a href="https://docs.oracle.com/database/121/COMSC/installation.htm#COMSC001">https://docs.oracle.com/database/121/COMSC/installation.htm#COMSC001</a> The sample schemas cannot be used directly without any custom adjustments of your own. Each group needs to have different database model.
2020-10-19(Mon), 2020-10-22(Thu)	<b>Assignment 2 - Database workload (spec.)</b> This assignment is to prepare the specification of the workload of your database application. The workload consists of SQL 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: 7 Transactions (3 queries, 4 changing the data). Each operation should be complex enough to promise the observable execution times.
2020-10-26(Mon), 2020-10-29(Thu)	<b>Assignment 3 - Database model (dev.)</b> Chose a DBMS. Install DBMS. Create a database and its structure. 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). Hint: It is recommended to use Oracle 18c Express edition
2020-11-09(Mon), 2020-11-05(Thu)	<b>Assignment 4 - Database workload (dev.)</b> Prepare the SQL script to run and measure the execution time of the workload. 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!).
2020-11-16(Mon), 2020-11-12(Thu)	<b>Assignment5 - Query plans (spec.)</b> Collect and document the query plans for all transactions from the workload. Analyse the query plans and point the most costly operations. Analyse 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
2020-11-23(Mon), 2020-11-19(Thu)	<b>Assignment 6 - Indexes (spec.)</b> Prepare the proposition of indexes for your database. Specify at least 5 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.
2020-11-30(Mon), 2020-11-26(Thu)	<b>Assignment 7 - Indexes (dev.)</b> 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!).
2020-12-07(Mon), 2020-12-03(Thu)	<b>Assignment 8 - Partitions (spec.)</b> Prepare the proposition of partitions for your database. Specify at least 5 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 that 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.
2020-12-14(Mon), 2020-12-10(Thu)	<b>Assignment 9 - Partitions (dev.)</b> 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!).
2020-12-21(Mon), 2021-12-17(Thu)	<b>Assignment 10 - Columnar store (spec.)</b> 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.
2021-01-11(Mon), 2021-01-07(Thu)	<b>Assignment 11 - Columnar store (dev.)</b> 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 compression methods. The running times will be verified on the classes (need to be repeatable!).
2021-01-18(Mon), 2021-01-14(Thu)	<b>12 - Summary (spec.)</b> Summary of results achieved for the optimization methods (indexes, partitions, column storage) compared in the previous phases of the project. Results to be presented in a table with the average execution times and costs of query plans for each transaction from the workload (rows), for no optimization and each optimization method (columns) The table

	should be based on the results of the previous phases of the project. Indicate the best results obtained. Prepare the selection of two methods that give the best results for each transaction.
2021-01-25(Mon), 2021-01-21(Thu)	<b>13 - Summary (dev.)</b> Prepare the configuration of the system with both optimization version proposed in previous phase. Collect runnig times for both versions. Compare the results and chose the method giving the best results for the whole workload. The report from this phase should consist of: Documentation of the execution times of each transaction for both optimization methods (table with number of runs, min, max, avg, running times per each transaction). Conclusions including analysis of results for different optimization methods. The running times will be verified on the classes (need to be repeatable!).