# Stage 3 – Indexing: Project Demo (30%)

**Graded by: Final Score (out of 30%):**

**Project members (student IDs):**

## Data Analysis – 30 minutes

Please hand out one student sheet to each group.

Two tables will be used in this test, “**user**” and “**trans**”.

table user

attr1 is user id, int, **uniform distribution**

attr2 is name, varchar

attr3 is salary, int, **exp distribution**\*

attr4 is IQ, int, **normal distribution**

attr5 is tax, int, **exp distribution**\*

\*salary and tax are correlated

table trans

attr1 is transaction ID, varchar

attr2 is user id, int

attr3 is stock id, int, **uniform distribution**

attr4 is buy price, int, **exp distribution**

attr5 is sell price, int, **exp distribution**

## Query time test – 3 minutes

In order to test if the query time reported by the application is reasonable and correct, the queries below must be executed. The tester must capture the time taken with a stopwatch and compare it with the time displayed by the application.

SELECT \*  
FROM user;

SELECT \*  
FROM user  
WHERE attr1 = 1;

|  |  |  |
| --- | --- | --- |
| **Query** | **Time reported by application** | **Time reported by stopwatch** |
| 1 |  |  |
| 2 |  |  |

## Query testing – 10 minutes

The following queries must be tested and executed correctly. Please write the execution times in section 4.

1. (1%) Equality query

SELECT \*  
FROM trans  
WHERE attr5 > 0;

1. (1%) Range and join query

SELECT COUNT(\*)  
FROM user, trans  
WHERE user.attr1 = trans.attr2 AND user.attr5 > 50000;

1. (1%) Range query

SELECT COUNT(\*)  
FROM user  
WHERE attr3 > 100000 AND attr3 < 200000;

1. (1%) Scan query

SELECT COUNT(\*)  
FROM trans;

1. (1%) Equality and range query

SELECT SUM(attr4)  
FROM user  
WHERE attr3 = 1510503 OR attr5 > 500000;

## Query speed

The speed of each query must be measured by displaying the query’s execution time (in milliseconds or seconds) on the screen. The speed must be written in the following table:

|  |  |
| --- | --- |
| **Query** | **Execution time** |
| 1 |  |
| 2 |  |
| 3 |  |
| 4 |  |
| 5 |  |

The standard for grading each group will be based on the distribution of execution times.

* **(15%)** For each query, the execution time will be calculated using the following rules:
  + **If the query is executed correctly:**
    - The top 20% will obtain full credit for that query
    - Top 40% to 20% get 80% credit for that query
    - Top 60% - top 40% get 60% credit for that query
    - Lower 40% get 40% credit for that query
  + **If the query is not executed correctly:**
    - 0% credit for that query, and the execution time given for this query will be 2 times the maximum execution time for that query in the whole class
* **(15%)** The execution time will be added up for all queries and use the same previous rules.