

**CE/CZ4123/SC4023 BIG DATA MANAGEMENT**

# SEMESTER GROUP PROJECT

Lee Juin <>

Ng Zhi Quan <U2120746>

Tey Kai Seong <>

**SCHOOL OF COMPUTER SCIENCE AND ENGINEERING**

**NANYANG TECHNOLOGICAL UNIVERSITY**

# REPORT FORMAT

## 1 Data Storage

In this section, explain how your program handles and stores the data. You may present your design and experience (whether success or failure) related to:

* How to store the data in the column-store approach[[1]](#footnote-1);
* How to design data columns for efficient processing‡;
* How to read and write the input/output files;
* How to handle exceptions such as empty qualified entries in the code.

## 2 Data Processing

In this section, explain how your program scans the data and finds the values. You may present contents related to:

* How to scan columns according to task conditions;
* How to decide and record the statistics (e.g. the minimum values);
* How to improve the efficiency in scanning columns‡;

## 3 Experiment Result

In this section, present the experimental results that your program successfully complete the tasks. The following contents are compulsory:

* Screenshots that your program executes and outputs results successfully;
* Evaluations that the output results are correct (You may put screenshots in the report comparing your output results with the correct results output by other tools such as Excel).

1

# CONTRIBUTION FORM

Group Number

|  |  |  |  |
| --- | --- | --- | --- |
| **Name** | **Matriculation** | **Detailed Individual Contribution** | **Percentage** |
|  | **Number** |  | **(100% in total)** |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |

**Name and Signature from all group members:**

Name and Signature of Member 1 Name and Signature of Member 2

Name and Signature of Member 3 Name and Signature of Member 4

2

1. Exploration and improvements on these aspects are encouraged. [↑](#footnote-ref-1)