Jan Ludwicki 188899, Adam Sobczuk 1888656

Hive Data Warehouse Project

This warehouse is based on previous worked we had done on Data Warehouse subject.

The Data warehouse is designed for Feeding kids business process. The process of feeding kids in the school canteen is as follows:

The day before cantina workers assess how much food they need to prepare for the next day. The kids each day can go to the school canteen and eat lunch there. The canteen workers prepare a list of students in excel who eat lunch every day. At the end of the day the remaining food is summed up and then if necessary thrown out.

Decisions Made:

Structured and Nested Data

Utilizing structured and nested data types like STRUCT and MAP in the meal table allows for a flexible representation of meal information. This approach enables the storage of courses as a structured entity and additional information as key-value pairs in a map.

Partitioning for Date, Country and Age Category

Partitioning the date_dim table by season, the kindergarten table by country and the student by age_category supports efficient data retrieval based on these key attributes. Partitioning is beneficial for optimizing query performance, especially when filtering on commonly used dimension

Bucketing for Eating Age Meal Serving ID

Using bucketing on the student table by meal_serving_id distributes the data evenly across the specified number of buckets. This can improve query performance for analytics that involve filtering or joining based on age categories.

Storage Formats for Data Files

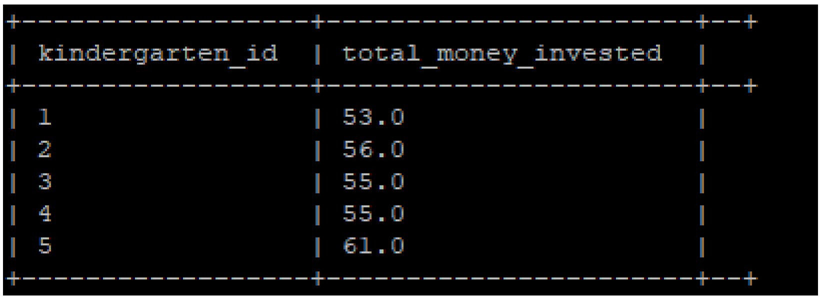
Choosing different file formats (SEQUENCE, PARQUET, ORC, TEXTFILE) for the tables is based on the trade-offs between storage efficiency, query performance, and compatibility with specific tools. For example TEXTFILE allows data modification as it's not a binary format.

External Table for Date Dimension

Creating an external table for date_dim allows flexibility in managing data storage locations. External tables are useful when the data is stored outside the Hive warehouse directory, and the schema can be projected onto the data stored in a different location.

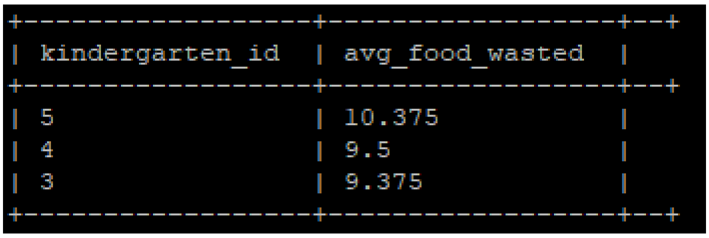
Competency questions:

1.Retrieve the total amount of money invested in meals for each kindergarten in the year 2023.



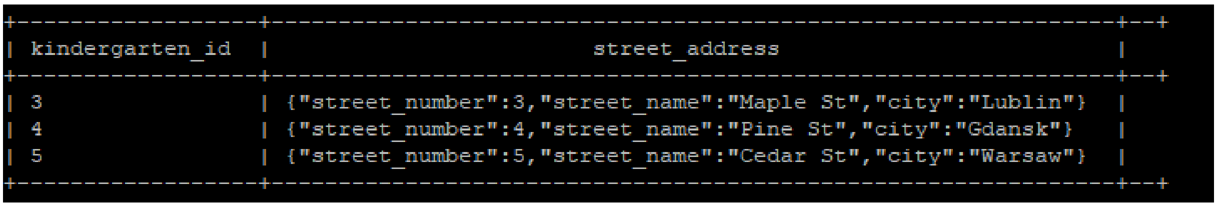
```
+------------------+-----------------------+--+
| kindergarten_id  | total_money_invested  |
+------------------+-----------------------+--+
| 1                | 53.0                  |
| 2                | 56.0                  |
| 3                | 55.0                  |
| 4                | 55.0                  |
| 5                | 61.0                  |
+------------------+-----------------------+--+
```

2.List the top 3 kindergartens with the highest average food wastage per meal.



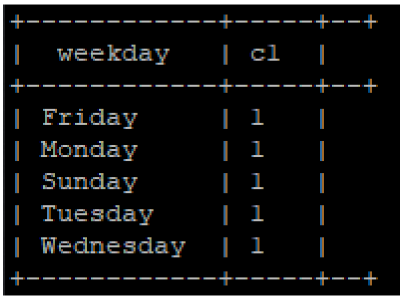
```
+------------------+-------------------+--+
| kindergarten_id  | avg_food_wasted   |
+------------------+-------------------+--+
| 5                | 10.375            |
| 4                | 9.5               |
| 3                | 9.375             |
+------------------+-------------------+--+
```

3.Retrieve the names and addresses of kindergartens in Poland.



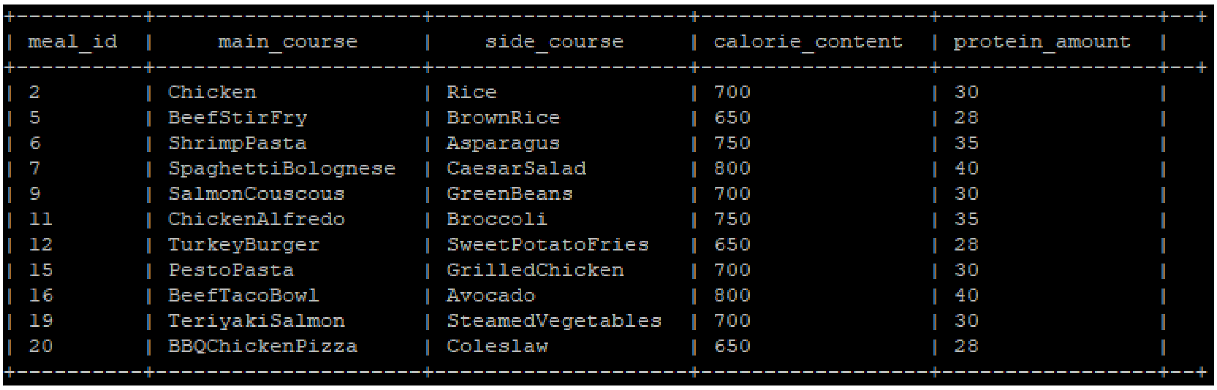
```
+------------------+----------------------------------------------------------------+--+
| kindergarten_id  |                         street_address                         |
+------------------+----------------------------------------------------------------+--+
| 3                | {"street_number":3,"street_name":"Maple St","city":"Lublin"}   |
| 4                | {"street_number":4,"street_name":"Pine St","city":"Gdansk"}    |
| 5                | {"street_number":5,"street_name":"Cedar St","city":"Warsaw"}   |
+------------------+----------------------------------------------------------------+--+
```

4.Find the total number of meals served on each day of the week during the Summer season.



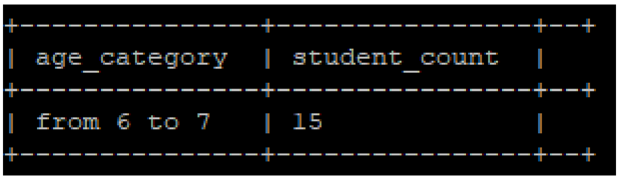
```
+------------+-----+--+
|  weekday   | c1  |
+------------+-----+--+
| Friday     | 1   |
| Monday     | 1   |
| Sunday     | 1   |
| Tuesday    | 1   |
| Wednesday  | 1   |
+------------+-----+--+
```

5.Retrieve the meals with their respective calorie content and protein amounts, filtering for meals with more than 25 grams of protein.



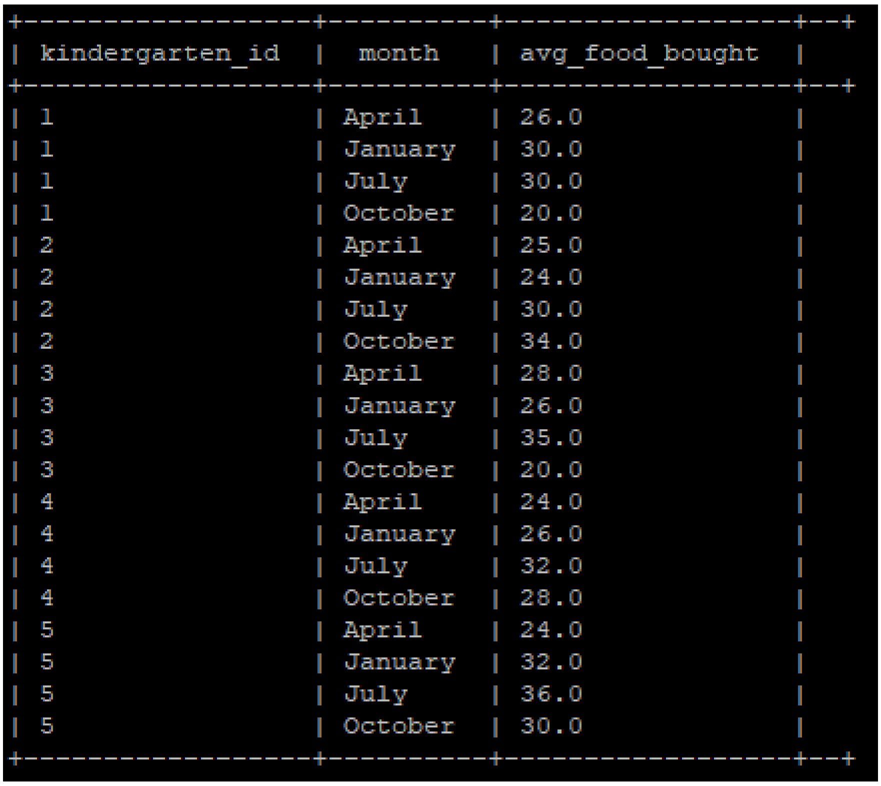
```
+----------+---------------------+---------------------+------------------+------------------+--+
| meal_id  |     main_course     |     side_course     | calorie_content  |  protein_amount  |
+----------+---------------------+---------------------+------------------+------------------+--+
| 2        | Chicken             | Rice                | 700              | 30               |
| 5        | BeefStirFry         | BrownRice           | 650              | 28               |
| 6        | ShrimpPasta         | Asparagus           | 750              | 35               |
| 7        | SpaghettiBolognese  | CaesarSalad         | 800              | 40               |
| 9        | SalmonCouscous      | GreenBeans          | 700              | 30               |
| 11       | ChickenAlfredo      | Broccoli            | 750              | 35               |
| 12       | TurkeyBurger        | SweetPotatoFries    | 650              | 28               |
| 15       | PestoPasta          | GrilledChicken      | 700              | 30               |
| 16       | BeefTacoBowl        | Avocado             | 800              | 40               |
| 19       | TeriyakiSalmon      | SteamedVegetables   | 700              | 30               |
| 20       | BBQChickenPizza     | Coleslaw            | 650              | 28               |
+----------+---------------------+---------------------+------------------+------------------+--+
```

6.Retrieve the age category with the highest number of current students.



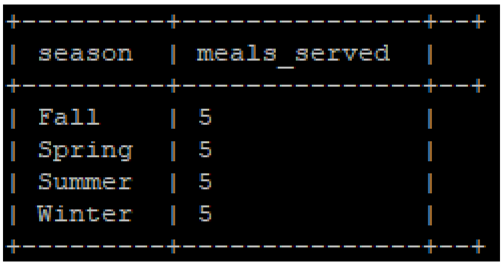
```
+---------------+----------------+--+
| age_category  | student_count  |
+---------------+----------------+--+
| from 6 to 7   | 15             |
+---------------+----------------+--+
```

7.Calculate the average amount of food bought per month for each kindergarten.



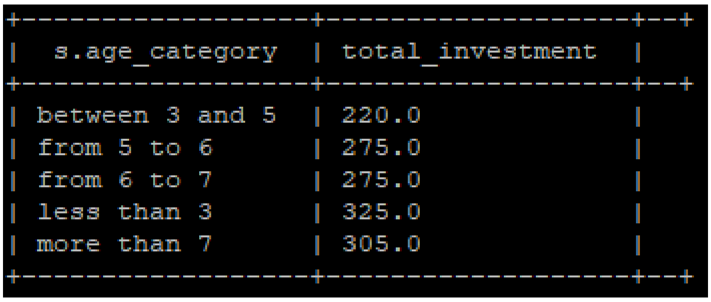
```
+------------------+----------+------------------+--+
| kindergarten_id  |  month   | avg_food_bought  |
+------------------+----------+------------------+--+
| 1                | April    | 26.0             |
| 1                | January  | 30.0             |
| 1                | July     | 30.0             |
| 1                | October  | 20.0             |
| 2                | April    | 25.0             |
| 2                | January  | 24.0             |
| 2                | July     | 30.0             |
| 2                | October  | 34.0             |
| 3                | April    | 28.0             |
| 3                | January  | 26.0             |
| 3                | July     | 35.0             |
| 3                | October  | 20.0             |
| 4                | April    | 24.0             |
| 4                | January  | 26.0             |
| 4                | July     | 32.0             |
| 4                | October  | 28.0             |
| 5                | April    | 24.0             |
| 5                | January  | 32.0             |
| 5                | July     | 36.0             |
| 5                | October  | 30.0             |
+------------------+----------+------------------+--+
```

8.Find the total number of meals served in each season



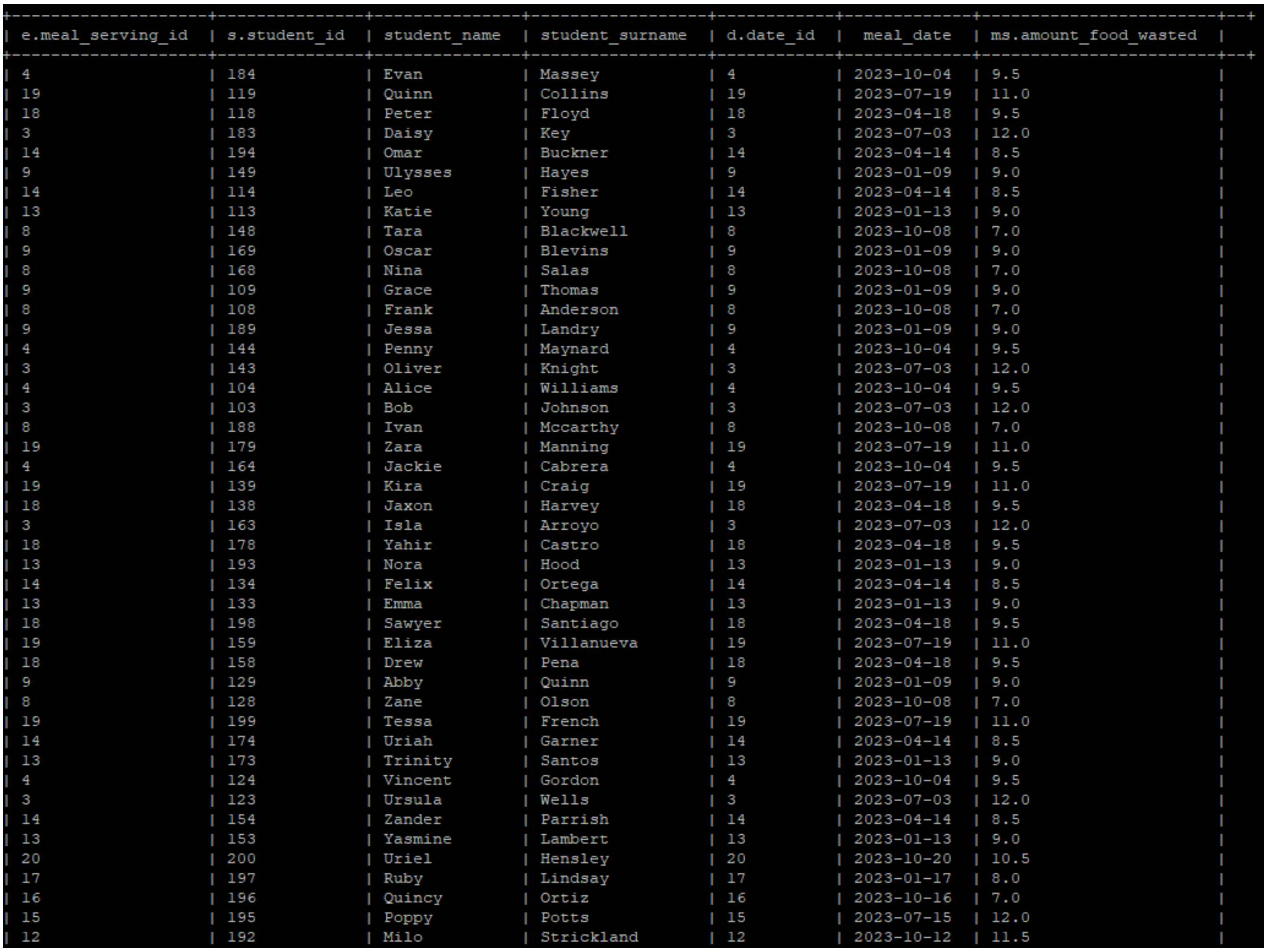
```
+---------+---------------+--+
| season  | meals_served  |
+---------+---------------+--+
| Fall    | 5             |
| Spring  | 5             |
| Summer  | 5             |
| Winter  | 5             |
+---------+---------------+--+
```

9.Find the total amount of money invested in meals for each age category of students.



```
+------------------+-------------------+--+
|  s.age_category  | total_investment  |
+------------------+-------------------+--+
| between 3 and 5  | 220.0             |
| from 5 to 6      | 275.0             |
| from 6 to 7      | 275.0             |
| less than 3      | 325.0             |
| more than 7      | 305.0             |
+------------------+-------------------+--+
```

10.Retrieve information about students who attended meals, including their names, the date of the meal, and the amount of food wasted for each meal.



```
+--------------------+--------------+---------------+-------------------+------------+-------------+------------------------+--+
| e.meal_serving_id  | s.student_id | student_name  | student_surname   | d.date_id  |  meal_date  | ms.amount_food_wasted  |
+--------------------+--------------+---------------+-------------------+------------+-------------+------------------------+--+
| 4                  | 184          | Evan          | Massey            | 4          | 2023-10-04  | 9.5                    |
| 19                 | 119          | Quinn         | Collins           | 19         | 2023-07-19  | 11.0                   |
| 18                 | 118          | Peter         | Floyd             | 18         | 2023-04-18  | 9.5                    |
| 3                  | 183          | Daisy         | Key               | 3          | 2023-07-03  | 12.0                   |
| 14                 | 194          | Omar          | Buckner           | 14         | 2023-04-14  | 8.5                    |
| 9                  | 149          | Ulysses       | Hayes             | 9          | 2023-01-09  | 9.0                    |
| 14                 | 114          | Leo           | Fisher            | 14         | 2023-04-14  | 8.5                    |
| 13                 | 113          | Katie         | Young             | 13         | 2023-01-13  | 9.0                    |
| 8                  | 148          | Tara          | Blackwell         | 8          | 2023-10-08  | 7.0                    |
| 9                  | 169          | Oscar         | Blevins           | 9          | 2023-01-09  | 9.0                    |
| 8                  | 168          | Nina          | Salas             | 8          | 2023-10-08  | 7.0                    |
| 9                  | 109          | Grace         | Thomas            | 9          | 2023-01-09  | 9.0                    |
| 8                  | 108          | Frank         | Anderson          | 8          | 2023-10-08  | 7.0                    |
| 9                  | 189          | Jessa         | Landry            | 9          | 2023-01-09  | 9.0                    |
| 4                  | 144          | Penny         | Maynard           | 4          | 2023-10-04  | 9.5                    |
| 3                  | 143          | Oliver        | Knight            | 3          | 2023-07-03  | 12.0                   |
| 4                  | 104          | Alice         | Williams          | 4          | 2023-10-04  | 9.5                    |
| 3                  | 103          | Bob           | Johnson           | 3          | 2023-07-03  | 12.0                   |
| 8                  | 188          | Ivan          | Mccarthy          | 8          | 2023-10-08  | 7.0                    |
| 19                 | 179          | Zara          | Manning           | 19         | 2023-07-19  | 11.0                   |
| 4                  | 164          | Jackie        | Cabrera           | 4          | 2023-10-04  | 9.5                    |
| 19                 | 139          | Kira          | Craig             | 19         | 2023-07-19  | 11.0                   |
| 18                 | 138          | Jaxon         | Harvey            | 18         | 2023-04-18  | 9.5                    |
| 3                  | 163          | Isla          | Arroyo            | 3          | 2023-07-03  | 12.0                   |
| 18                 | 178          | Yahir         | Castro            | 18         | 2023-04-18  | 9.5                    |
| 13                 | 193          | Nora          | Hood              | 13         | 2023-01-13  | 9.0                    |
| 14                 | 134          | Felix         | Ortega            | 14         | 2023-04-14  | 8.5                    |
| 13                 | 133          | Emma          | Chapman           | 13         | 2023-01-13  | 9.0                    |
| 18                 | 198          | Sawyer        | Santiago          | 18         | 2023-04-18  | 9.5                    |
| 19                 | 159          | Eliza         | Villanueva        | 19         | 2023-07-19  | 11.0                   |
| 18                 | 158          | Drew          | Pena              | 18         | 2023-04-18  | 9.5                    |
| 9                  | 129          | Abby          | Quinn             | 9          | 2023-01-09  | 9.0                    |
| 8                  | 128          | Zane          | Olson             | 8          | 2023-10-08  | 7.0                    |
| 19                 | 199          | Tessa         | French            | 19         | 2023-07-19  | 11.0                   |
| 14                 | 174          | Uriah         | Garner            | 14         | 2023-04-14  | 8.5                    |
| 13                 | 173          | Trinity       | Santos            | 13         | 2023-01-13  | 9.0                    |
| 4                  | 124          | Vincent       | Gordon            | 4          | 2023-10-04  | 9.5                    |
| 3                  | 123          | Ursula        | Wells             | 3          | 2023-07-03  | 12.0                   |
| 14                 | 154          | Zander        | Parrish           | 14         | 2023-04-14  | 8.5                    |
| 13                 | 153          | Yasmine       | Lambert           | 13         | 2023-01-13  | 9.0                    |
| 20                 | 200          | Uriel         | Hensley           | 20         | 2023-10-20  | 10.5                   |
| 17                 | 197          | Ruby          | Lindsay           | 17         | 2023-01-17  | 8.0                    |
| 16                 | 196          | Quincy        | Ortiz             | 16         | 2023-10-16  | 7.0                    |
| 15                 | 195          | Poppy         | Potts             | 15         | 2023-07-15  | 12.0                   |
| 12                 | 192          | Milo          | Strickland        | 12         | 2023-10-12  | 11.5                   |
```

(Only part shown)