Q.1

a-)

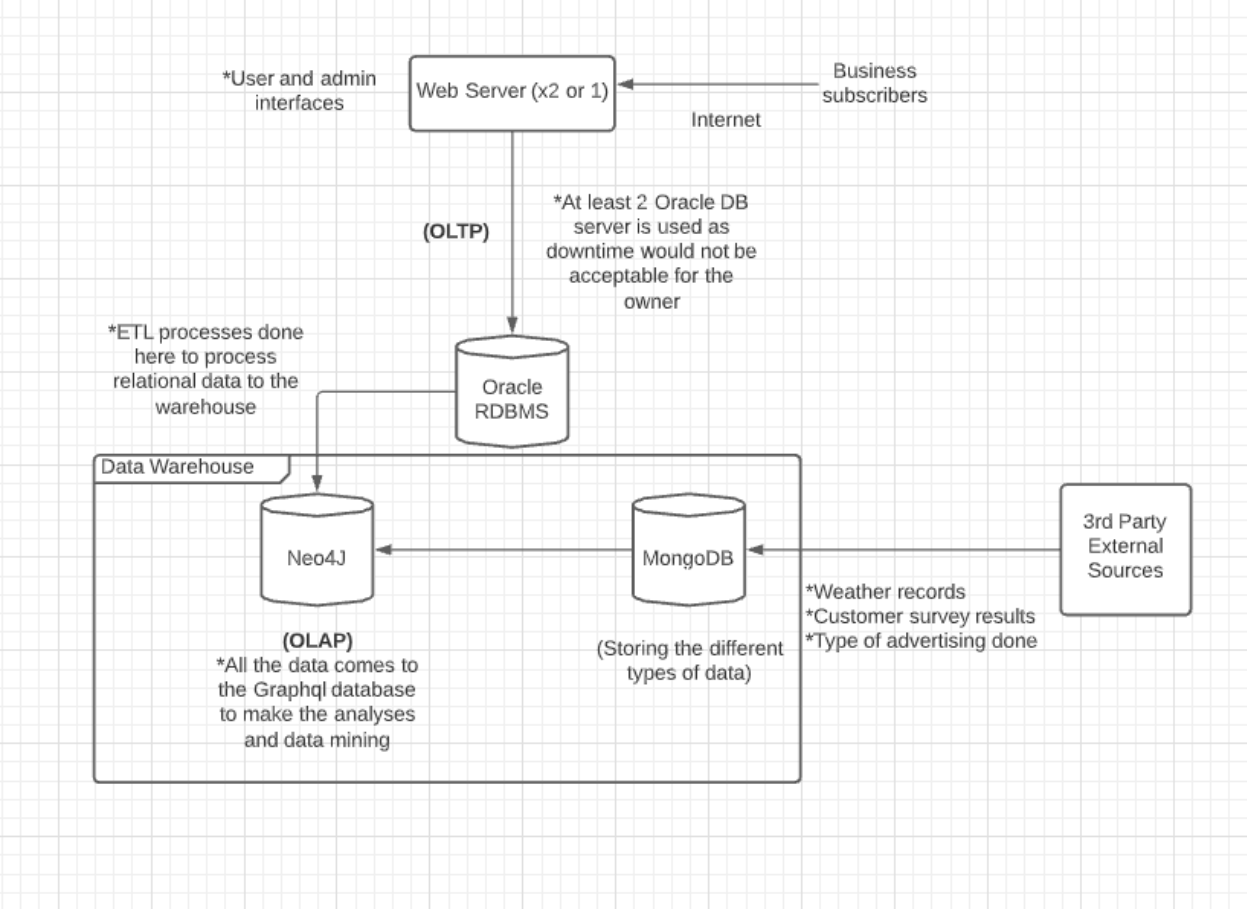
Firstly, the system needs a huge amount of data from different resources, and there will be unstructured data, so that MongoDB will be the first choice to gather and store the data because of its convenience to store humongous, different formats and kinds of data. All can be converted to BSON format easily.

Furthermore, the system's main purpose is to do OLAP, the predictions based on the patterns found in the data. Since the relationships will be important in this case than the actual data itself, a graph database, so Neo4j will be the choice for that purpose.

Lastly, the system also needs to make transactions and extracts on its partial data, so OLTP. For making those operations on a relatively smaller amount of data, a relational database can be used for its well-structured nature. Assuming the importance of reliability and performance, Oracle will be the last choice.

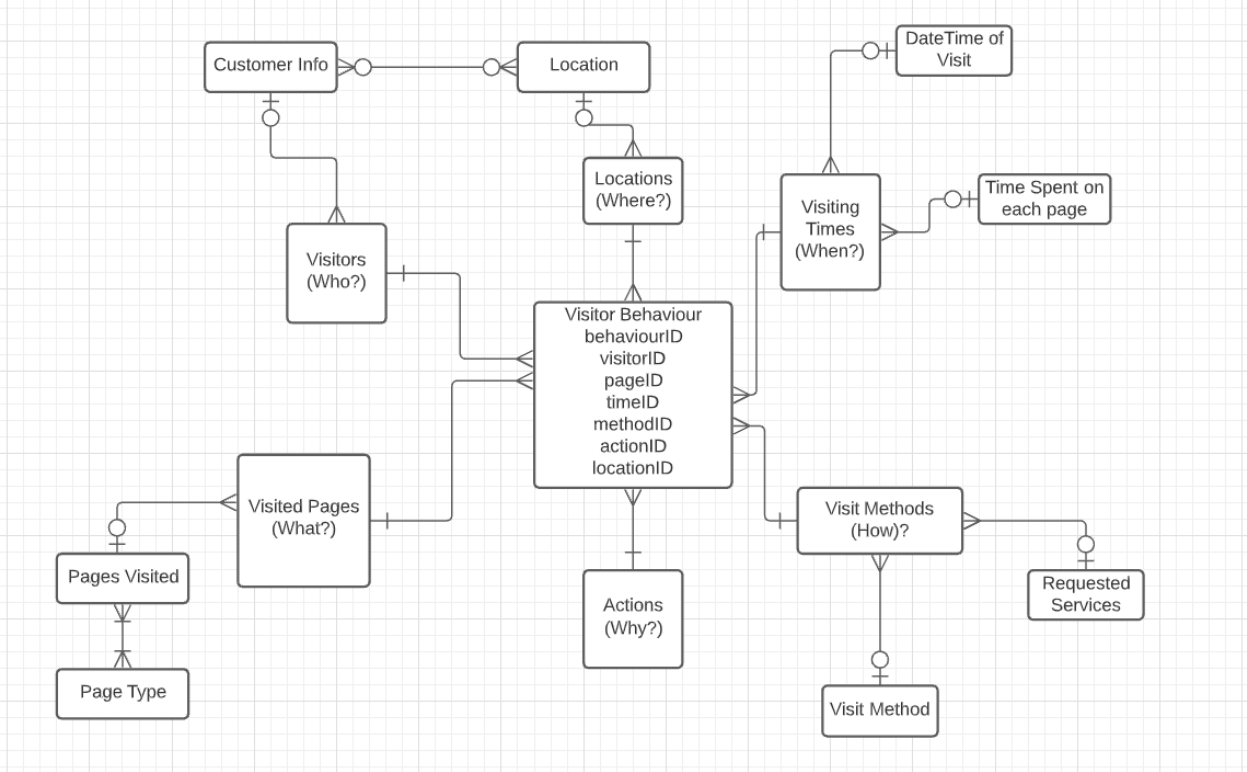
Note: As indicated, there is no hierarchy between those choices, and they all will be used somehow standalone for different purposes.

b-)



Q.2

a-)



Note: In the Visitor Behaviour table behaviourID is the primary key and all others are the foreign keys which are the primary keys of the referring dimension tables causing the One-to-Many relationships. Those dimension tables are thought to be in the fifth normal form which the fields are considered as separate entities. Only the mentioned attributes are included to the schema, but more attributes can be associated.

b-)

The attributes to be associated with the main fact table are:

Who - Customer profile: Job, age, gender

What – Pages Visited: Type of pages, categories

Where – Location: IP address, country, region

When – Time: Datetime of visit, time spent on each page

How – Visit method, requested services, device to visit