1.project explantion

explain project end to end with tools which you are using in it

2.what is the oracle hints

A [hint](https://docs.oracle.com/cd/E11882_01/server.112/e41573/glossary.htm#BGBHDFAC) is an instruction to the optimizer. When writing SQL, you may know information about the data unknown to the optimizer. Hints enable you to make decisions normally made by the optimizer, sometimes causing the optimizer to select a plan that it sees as higher cost.

**Types of Hints**

Hints can be of the following general types:

* Single-table

Single-table hints are specified on one table or view. [INDEX](https://docs.oracle.com/cd/E11882_01/server.112/e41084/sql_elements006.htm#SQLRF50405) and [USE\_NL](https://docs.oracle.com/cd/E11882_01/server.112/e41084/sql_elements006.htm#SQLRF50701) are examples of single-table hints.

* Multi-table

Multi-table hints are like single-table hints, except that the hint can specify one or more tables or views. [LEADING](https://docs.oracle.com/cd/E11882_01/server.112/e41084/sql_elements006.htm#SQLRF50705) is an example of a multi-table hint. Note that [USE\_NL(table1 table2)](https://docs.oracle.com/cd/E11882_01/server.112/e41084/sql_elements006.htm#SQLRF50701) is not considered a multi-table hint because it is a shortcut for USE\_NL(table1) and USE\_NL(table2).

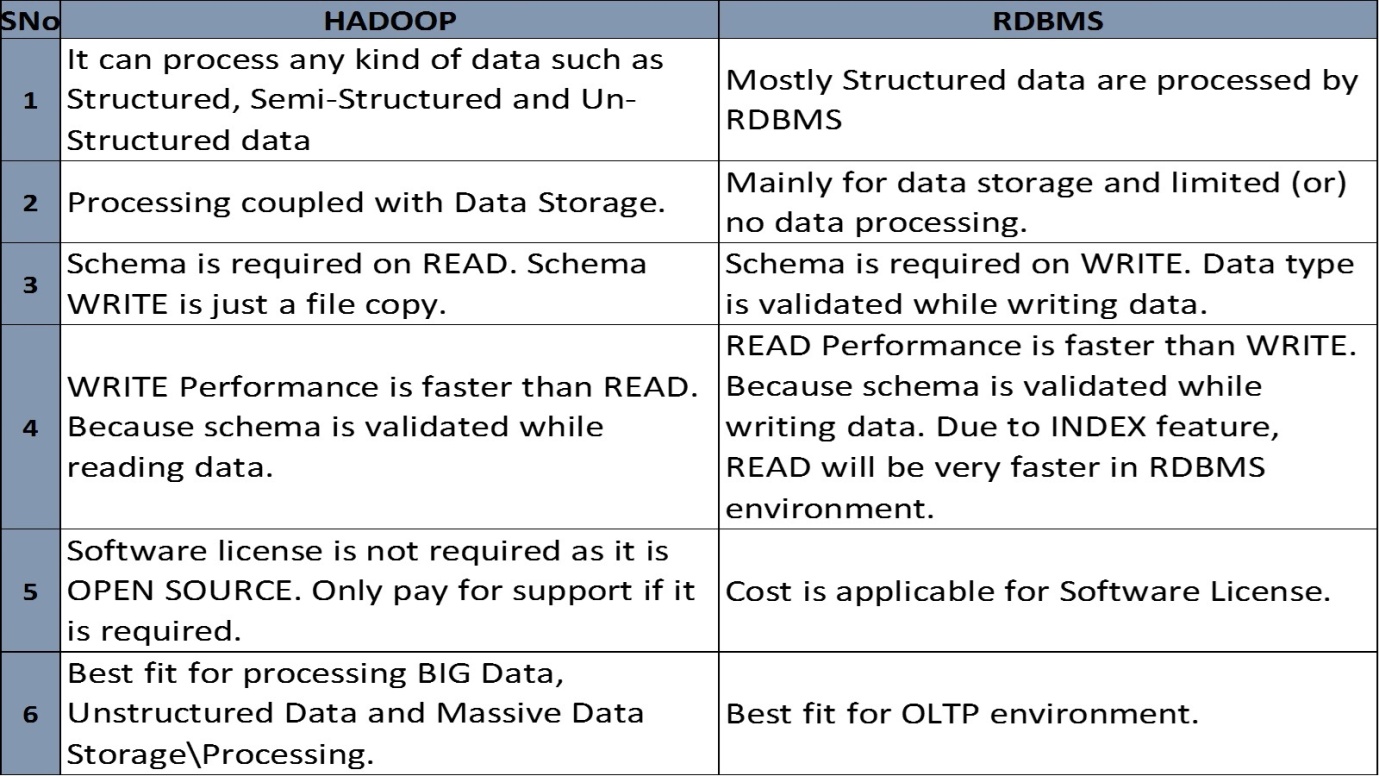
* Query block

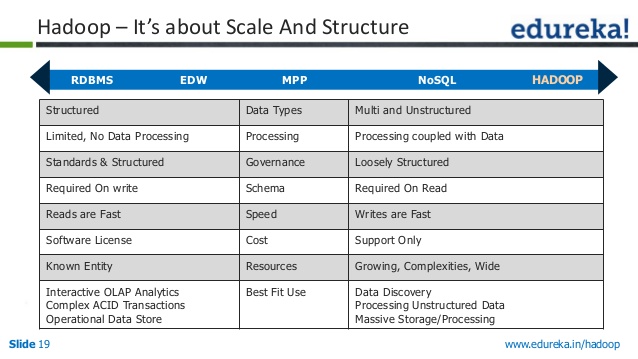
Query block hints operate on single query blocks. [STAR\_TRANSFORMATION](https://docs.oracle.com/cd/E11882_01/server.112/e41084/sql_elements006.htm#SQLRF50508) and [UNNEST](https://docs.oracle.com/cd/E11882_01/server.112/e41084/sql_elements006.htm#SQLRF50905) are examples of query block hints.

* Statement

Statement hints apply to the entire SQL statement. [ALL\_ROWS](https://docs.oracle.com/cd/E11882_01/server.112/e41084/sql_elements006.htm#SQLRF50301) is an example of a statement hint.

3.what is the deference between hadoop and databases





4, what is the deference between where clause and having clause

WHERE **clause** is used for filtering rows and it applies on each and every row, while **HAVING clause** is used to filter groups in SQL. 3) One syntax level **difference between** WHERE and **HAVING clause** is that, former is used before GROUP BY **clause**, while later is used after GROUP BY **clause**.

5.how can your rate in sql out of 5

6. what is the data ware house

A **data warehouse** is a relational database that is designed for query and analysis rather than for transaction processing. It usually contains historical **data** derived from transaction **data**, but it can include **data** from other sources.

The **Benefits of Data Warehousing** and ETL. **Data warehouses** are centralized **data** storage systems that allow your business to integrate **data** from multiple applications and sources into one location. This provides an environment that is designed for decision support, analytics reporting, and **data** mining.

7,what is order by and sort by and which is the best

Hive supports SORT BY which sorts the data per reducer. The difference between "order by" and "sort by" is that the former guarantees total order in the output while the latter only guarantees ordering of the rows within a reducer. If there are more than one reducer, "sort by" may give partially ordered final results.

Note: It may be confusing as to the difference between SORT BY alone of a single column and CLUSTER BY. The difference is that CLUSTER BY partitions by the field and SORT BY if there are multiple reducers partitions randomly in order to distribute data (and load) uniformly across the reducers.

Basically, the data in each reducer will be sorted according to the order that the user specified. The following example shows

SELECT key, value FROM src SORT BY key ASC, value DESC

8.what is serde and explantions

A SerDe is a combination of a Serializer and a Deserializer (hence, **Ser-De**). The Deserializer interface takes a string or binary representation of a record, and translates it into a Java object that Hive can manipulate

**Hive** uses Files systems like HDFS or any other storage (FTP) to store data, data here is in the form of **tables** (which has rows and columns). **SerDe** - Serializer, Deserializer instructs **hive** on how to process a record (Row)

9.types of serdes

Lazy simple serde(single delimiter)

regex serde(2 delimiters)

JsonSerDe(3 delimiters)

10. How can you handle multi delimiters

using regexSerde and custom SerDe

11 what you use for hive optimization

1. map-side join

2.proper using of partitions and bucketization

3.enabling vetorization

4.storing orc format

5.parallel execution

 set hive.exec.parallel=true;

 set hive.exec.parallel.thread.number=8;

12.hadoop support update

hadoop 1.x no support

hadoop 2.x support but only append( adding bottom of the file)