TASK - 3

DIABETES PATIENT ANALYSIS

USING SQL

BY RIYAVARSHINI ARUNPRASAD

POWERED BY

PSYLIQ

INTRODUCTION

- Diabetes patient analysis project is a third task provided by PSYLIQ during my internship.
- This project aims in demonstrating intern's SQL skills at all levels.
- The primary objective is to analyze a dataset consisting health records of diabetes patients.
- The entire project is executed using Oracle SQL Developer, providing a hands-on opportunity to showcase and refine my SQL expertise.

PSYLIQ

AGENDA

1 DATA OVERVIEW

2 DATA QUERIES

3 DATA INSIGHTS

SCHEMA ENHANCEMENT 4

QUERY OPTIMIZATION 5

CONCLUSION 6

PSYLIQ

DATA OVERVIEW

- **This dataset .xlxs format and has 1,00,000 rows and 11 columns making it ideal for analysis.**
- **Attributes and its datatype :**

Sl.No	Attribute Name	Data Type	Description
1	Patient_name	VARCHAR2(50)	Name of the patient
2	Patient_id	VARCHAR2(20)	Unique identifier of each patient
3	Gender	VARCHAR2(10)	Gender of Patient
4	Age	NUMBER(3)	Age of patient
5	Hypertension	NUMBER(1)	Indicates whether the patient has hypertension (1 for yes, 0 for no)
6	Heart_disease	NUMBER(1)	Indicates whether the patient has heart diseases (1 for yes, 0 for no)
7	Smoking_history	VARCHAR2(15)	Smoking history of patient
8	BMI	NUMBER(5,2)	Body Mass Index of patient
9	HbA1c_level	NUMBER(3,1)	Haemoglobin A1c level of patient
10	Blood_glucose_level	NUMBER(4)	Blood_glucose_level of patient
11	Diabetes	NUMBER(1)	Indicates whether the patient has Diabetes (1 for yes, 0 for no)

Note : While importing data into SQL developer an incorrect column name found and was renamed. Employeename → Patient_name

PSYLIQ

DATA QUERIES

1 Retrieve the patient_id and ages of all patients

```
SELECT patient_id, age
FROM health_record;
```

- Patient ID and Age of all patients were retrieved.

Worksheet | Query Builder

```
Describe health_record;

SELECT patient_id,age
FROM health_record;
```

Script Output | Query Result

SQL | Fetched 50 rows in 0.012 seconds

	PATIENT_ID	AGE
1	PT591	34
2	PT592	29
3	PT593	46
4	PT594	59
5	PT595	60
6	PT596	4
7	PT597	70
8	PT598	80
9	PT599	62
10	PT600	75

PSYLIQ

DATA QUERIES

2 Select all female patients who are older than 40.

```
SELECT patient_id,patient_name,gender,age
FROM health_record
WHERE gender = 'Female' AND age > 40;
```

- 29626 patients found to be female and are above the age of 40

Worksheet | Query Builder

```
SELECT patient_id,patient_name,gender,age
FROM health_record
WHERE gender = 'Female' and age > 40;
```

Script Output | Query Result

SQL | Fetched 50 rows in 0.165 seconds

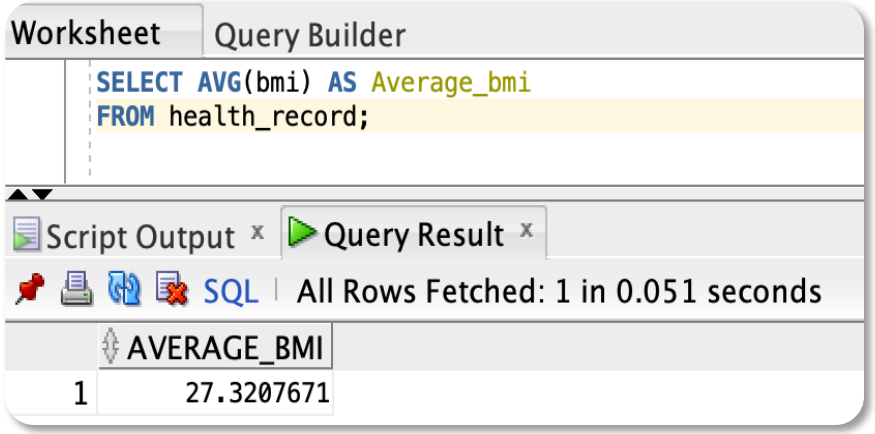
	PATIENT_ID	PATIENT_NAME	GENDER	AGE
1	PT593	JEFF COLUMBINI	Female	46
2	PT594	ANDRE WILLIAMS	Female	59
3	PT595	FLOYD ROLLINS	Female	60
4	PT598	KANDACE BENDER	Female	80
5	PT599	ROBERT SHAW	Female	62
6	PT605	ELLEN LEVIN	Female	78
7	PT606	TUAMELIE MOALA	Female	76
8	PT608	JULIA M C FRIEDLANDER	Female	69
9	PT619	NOREEN AMBROSE	Female	67
10	PT621	KIMIKO BURTON	Female	58

PSYLIQ

DATA QUERIES

3 Calculate the average BMI of patients

```
SELECT AVG(bmi) AS Average_bmi
FROM health_record;
```



- The average Body Mass Index of patients found to be approximately 27.3.

DATA QUERIES

4 List patients in descending order of blood glucose levels

```
SELECT
patient_id,patient_name,blood_glucose_level
FROM health_record
ORDER BY blood_glucose_level DESC;
```

- The patient details have been sorted in descending order of blood glucose level.

Worksheet | Query Builder

```
SELECT patient_id,patient_name,blood_glucose_level
FROM health_record
ORDER BY blood_glucose_level DESC;
```

Script Output | Query Result

SQL | Fetched 50 rows in 0.111 seconds

	PATIENT_ID	PATIENT_NAME	BLOOD_GLUCOSE_LEVEL
1	PT89191	Jacqueline M Phillips	300
2	PT91095	Bonnie S Ma	300
3	PT92871	Mary Ann Moran	300
4	PT91135	Zandra L Thompson	300
5	PT91144	Editha J Pascual	300
6	PT98454	Lenora G Banks	300
7	PT98461	Dante Rogayan	300
8	PT98500	Tinisha C Bishop	300
9	PT98538	Tualatai Auimatagi	300
10	PT95386	Cristina T Caceres	300

PSYLIQ

DATA QUERIES

5 Find patients who have hypertension and diabetes

```
SELECT patient_id, patient_name, hypertension,
diabetes
FROM health_record
WHERE hypertension = 1 and diabetes = 1;
```

- 1730 patients in dataset found to have hypertension and diabetes.

Worksheet | Query Builder

```
SELECT patient_id,patient_name,hypertension, diabetes
FROM health_record
WHERE hypertension = 1 and diabetes = 1;
```

Script Output | Query Result

SQL | Fetched 50 rows in 0.07 seconds

	PATIENT_ID	PATIENT_NAME	HYPERTENSION	DIABETES
1	PT632	ROBERT BONNET	1	1
2	PT139	JONES WONG	1	1
3	PT205	PATRIC STEELE	1	1
4	PT343	ARTHUR STELLINI	1	1
5	PT355	CHAD LAW	1	1
6	PT451	CATHERINE JAMES	1	1
7	PT565	JOHN HART	1	1
8	PT567	JOHN BARKER	1	1
9	PT1557	JOSEPH MCFADDEN	1	1
10	PT1577	DAVID HAMILTON	1	1

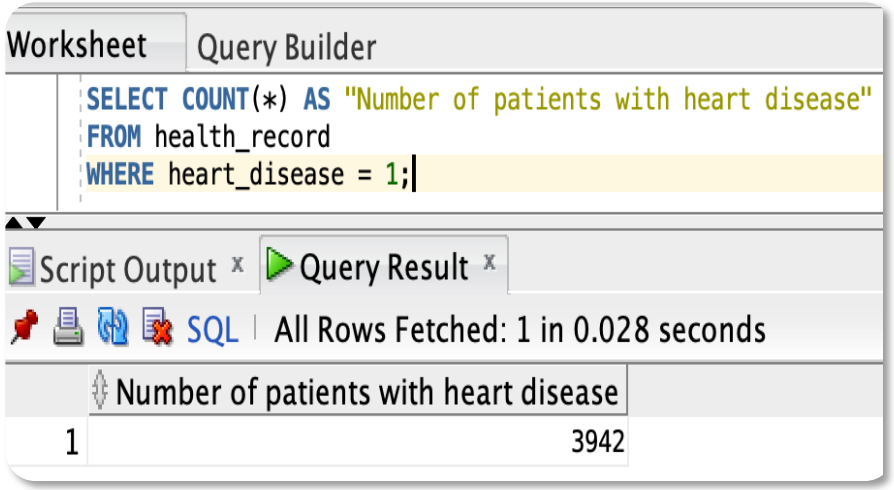
PSYLIQ

DATA QUERIES

6 Determine the number of patients with heart disease

```
SELECT COUNT(*) AS "Number of patients with
heart disease"
FROM health_record
WHERE heart_disease = 1;
```

- 3942 patients in dataset found to have heart disease



PSYLIQ

DATA QUERIES

7 Group patients by smoking history and count how many patients are there ?

```
SELECT SMOKING_HISTORY, COUNT(*) AS "TOTAL
PATIENT"
FROM health_record
GROUP BY smoking_history;
```

- There were five distinct entries for denoting patients smoking history and about 35816 entries of No info.
- Out of five, Most number of patients about 35095 people never smoked in their life.

Worksheet | Query Builder

```
SELECT SMOKING_HISTORY, COUNT(*) AS "TOTAL PATIENTS"
FROM health_record
GROUP BY smoking_history;
```

Script Output | Query Result

SQL | All Rows Fetched: 6 in 0.056 seconds

	SMOKING_HISTORY	TOTAL PATIENTS
1	No Info	35816
2	ever	4004
3	former	9352
4	current	9286
5	not current	6447
6	never	35095

PSYLIQ

DATA QUERIES

8 Retrieve the Patient_ids of patients who have a BMI greater than the average BMI

```
SELECT patient_id
FROM health_record
WHERE bmi > (
                SELECT AVG(BMI)
                FROM health_record );
```

- The Body Mass Index (BMI) of the respective patient greater than the average BMI is calculated from the subquery.
- 56802 records are retrieved.

Worksheet | Query Builder

```
SELECT patient_id
FROM health_record
WHERE bmi > (SELECT AVG(BMI) FROM health_record);
```

Script Output | Query R...

SQL | Fetched 50 rows in 0.051 seconds

	PATIENT_ID
1	PT592
2	PT594
3	PT595
4	PT598
5	PT600
6	PT603
7	PT608
8	PT613
9	PT615
10	PT616

PSYLIQ

DATA QUERIES

9 Find the patient with the highest HbA1c level and the patient with the lowest HbA1clevel.

- The Common Table Expression selects patient details for both the maximum and minimum HbA1c levels using the UNION operator.
- The CTE separates the logic for finding the highest and lowest HbA1c levels, providing a cleaner and more organized structure.

Worksheet | Query Builder

```sql
WITH highest_lowest_hba1c_level as(
SELECT patient_id,patient_name,hba1c_level
FROM health_record
WHERE hba1c_level = (SELECT MAX(hba1c_level) FROM health_record)
UNION
SELECT patient_id,patient_name,hba1c_level
FROM health_record
WHERE hba1c_level = (SELECT MIN(hba1c_level) FROM health_record))

SELECT patient_id,patient_name,hba1c_level
FROM highest_lowest_hba1c_level
ORDER BY hba1c_level DESC;
```

Script Output | Query Result

SQL | Fetched 1,000 rows in 0.343 seconds

	PATIENT_ID	PATIENT_NAME	HBA1C_LEVEL
652	PT15868	WILSON LAU	9
653	PT16159	JOSEPH LIEBMAN	9
654	PT16324	WILLIAM WOOD	9
655	PT100000	Meredith H Reddoch-Ho	3.5
656	PT100008	Tara L Croan	3.5
657	PT10001	ROBERTO VALLADARES	3.5

PSYLIQ

DATA QUERIES

10 Calculate the date of birth of patients in years (assuming the current date as of now).

```
SELECT
    patient_id, patient_name, age,
    EXTRACT(YEAR FROM SYSDATE) - age AS "Year
of Birth"
FROM health_record;
```

- Subtracting the age of patient from current date provides year of birth of patients.
- SYSDATE provides current date in Oracle SQL developer.

Worksheet | Query Builder

```
SELECT
    patient_id,
    patient_name,
    age,
    EXTRACT(YEAR FROM SYSDATE) - age AS "Year of Birth"
FROM health_record;
```

Script Output | Query Result

SQL | Fetched 50 rows in 0.016 seconds

	PATIENT_ID	PATIENT_NAME	AGE	Year of Birth
1	PT591	CARL FABBRI	34	1990
2	PT592	JILL LECOUNT	29	1995
3	PT593	JEFF COLUMBINI	46	1978
4	PT594	ANDRE WILLIAMS	59	1965
5	PT595	FLOYD ROLLINS	60	1964
6	PT596	JOSE CASTILLO	4	2020
7	PT597	THOMAS KOHMANN	70	1954
8	PT598	KANDACE BENDER	80	1944
9	PT599	ROBERT SHAW	62	1962
10	PT600	LEONARDO FERMIN	75	1949

PSYLIQ

DATA QUERIES

11 Rank patients by blood glucose level within each gender group.

```
SELECT
patient_id, patient_name, gender, blood_glucose_level,
DENSE_RANK() OVER(partition by gender order by
blood_glucose_level desc) as "RANK by blood glucose level"
FROM health_record;
```

- DENSE_RANK() window function to assign a rank to each record in the health_record table based on the descending order of the blood_glucose_level within each gender partition.

Worksheet | Query Builder

```
SELECT
    patient_id,
    patient_name,
    gender,
    blood_glucose_level,
    DENSE_RANK() OVER(partition by gender order by blood_glucose_level desc) as "RANK by blood glucose leve
FROM health_record;
```

Script Output | Query Result

SQL | Fetched 50 rows in 0.082 seconds

	PATIENT_ID	PATIENT_NAME	GENDER	BLOOD_GLUCOSE_LEVEL	RANK by blood glucose level
1	PT6227	WINSON SETO	Female	300	1
2	PT7029	PAUL FOSTER	Female	300	1
3	PT7040	SPENCER LEE	Female	300	1
4	PT5536	STEPHEN ZOLLMAN	Female	300	1
5	PT5573	CARLOS CARRILLO	Female	300	1
6	PT5302	JOHN SCOTT	Female	300	1
7	PT4652	MARK RUDNICKI	Female	300	1
8	PT5233	MARY DONOVAN	Female	300	1
9	PT3675	DANIEL LUI	Female	300	1
10	PT4197	JOSE MEJIA	Female	300	1

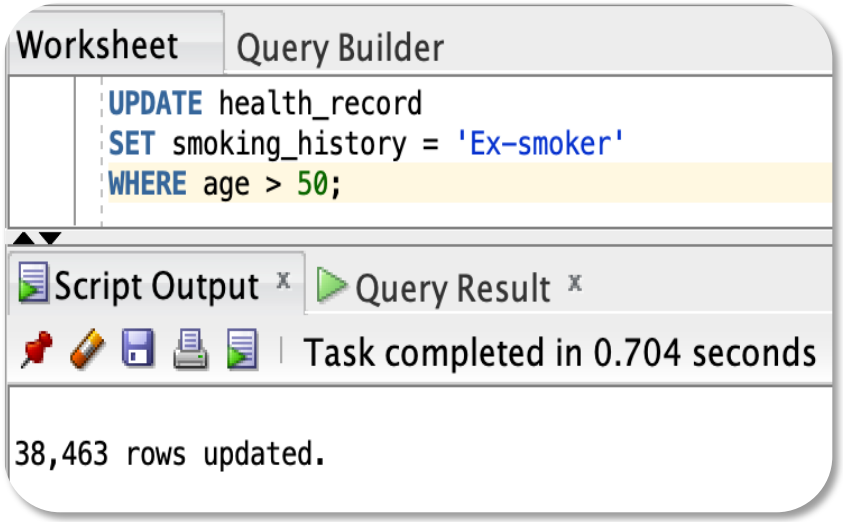
PSYLIQ

DATA QUERIES

12

Update the smoking history of patients who are older than 50 to "Ex-smoker."

```
UPDATE health_record
SET smoking_history = 'Ex-smoker'
WHERE age > 50;
```



- Data Manipulation Language (DML) statements are responsible for manipulating data stored in the database.
- UPDATE performs updation in table health_record where criteria meet.

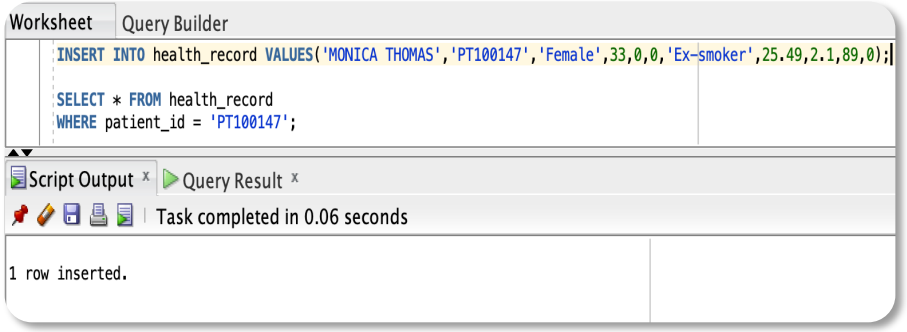
PSYLIQ

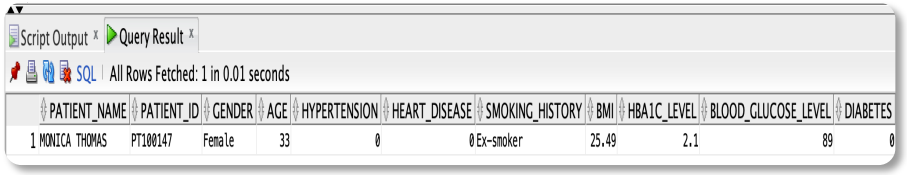
DATA QUERIES

13

Insert a new patient into the database with sample data.

```
INSERT INTO health_record
VALUES('MONICA THOMAS', 'PT100147', 'Female',
33,0,0,'Ex-smoker',25.49,2.1,89,0);
```





	PATIENT_NAME	PATIENT_ID	GENDER	AGE	HYPERTENSION	HEART_DISEASE	SMOKING_HISTORY	BMI	HBA1C_LEVEL	BLOOD_GLUCOSE_LEVEL	DIABETES
1	MONICA THOMAS	PT100147	Female	33	0	0	Ex-smoker	25.49	2.1	89	0

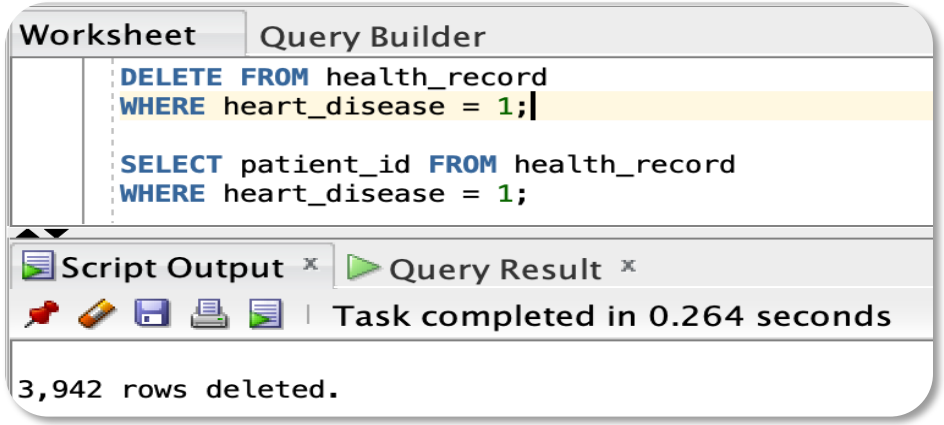
- Data Manipulation Language (DML) statement INSERT is responsible for adding new row of values into table health_record.

DATA QUERIES

14 Delete all patients with heart disease from the database.

```
DELETE FROM health_record
WHERE heart_disease = 1;
```

- Data Manipulation Language (DML) statement DELETE removes the rows where criteria meets.
- 3942 records of patients with heart_diseases were deleted from health_record.



PSYLIQ

DATA QUERIES

15

Find patients who have hypertension but not diabetes using the EXCEPT operator

- Equivalent of EXCEPT in Oracle SQL Developer is MINUS
- The MINUS operator is used to subtract the result set of the second query from the result set of the first query.

Worksheet | Query Builder

```
SELECT patient_id,patient_name,hypertension,diabetes FROM health_record
WHERE hypertension = 1
MINUS
SELECT patient_id,patient_name,hypertension,diabetes FROM health_record
WHERE diabetes = 1;
```

Script Output | Query Result

SQL | Fetched 50 rows in 0.045 seconds

	PATIENT_ID	PATIENT_NAME	HYPERTENSION	DIABETES
1	PT100009	Brian M DeNave	1	0
2	PT100010	Jennifer J Pascual	1	0
3	PT100015	Joshua R Mcdonald	1	0
4	PT100049	Mimi Su	1	0
5	PT100054	Marisa E Lott	1	0
6	PT100064	Ronald Lee	1	0

PSYLIQ

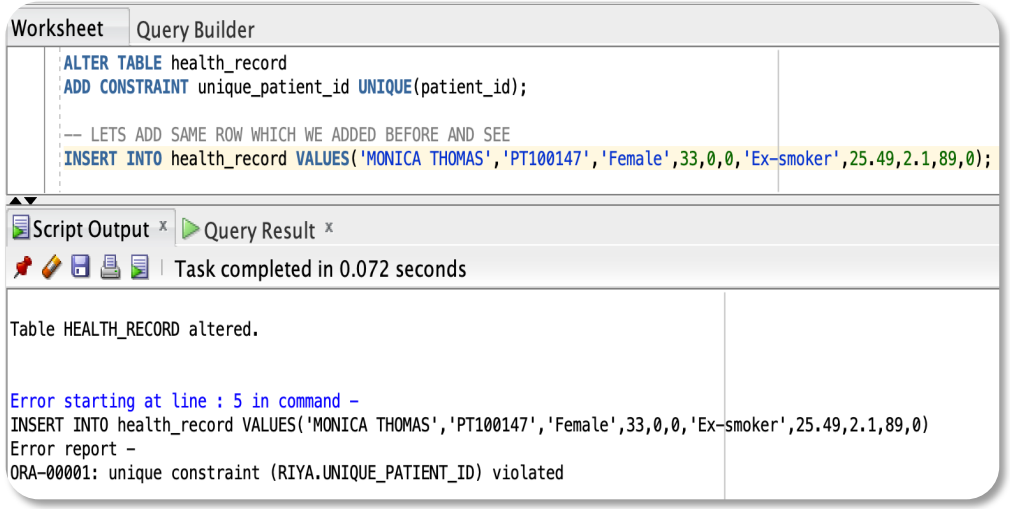
DATA QUERIES

16

Define a unique constraint on the "patient_id" column to ensure its values are unique.

```
ALTER TABLE health_record
ADD CONSTRAINT unique_patient_id
UNIQUE(patient_id);
```

- Data Definition Language statement ALTER TABLE ... ADD CONSTRAINT is used to add or modify the constraints in columns of table.



PSYLIQ

DATA QUERIES

17

Create a view that displays the Patient_ids, ages, and BMI of patients.

```
CREATE VIEW patient_bmi AS
(    SELECT patient_id, age, bmi
     FROM health_record
     );
SELECT patient_id, age, bmi
FROM patient_bmi;
```

Worksheet | Query Builder

```
CREATE VIEW patient_bmi AS
(    SELECT patient_id, age, bmi
     FROM health_record
     );

SELECT patient_id, age, bmi FROM patient_bmi;
```

Script Output | Query Result

Task completed in 0.098 seconds

```
View PATIENT_BMI created.
```

Script Output | Query Result

SQL | Fetched 50 rows in 0.016 seconds

	PATIENT_ID	AGE	BMI
1	PT591	34	23.91
2	PT592	29	36.17
3	PT593	46	21.66
4	PT594	59	29.93
5	PT595	60	28.82
6	PT596	4	15.64

- The view patient_bmi is created to provide a convenient way to query and analyze data from the health_record table, specifically focusing on the patient_id, age, and bmi columns.

PSYLIQ

SCHEMA ENHANCEMENT

NORMALIZATION

Normalizing schema helps to reduce redundancy and enhance data integrity

Normalization in given dataset: separating columns into three tables as Patient_details,Health_parameters, Health_history

ADD CONSTRAINTS

Creating constraints on columns ensure data validation, reduce redundancy, reduce missing values, and in some case, help reduce extreme outliers

Make patient_id as primary key in one table and as foreign key in another table. Add check constraint in health_parameter table to avoid extreme outliers.

PSYLIQ

QUERY OPTIMIZATION

Query optimization is a crucial aspect of database performance tuning.

Some of the optimisation techniques:

- LIMIT THE RESULTS
- REDUCE USING DISTINCT KEYWORD OFTEN
- AVOID USING SELECT * INSTEAD RETRIEVE SELECTIVE COLUMNS
- IMPLEMENT INDEXING ON MOST QUERYING COLUMNS

PSYLIQ

THANK YOU

PSYLIQ