Ciencia de Datos

Introducción a SQL

R2

Tech for all

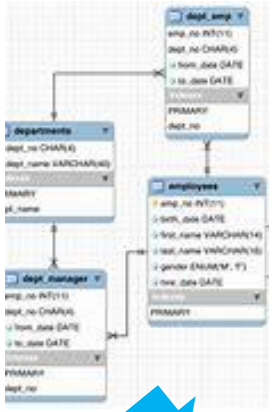
Bases de Datos- SQL

Importancia del SQL

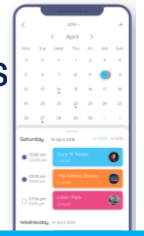
Proceso de Transformación

Requisitos
-Cliente

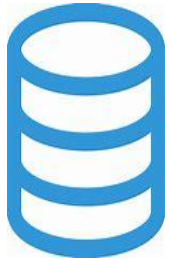
Diseño BD



APPs



BD

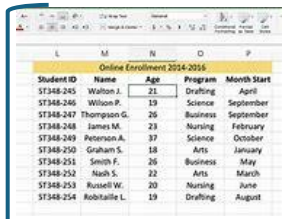


Data

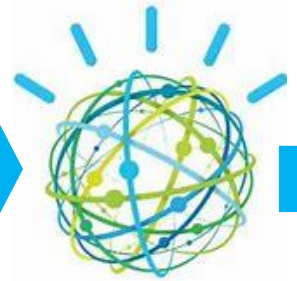
Consultas
DML
DDL

SQL

Dataset



Análisis
-Algoritmos
-Transformación
-IA (ML,DL...)



Data Mining
Data Science
Big Data
-

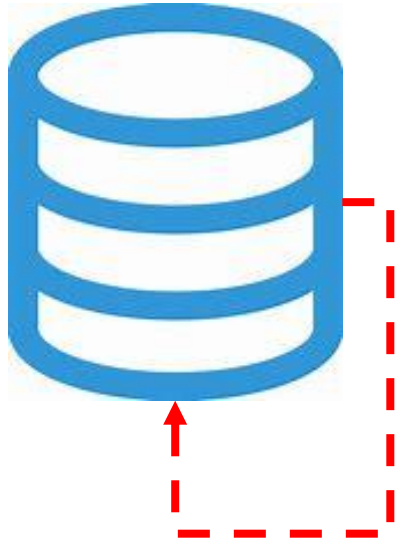
Resultados

Dinámico

Bases de Datos- SQL

Efecto SQL en la BD

SQL



SQLQuery3.sql - MA...te (MAC\user (52))*

```
Insert Into Funcionarios(Nome,Sobrenome,Salario,Sexo,Cidade) values
Select * from funcionarios
```

100 %

Results | Messages

	Id	Nome	Sobrenome	Salario	Sexo	Cidade
1	1	Pedro	Ribeiro	5000	M	Brasilia
2	2	João Paulo	Santos	1800	M	Curitiba
3	3	Maria Rita	Lousada	3800	F	Santos
4	4	Domingos	Ramirez	2500	M	Rio de Janeiro
5	5	Carlos	Bueno	2000	M	São Paulo
6	6	Vera Lucia	Silva	5000	F	Londrina
7	7	Camila	Sanches	4000	F	São Paulo
8	8	Jefferson Andre	Silva	3500	M	Brasilia
9	9	Marta Ribeiro	Soares	1500	F	Campinas
10	10	Amaldo	Giacomelli	4200	M	Santos
11	11	Suzana	Vieira	3200	F	São Paulo
12	12	Miriam Rocha	Silveira	2100	F	Ribeirão Preto
13	13	Helena	Ribeiro	1500	F	Belo Horizonte
14	21	Xavier	de Toledo	2500	M	Jundiai

Bases de Datos- SQL

SQL – Structured Query Language

"Lenguaje Estructurado de Consulta "

- Uso en BD relacionales
- Define estructuras
- Almacenar
- Procesar (CRUD)
- Gestionar

- Simple, en estructura y complejidad
- Estructura similar a ingles corriente

ISO/IEC 9075

formaliza las estructuras y comportamientos de la sintaxis SQL

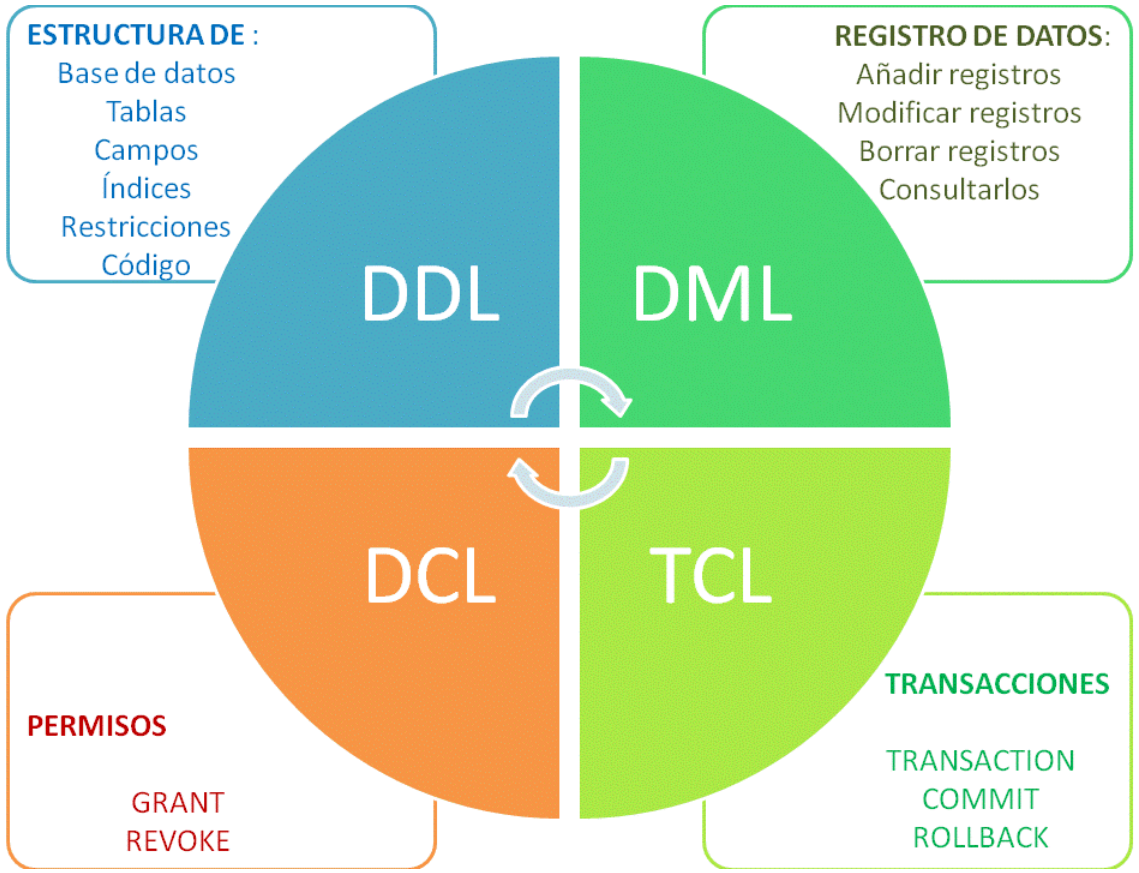
Importancia de SQL“

- Almacenamiento persistente
- Inter operatividad
- Análisis de datos
- Automatización
- Seguridad

Bases de Datos- SQL

Sub-lenguajes

- DDL -Data Definition Language
- DML- Data Manipulation Language
- DCL -Data Control Language
- TCL - Transaction Control Language



Bases de Datos- SQL

DDL -Data Definition Language

```
CREATE TABLE empleados (
    id INT PRIMARY KEY,
    nombre VARCHAR(100),
    puesto VARCHAR(50),
    salario DECIMAL(10, 2)
);
```

```
ALTER TABLE empleados
ADD fecha_contratacion DATE;
```

```
DROP TABLE empleados;
```

```
/*==============================================================
/* Table: Formadores
/*==============================================================
create table Formadores (
   Cod_formador          integer                identity,
   Nombre_empresa        char(60)               null,
   Nit                   char(20)               null,
   Telefono              char(25)               null,
   Web                   char(60)               null,
   email                 char(60)               null,
   contacto              char(60)               null,
   Descripcion           char(200)              null,
   constraint PK_FORMADORES primary key (Cod_formador)
)
```

```
CREATE INDEX idx_nombre ON empleados (nombre);
```

Bases de Datos- SQL

DML -Data Manipulation Language

Operaciones CRUD con una BD

Operación	Instrucción SQL	Descripción
CREATE	`INSERT INTO tabla (columnas) VALUES (valores);`	Insertar un nuevo registro en la tabla.
READ	`SELECT columnas FROM tabla WHERE condición;`	Leer datos de la tabla.
UPDATE	`UPDATE tabla SET columna = valor WHERE condición;`	Actualizar registros existentes en la tabla.
DELETE	`DELETE FROM tabla WHERE condición;`	Eliminar registros de la tabla.

Bases de Datos- SQL

DML -Data Manipulation Language

Tabla empleados

```sql
CREATE TABLE empleados (
    id_empleado INT PRIMARY KEY,
    nombre VARCHAR(100),
    departamento VARCHAR(100),
    salario DECIMAL(10, 2)
);
```

Insertar 3 registros en la tabla

```sql
INSERT INTO empleados (id_empleado, nombre, departamento, salario)
VALUES (1, 'Juan Pérez', 'Ventas', 3000.00),
       (2, 'Ana Gómez', 'Marketing', 3500.00),
       (3, 'Luis Martínez', 'IT', 4000.00);
```

Bases de Datos- SQL

DML -Data Manipulation Language

```sql
-- Leer todos los registros
SELECT * FROM empleados;


-- Leer un registro específico
SELECT * FROM empleados WHERE id_empleado = 1;
```

Clausula **WHERE**

Se utiliza para filtrar el alcance de la consulta.

Bases de Datos- SQL

DML -Data Manipulation Language

```sql
-- Actualizar el salario de un empleado
UPDATE empleados
SET salario = 3200.00
WHERE id_empleado = 1;

-- Actualizar el departamento de un empleado
UPDATE empleados
SET departamento = 'Recursos Humanos'
WHERE id_empleado = 2;
```

Clausula **WHERE**
Se utiliza para Limitar o Filtrar-los registros que se actualizan.

Bases de Datos- SQL

DML -Data Manipulation Language

```sql
-- Eliminar un empleado por su ID
DELETE FROM empleados
WHERE id_empleado = 3;
```

Clausula **WHERE**
Se utiliza para Limitar o Filtrar-los registros que se eliminan.

Advertencia:
Si una instrucción **DELETE** no incluye la sentencia **WHERE** podria elimina todos lso registros de la tabla

WHERE, ORDER BY, GROUP BY

WHERE

Se utiliza para Limitar los registros involucrados.

```sql
UPDATE empleados
SET departamento = 'Recursos Humanos'
WHERE id_empleado = 2;
```

ORDER BY

ordenar los resultados de una consulta en orden ascendente o descendente según una o más columnas.

```sql
SELECT nombre, departamento, salario
FROM empleados
ORDER BY departamento ASC, salario DESC;
```

GROUP BY

Agrupar filas que tienen valores iguales en columnas específicas, frecuente con funciones de agregación.

```sql
SELECT departamento, COUNT(*) AS numero_empleados
FROM empleados
GROUP BY departamento;
```

AS, define un alias de encabezado en resultado.

Funciones de Agregación

- Son funciones que toman un conjunto de valores y devuelven un único valor,
- Algunas veces aplica cláusula GROUP BY

GROUP BY
Para agrupar resultados de una consulta.

Función	Descripción	Ejemplo de Uso Común en Ciencia de Datos
`COUNT`	Cuenta el número de filas en un conjunto de resultados.	`SELECT COUNT(*) FROM empleados;`
`SUM`	Calcula la suma total de un conjunto de valores.	`SELECT SUM(salario) FROM empleados;`
`AVG`	Calcula el promedio de un conjunto de valores.	`SELECT AVG(salario) FROM empleados;`
`MIN`	Devuelve el valor mínimo de un conjunto de valores.	`SELECT MIN(salario) FROM empleados;`
`MAX`	Devuelve el valor máximo de un conjunto de valores.	`SELECT MAX(salario) FROM empleados;`

Funciones de Agregación

Otras funciones

`STDDEV`	Calcula la desviación estándar de un conjunto de valores.	`SELECT STDDEV(salario) FROM empleados;`
`VARIANCE`	Calcula la varianza de un conjunto de valores.	`SELECT VARIANCE(salario) FROM empleados;`
`MEDIAN`	Calcula la mediana de un conjunto de valores.	`SELECT MEDIAN(salario) FROM empleados;`
`PERCENTILE_CONT`	Calcula el percentil continuo de un conjunto de valores.	`SELECT PERCENTILE_CONT(0.5) WITHIN GROUP (ORDER BY salario) FROM empleados;`
`PERCENTILE_DISC`	Calcula el percentil discreto de un conjunto de valores.	`SELECT PERCENTILE_DISC(0.5) WITHIN GROUP (ORDER BY salario) FROM empleados;`

Integración SQL Server Pyhton ML

Servicios de Machine Learning

- rxLinMod
- rxLogit
- rxGlm
- rxBTrees
- rxDTree
- rxNaiveBayes
- rxKmeans
- rxNeuralNet
- rxPredict

```
EXEC sp_execute_external_script
  @language = N'Python',
  @script = N'
    import pandas as pd
    from sklearn.linear_model import LinearRegression

    # Load data
    data = InputDataSet

    # Prepare the model
    model = LinearRegression()
    model.fit(data[["Age", "Education"]], data["Salary"])

    # Output the coefficients
    OutputDataSet = pd.DataFrame(model.coef_, columns=["Coefficient"])
  ',
  @input_data_1 = N'SELECT Age, Education, Salary FROM Employees',
  @output_data_1_name = N'OutputDataSet';
```

Ciencia de Datos

Taller

Gracias