Primera clase 09/01/23

Curso Big Data

Es transversal y **escalable (dos palabras claves y fundamentales en BD**)

Ver Skyscanner

Disrupciòn: innovación que destruye

ROLES Y TRABAJOS:

Big Data architect/ Developer

Data Engineer

Data Analyst

Data scientist



Big Data architect/ Developer

Data Engineer

Data Analyst

Data scientist

Nombres empresas

Salarios

Tecnologias

Que hacen

Buscar empleos de las empresas con estos perfiles

Big Data architect/ Develope

**Integrated Worlds GmbH**

What Doing?

* .

Salario

* Salary

**Leer sobre el algoritmo de amazon, método de marketing.**

Las 4V del Big Data

* Volumen
* Velocidad
* Variedad
* **Valor ( Es lo que todos buscan, disminuir coste, aumentar productividad, …)**

Segunda Clase 10/01/23

Business Intelligence BI

**Buscar librería Leaflit (es para visualizar los mapas, es código libre)**

Tipos de datos;

**Estructurado** (Esquema: Tabls y campos). Rigidos

**Semi estructurado**: Hay algo de estructura. Eje: CSV, JSON

CSV: Es el màs rigido de este tipo de datos, es el más interesante, pero hay que tener control con el tipo de datos.

JSON: Es mas flexible, ante la incertidumbre d elos datos es mucho mejor. Es un fichero generalmente entre corchetes, del tipo:

[

{ “Nombre”: “Jhon”

“Edad”: “28”

“Notas”: “9” }

,

{ “Nombre”: “Santiago”

“Edad”: “28”

“Notas”: “10” }

]

**ESTUDIAR CSV Y JSON**

**No estructurado**

Son los más complejos, se deben modelar

**CUADROS DE MANDO, PANEL O DASHBOARD**

**KPI** es información agrupada con un dibujo. Deben ser específicos y relevantes.

Un dashboard se compone de uno a varios KPI (gráficos u indicadores)

Se utilizan por diversas ventajas:

* Permiten obtener **información valiosa y útil**
* Medir determinadas variables y resultados a partir de la información
* **Analizar la información** y efectos de unas determinadas estrategias.

CRM Customer Relationship management

ERP …

Herramientas para BI TOP 3

**Power BI; Tableau**; Qlik

Descargar sublime y entrar en Kaggle desde internet

Comandos rapidos sublime

Window + L cierra sesión

Ctrl+shift+n carpeta nueva

Window + R ejecutador (cmd, Paint, notepad)

Window + E: explorador

Ctrl +W: cerrar pestaña, ventana

Ctrl+ shift+t: Devuelve la acción en el buscador der internet

Como hacer el dashboard desde Qlik

* 1. Diseñar una App analítica
  2. Subir datos
  3. Diseñar Panel

**Clase 3.**

**Presentaciones de programación, Big Data, inteligencia artificial**

Todo en programación debe ser:

**Secuencial y determinista**

Para programar primero se debe pensar en el problema y luego resolverlo

Propiedades ACID de SQL (Estudiar)

SQL (Search querines language)

HADOOP (framework). Un Framework

**HDFS Hadoop distributed file system (almacena info)**

**MapReduce (Programaciòn)**

**Batch processing** ( grandes volúmenes de datos que se trabajan para tener pronta disponibilidad)

**Estudiar Deep learning y machine learning.**

**Machine learning 🡪 son supervisados y no supervisados**

**Si quiero predecir algo, todo lo qie tenga que ver con categorías (se hacen con algoritmos de clasificación)**

**No supervisados: los que no se pueden clasificar.**

**Protocolos de comunicación:**

**Beat: mínima unidad de almacenamiento**

**Cuarta clase módulo 1.**

**Programación**

**Clasificaciòn de lenguajes:**

**Alto nivel (integra el compilador dentro del pc)**

**Medio nivel ( Compilador, transformaba las líneas de código en 0 y 1)**

**Bajo nivel ( 0 y 1)**

**Operación de asignación** es cuando le damos el valor a una variable

Todos los lenguajes pueden hacer propiedades aritméticas

Tipos de datos:

Primarios o simples: Son los **números (integer (enteros); Float (números con decimales (reales);** , luego los **booleanos** (se basan en unos y ceros) y los **textos (del tipo “”; ‘’, string)**

**Antes de Programar hay que pensar que vamos hacer**

**Algoritmo de diseñar para programar:**

**Paso 1: Identificamos el problema**

**Paso 2: Localizamos la formula**

**Paso 3: Observar cuales son las variables. Identificarlas**

**Ubicuidad: integración de dispositivos inteñigentes a nuestro entorno**

**Ejercicio 10**

**Desarrollar un programa que calcula el voltaje que cae en una resistencia cuando los valores de la intensidad y la resistencia son conocidos.**

**1. La Ley de Ohm nos indica que el voltaje = intensidad \* resistencia**

**2.V= I\*R**

**3. Datos de entrada son intensidad y resistencia**

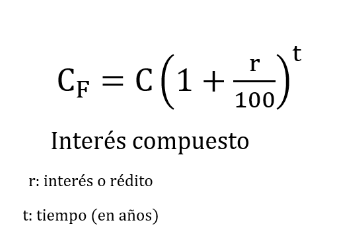
**4. pruebas**

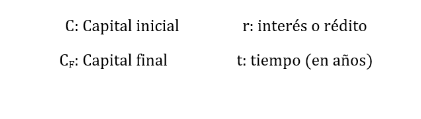
**5. Resultados**

**6. Banco de pruebas**

Otro ejercicio

**Calcular los intereses que tendrás si inviertes 100.000€ (capital Inicial) a un 2% (interes) de interés en 10 años (ciclos) debido a la fórmula del interés compuesto. Pista: Tienes un montón de información al respecto en Internet. Localiza la fórmula.**





**Pasos para programar en el lenguaje X. Utilizar siempre estos pasos**

1. Comentarios: Autor, fecha y objetivo (# en Python)
2. Zona de librerías
3. Zona funciones
4. Zona variables
5. Algoritmo
6. Print (). Resultado

**Nota: Siempre comentar todo**

**Estructuras de control**

Caracteres lógicos en programación

O= ||

Y= and, &

Not= ¡

Estructuras condicionales y estructuras repetitivas

Condicionales: **IF, ELIF, ELSE**

Repetitivas: **FOR, WHILE**

**Código Ejercicio Hipotenusa**

'''

1. comentarios

Autor: Santiago Gutierrez Cardona

Fecha: 12/01/23

Objetivo: Calcular la hipotenusa a partir de los datos de entrada

'''

#2.Zona libreria

#3. Funciones

#4. Variables

#in

a= 1

b= 0

#out

c= 0

#5. Algoritmo

a= a\*\*2

b= b\*\*2

c= (a+b)\*\*0.5

#6. Resultado

print (c)

Ejercicio anterior pero con entrada manual de datos

'''

1. comentarios

Autor: Santiago Gutierrez Cardona

Fecha: 12/01/23

Objetivo: Calcular la hipotenusa a partir de los datos de entrada

'''

#2.Zona libreria

#3. Funciones

#4. Variables

#in

textoa= input ('selecciones el valor de a')

textob= input ('selecciones el valor de b')

a=int(textoa)

b=int(textob)

#out

c= 0

#5. Algoritmo

#a= a\*\*2

#b= b\*\*2

c= (a+b)\*\*0.5

#6. Resultado

print (c)

**Código Ejercicio Voltaje**

'''

1. comentarios

Autor: Santiago Gutiérrez Cardona

Fecha: 12/01/23

Objetivo: Calcular la hipotenusa a partir de los datos de entrada

'''

#2.Zona libreria

#3. Funciones

#4. Variables

#in

i= 2

r=1

#out

v= 0

#5. Algoritmo

v= i\*r

#6. Resultado

print (v)

tips de lenguajes de programación

**Ejercicio 7**

Escribir un programa que calcule la Circunferencia de un círculo y muestre por pantalla el resultado. Si puedes con eso, hazlo para que también calcule el diámetro.

**C= 2\*Pi\*r**

**R: radio**

**Pi: numero pi**

**C=perímetro**

**Ejercicio 4**

'''

Haz una calculadora básica pida al usuario dos valores, a y b.

Según la opción que desean, realizar la operación:

Si operación es 1 entonces debemos ver el resultado de a + b

Si operación es 2 entonces debemos ver el resultado de a \* b

Si operación es 3 entonces debemos ver el resultado de a - b

Si operación es 4 entonces debemos ver el resultado de a / b

1. comentarios

Autor: Santiago Gutierrez Cardona

Fecha: 12/01/23

Objetivo:

'''

a=float(input("ingrese el primer valor"))

b=float(input("ingrese el segundo valor"))

resultado=0

operacion=int(input("Selecciones la operaciòn a ejecutar, 1 suma, 2 multiplicaciòn, 3 resta, 4 divisiòn"))

if operacion== 1:

resultado=a+b

if operacion== 2:

resultado=a\*b

if operacion== 3:

resultado=a-b

if operacion== 4:

resultado=a/b

print("El resultado de la operaciòn es:")

print(resultado)

**Ejercicio 9**

'''

Partiendo de la tarifa anual (que puede cambiar), nos piden que debemos calcular el precio de la tarifa de nuestro polideportivo, sabiendo las siguientes condiciones:

Criterio 1: Si es mayor de edad y está trabajando -> Paga el 100%

Criterio 2: Si es menor de edad y está trabajando -> Paga el 95%

Criterio 3: Si es mayor de edad y no está trabajando -> Paga el 75%

Criterio 4: Si es menor de edad y no está trabajando -> Paga el 50%

'''

preciototal= 1000

edad=int(input("ingrese su edad"))

trabajo=input("esta trabajando actualmente, responda si o no")

if edad>=18 and trabajo== "si":

tarifa= preciototal

print("Debe pagar:")

print(tarifa)

elif edad<18 and trabajo== "si":

tarifa= preciototal\*0.95

print("Debe pagar:")

print(tarifa)

elif edad>=18 and trabajo== "no":

tarifa= preciototal\*0.75

print("Debe pagar:")

print(tarifa)

elif edad<18 and trabajo== "no":

tarifa= preciototal\*0.5

print("Debe pagar:")

print(tarifa)

else:

print ("error, verifique el valor de la edad y la respuesta de empleo")

print("fin del programa")

**Ejercicio 10**

'''

La pizzería Bella Napoli ofrece pizzas vegetarianas y no vegetarianas a sus clientes.

Los ingredientes para cada tipo de pizza aparecen a continuación.

Ingredientes vegetarianos: Pimiento y tofu.

Ingredientes no vegetarianos: Peperoni, Jamón y Salmón.

Escribir un programa que pregunte al usuario si quiere una pizza vegetariana o no y en función de su respuesta

le muestre un menú con los ingredientes disponibles para que elija.

Solo se puede eligir un ingrediente además de la mozzarella y el tomate que están en todas la pizzas.

Al final se debe mostrar por pantalla si la pizza elegida es vegetariana o no y todos los ingredientes que lleva.

Autor: Santiago Gutierrez Cardona

Fecha: 12/01/23

'''

pizza= int (input ("Que tipo de pizza desea?, Escriba 1 para vegetariana y 2 para no vegetariana"))

if pizza==1:

ingredientes= int (input ("Que ingredientes desea adicionar ademas de mozzarella y tomate?, Escriba 1 para Pimiento y 2 para Tofu"))

if ingredientes==1:

print("Su pizza es vegetariana con mozarella, tomate y Pimiento")

print("Gracias por ordenar")

elif ingredientes==2:

print("Su pizza es vegetariana con mozarella, tomate y Tofu")

print("Gracias por ordenar")

elif pizza==2:

ingredientes= int (input ("Que ingredientes desea adicionar ademas de mozzarella y tomate?, Escriba 1 para Pepperoni, 2 para salmon y 3 para jamon"))

if ingredientes==1:

print("Su pizza es no vegetariana con mozarella, tomate y Pepperoni")

print("Gracias por ordenar")

elif ingredientes==2:

print("Su pizza es no vegetariana con mozarella, tomate y salmon")

print("Gracias por ordenar")

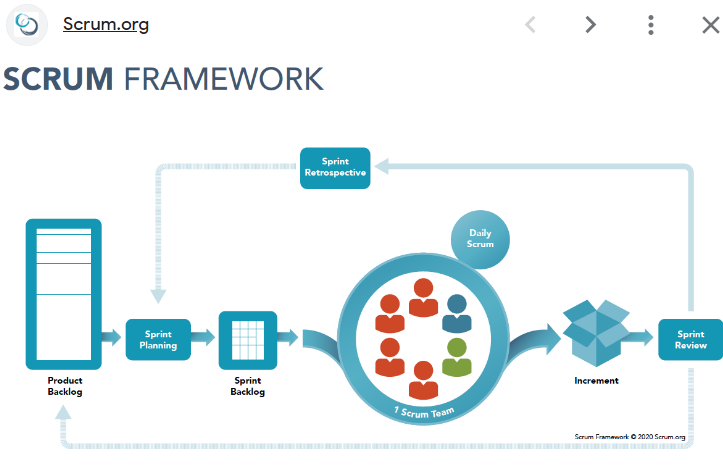
elif ingredientes==3:

print("Su pizza es no vegetariana con mozarella, tomate y jamon")

print("Gracias por ordenar")

**Clase 5. Evolución de bases de datos**

**Scrum** Framework



SCRUM es para gestionar el personal y las tareas

Product order

Product Manager

Backlog= Dividir en pequeñas tareas el problema para repartirlo entre el equipo.

**Algoritmo back propagation de una red neuronal. Investigar**

How to lie with stadistics. Ver y leer libro

GIT (Palabra que utilizan los de RRHH para saber si pasa filtro o no. Tener esta palabra en el curriculum. El git es para gestionar proyectos

Ver github

**ECOSISTEMA DATA**

Bases de datos transaccionales- OLTP

OLTP ( On-Line transaction processing) se caracteriza por bases de datos transaccionales con una gran cantidad de transacciones cortas (INSERT, READ, UPDATE, DELETE) online, suelen ser BD de sistemas críticos como ERP, CRM…

**Transaccion** es un conjunto de operaciones que se reralizan si todas se dan. Si no se da alguna de las operaciones ninguna se realiza.

**SISTEMAS ANALITICOS- OLTP**

No transaccionales

Tiempo de respuesta inmediato, de lectura (no se modifica el contenido normalmente)

Es orientada a análisis de:

Analisis de la organización

Previsiones de evolución

Diseño de estrategias

Operaciones pesadas de carga fuertes **ETL y ELT**

**ETL**

La mayoría de las empresas almacenan grandes cantidades de info

Existen dos enfoques posibles para procesar la información en una etl

BATCH PROCESING

STREAM PROCESING (Real time. Es mucho màs costoso)

¿Cuál utilizar?

Dependiendo del problema

BATCH PROCESING

Cuando es una herramienta analítica

No es importante…

**Tipos de bases de datos noSQL**

* **Basada en grafos** ( pelotas y flechas)
* **Documentales** y Columnarias (Apache es bueno para este tipo),
* **Clave-Valor** ( Mongo DB es buena para este tipo), ejemplo: motor de búsqueda

**Machine Learning. La IA**

Leer sobre:

* Algoritmos genéticos (Para eficiencia y optimización). Libros de Stanford.
* NLP

El Machine learning puede ser:

**supervisado** (solo se usan cuando necesito un **target claro** a predecir. Necesito tener los datos necesarios, organizados o preparados, los tipos de algoritmos supervidados son de clasificación y regresión). De clasificación son los algoritmos que necesito para predecir un número.

**no supervisado:**  Cuando necesito descubrir algo.

No necesita la intervención humana

Hacen parte de ellos:

Clustering

PCA.

Un IA es un modelo que se compone de datos y algoritmos

**Interview Warmup de Google**

No indexar hasta tener minino 15 publicaciones. Y ponerla .com

Módulo 3 BBDD y Python

Filas o registros

Columnas o campos

Condiciones primary KEY

* Tiene que ser único (código, DNI, Email, Numero de telofono
* No debe tener un campo vacío (no nublado)

Luego de haber decidido la primary Key hay que ver como relacionar la tabla

Foreing Key

Puede repetirse, no es única

**Tipos de relaciones**

Relación 1 a muchos (1 -> n) Solo se tiene una vez en una tabla, pero se puede repetir varias veces en las otras)

Relación varios a varios n 🡪 n

Relación 1 🡪1

**Tipos de interface**

GUI (Grafic user interface)

CLI (Comand Line Interface)

**Secciòn 2 (martes)**

SQLite

SELECT-FROM

**Select** ( que campos queremos seleccionar de que tabla)

**From ( NOMBRE TABLA)**

**Limit** (limita el número de campos que quieres ver). **Limit** siempre va a final

**AS** renombra los nombres de las columnas de la tabla a cambiar

**Distinc** Muestra el contenido único que hay dentro de un campo

**WHERE** Establece condiciones de filtrado para las filas en nuestro resultado final. Las condiciones están separados por operadores **AND/OR** ( las condiciones de where no están separados por comas),

**Si el contenido es en texto, se debe comer entre comilla simple**

**Order by,** va después del where, es para ordenar los campos siempre que existan en la tabla, puede ser ascendente o descendente, por defecto viene en ascendente.

**CASE WHEN STATEMENT**

El CASE WHEN statement permite crear campos con condiciones específicas, donde el valor que adquiera cada fila dependerá del criterio establecido. El statement se invoca utilizando la siguiente sintaxis

