**CMT 434 PYTHON FOR DATA SCIENCE**

**CONTINUOUS ASSESSMENT TEST**

**Student No: 1045858**

**Name: Brian Omondi**

Q1. a)

i. Artificial Intelligence analysis refers to the process of using AI models and algorithms to interpret and extract insights from data.

ii. Data Analysis is a process of inspecting, cleaning, transforming and modeling data with the goal of discovering useful information, suggesting conclusions and supporting decision-making.

iii. Big data is the collective name for the large amount of registered digital data and the equal growth thereof. The aim is to convert this stream of information into valuable information for the company.

b)

i.

- Structured Data: This type of data is highly organized, often stored in relational databases, and easily searchable due to its format (e.g., tables with rows and columns).
- Unstructured Data: This is data that lacks a predefined format, making it more challenging to process and analyze.
- Semi-Structured Data: This data has some level of organization but does not conform to strict models like structured data

ii. Big data

c) The variety of Big Data has led to the development of new Database Management Systems (DBMS) designed to handle complex data types. Traditional relational databases struggle with unstructured and semi-structured data. As a result:

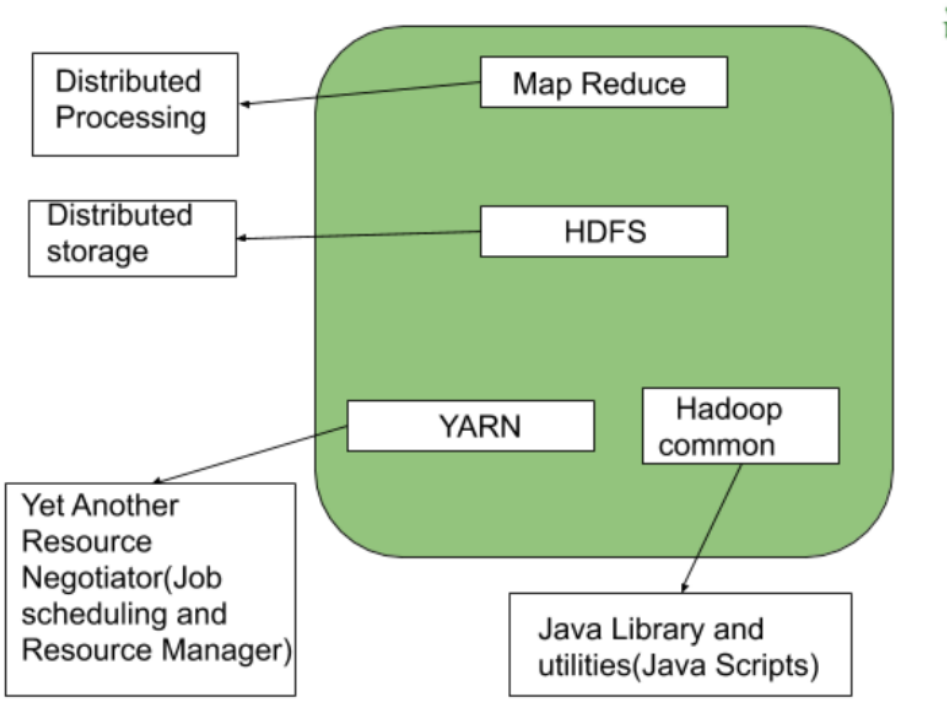
i. NoSQL Databases: These databases, such as MongoDB and Cassandra, were developed to manage unstructured and semi-structured data efficiently. They offer flexible schema design and are capable of handling large volumes of varied data types, such as documents, graphs, and key-value pairs.

ii. NewSQL Databases: Combining the strengths of SQL and NoSQL, NewSQL databases like Google Spanner aim to provide high scalability while supporting complex queries and maintaining the consistency associated with relational databases.

iii. Graph Databases: Used for managing data with complex relationships (e.g., social networks), graph databases like Neo4j can handle interconnected data that relational databases struggle with, enabling efficient processing of linked data.

iv. Cloud-based Databases: These solutions, offered by providers like AWS and Azure, allow for scalability and distributed data storage, addressing the need for storage and processing power that Big Data requires.

Q2. a)



- HDFS: Hadoop Distributed File System
- YARN: Yet Another Resource Negotiator
- MapReduce: Programming based Data Processing
- Spark: In-Memory data processing
- PIG, HIVE: Query based processing of data services
- HBase: NoSQL Database
- Mahout, Spark MLLib: Machine Learning algorithm libraries
- Solar, Lucene: Searching and Indexing
- Zookeeper: Managing cluster
- Oozie: Job Scheduling

b)

i. **Extensive Libraries**: Python has numerous libraries like Pandas, NumPy, Matplotlib, and Scikit-Learn, which simplify data manipulation, visualization, and machine learning.

ii. **Ease of Learning and Readability**: Python's simple syntax and readability make it accessible for beginners and allow data scientists to quickly write and understand code.

iii. **Strong Community Support**: Python has a large and active community that contributes to continuous library development and support for troubleshooting.

iv. **Integration and Compatibility**: Python integrates easily with other languages and databases, allowing data scientists to combine tools and manage diverse data sources effectively.

Q3. a)

- **Performance:** NumPy arrays are more memory-efficient and faster for numerical computations than Python lists. They are stored in contiguous memory blocks, allowing efficient processing.
- **Broad Functionality for Mathematical Operations**: NumPy provides optimized functions and operations (e.g., addition, multiplication) that can be applied to entire arrays at once, unlike Python lists, which require element-wise loops.

- **Multidimensional Array Support:** Unlike lists, NumPy arrays can handle multidimensional data (matrices, tensors) seamlessly, making them ideal for data science tasks involving matrices or higher dimensions.

b)

i. Machine Learning (ML) is that field of computer science with the help of which computer systems can provide sense to data in much the same way as human beings do. In simple words, ML is a type of artificial intelligence that extract patterns out of raw data by using an algorithm or method.

ii. A Restricted Boltzmann Machine (RBM) is a generative stochastic neural network used for dimensionality reduction, classification, and feature learning. It consists of visible and hidden layers with no intra-layer connections, meaning nodes within the same layer do not connect to each other. This lack of intra-layer connections is why it is termed "restricted," as it limits connections to only occur between layers, reducing computational complexity and making it easier to train.

iii. Backpropagation is an algorithm used to train neural networks by minimizing error. It works by calculating the gradient of the loss function and updating weights in the network from the output layer back to the input layer.

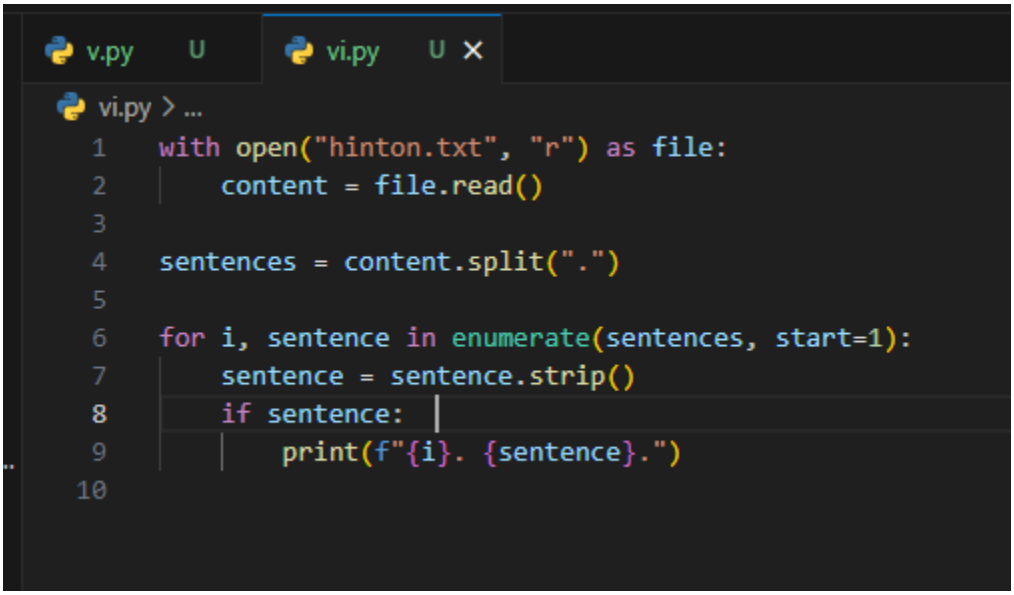
iv.

| **Aspect** | **Machine Learning-Based Systems** | **Knowledge-Based Systems** |
|---|---|---|
| Learning Approach | Learns from data through patterns and statistical methods | Relies on rules and domain knowledge |
| Data Dependence | Requires large datasets for training | Requires domain knowledge but not large datasets |
| Adaptability | Adapts and improves with more data | Limited adaptability; updates need manual effort |
| Decision Making | Based on probabilities and predictive analytics | Based on predefined rules and logic |

v.


v.py U
vi.py U
v.py > ...
print("Enter the narration (press Enter twice to end):")
narration = ""
while True:
line = input()
if line == "":
break
narration += line + "\n"
with open("hinton.txt", "w") as file:
file.write(narration)

vi.


v.py U
vi.py U
vi.py > ...
with open("hinton.txt", "r") as file:
content = file.read()
sentences = content.split(".")
for i, sentence in enumerate(sentences, start=1):
sentence = sentence.strip()
if sentence:
print(f"{i}. {sentence}.")