21CSE221T-Big Data T&T

CT1-Key

Part A

Q1. a) Semi-Structured. (Data with a discernible pattern)
b) Structured (Data in row-col format)
c) Unstructured (No inherent structure)
d) Quasi-structured (Effort and time to process and understand links)
1 mark each.

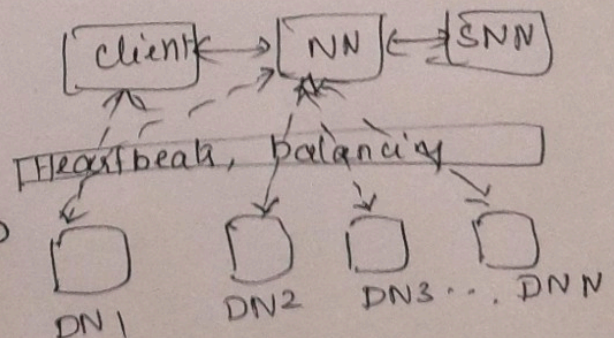Q2. Infrastructure as a Service (IaaS) (2marks)

Other options (2marks)

Paas

Saas

Q3. When there is no active stand by, it takes more time for the back up name node to start, when the primary name node fails in Hadoop 2 (2marks)

This was overcome in Hadoop 3 by having an active standby name node, which can immediately start functioning (2marks)

Part-B

Q4. HDFS Architecture.

Diagram – (4 marks)
Ref Unit1 ppt - Slide 57, 60

NameNode

DataNode

Secondary Name Node

Explanation on above (9 marks - 3 marks each)

(OR)

Q5.

i) Enterprise Big Data Mining & Steps (6 marks)

Big Data Mining - Entire life cycle. (2 marks)

- from procurement till impl.

Steps (4 marks)

1 to 5.

Q.5.ii) Process of selecting h/w and s/w stack. (7 marks)

Three broad models of h/w architecture (3 marks)

Selection of s/w stack - Options

Hadoop Ecosystem, Spark, NoSQL Cloud based solutions