A. Answer to the following questions (in one or a few sentences). (1*5=5)

1. What are the characteristics of big data?

Automatically generated by a machine

New source of data, such as the Internet

Not designed to be friendly

May not have many values

5 V's

Large and complex datasets that cannot be processed by traditional database applications.

Scale/Volume – data is increasing exponentially.

Complexity/Variety – data can be in many types such as strings, numbers, audio or images.

Speed/Velocity – Data is generated quickly and must be processed quickly. For example, data can be generated and processed in real time within sensors and social media.

Veracity – This is the amount of doubt that may be in the data. This can come in the format of bad or inconsistent data.

Value – This is the value of the data being held.

2. Give three differences between traditional and advanced security approaches.

Conventional	Advanced

Protect all information assets Preventive Controls No attackers get into the network	Focus protection on crown jewels Detective Controls Attackers sometimes get in, but are detected as early as possible and impact is minimized

3. How can Crypto be used for Big Data security?

Data-centric Security

Key Management

Data integrity and poisoning concerns

Searching/Filtering encrypted data

Security data collection/aggregation

Secure Collaboration

Proof of data storage

Secure outsourcing of computation

4. What is RTAP?

Real-Time Analytics Processing – Data is processed in real time such as streaming data. The response to this real time consumption of data is within milliseconds or seconds.

5. Tell a few advanced threats for Big Data.

Increase in advanced malware

Email based attacks are increasing with links and attachments that spread malware

Cybercriminals are increasing

Malicious emails are growing more diverse, evading detection.

B. Select True or False for the following statements (1*5=5)

6. Traditional security solutions cannot bridge the gaps between data breach protection and compliance.

True

7. Pattern matching is a function of big data.

True

8. The storage part of Hadoop is called MapReduce.

False, Hadoop File System (HDFS)

9. The goal of designing Hadoop is to process big data with reasonable cost and time.

True

10. Big Data is designed for access, not security.

True

C. Select appropriate answer(s) (1*5=5)

11. Which dimension of big data may correspond to the analysis of data security and intelligence

(c) Veracity

12. As organizations started adopting Hadoop at an increasing rate, the future use of Hadoop may need for additional capabilities, including:

(d) Improved security, workload management, and SQL support

13. Which statement is wrong in terms of Big data

b) Data is structured

14. What does "Velocity" in Big Data mean?

d) Speed of storing and processing data

15. Which of the following is true?

c) Map >Combine >Reduce