


Name:			
Enrolment No:			
<div><div>UPES</div><div>End Semester Examination, May 2025</div><div><div>Course: Big Data Security</div><div>Semester: 6th</div><div>Program: B. Tech CSE (H & NH) (BIG DATA)</div><div>Course Code: CSBD 3014</div><div>Calculator allowed: Yes</div><div>Instructions: Please attempt according to the time provided and given weightage.</div></div><div><div>Time : 03 hrs.</div><div>Max. Marks: 100</div></div></div>			
SECTION A (5Qx4M=20Marks)			
S. No.		Marks	CO
Q 1	Define the 3-2-1 backup rule and explain how the Principle of Least Privilege (POLP) should be applied to secure backup systems under this rule.	4	CO3
Q 2	Describe briefly the integrity checks in ingestion pipelines.	4	CO1
Q 3	Explain the two major weaknesses of Discretionary Access Control.	4	CO1
Q 4	What is data lineage? Explain the challenges in Data lineage.	4	CO1
Q 5	A distributed healthcare system processes patient records across cloud and edge nodes. Outline two key context elements needed to prioritize risks effectively and justify their impact on control selection.	4	CO4
SECTION B (4Qx10M= 40 Marks)			
Q 6	<p>Information Asset B has an asset value of 100. It is exposed to the following two vulnerabilities: Vulnerability 1 has a likelihood of 0.5, and current controls are in place that mitigate 50% of its risk. Vulnerability 2 has a likelihood of 0.1, but currently, no controls are applied. The accuracy of the assessment data and assumptions used in the analysis is estimated to be 80%.</p> <p>a) Calculate the residual likelihood for both vulnerabilities after considering the effectiveness of controls. b) Compute the residual risk score for each vulnerability. c) Calculate the total residual risk score for Asset B. d) Apply the 80% data accuracy factor to determine the final adjusted risk score. e) Estimate the range of possible risk scores by assuming a $\pm 10\%$ uncertainty in both the asset value and the control effectiveness.</p>	[5*2=10]	CO4

Q 7	<p>(a) For a uniform 9-bit hash function (output range: 0 to 511), calculate the minimum number of unique inputs (N) required to achieve at least a 77% probability of a collision.</p> <p>(b) Consider the hash function f defined as:</p> $f(x) = \begin{cases} 0 \parallel x, & \text{if } x \text{ is exactly } n \text{ bits long} \\ 1 \parallel g(x), & \text{otherwise} \end{cases}$ <p>where $g()$ is a collision-resistant hash function (e.g., SHA-3), and \parallel denotes concatenation.</p> <ol style="list-style-type: none"> For $n=4$, compute: $f(1101)$ and $f(10101)$ Prove that f is not pre-image-resistant for n-bit inputs but is second pre-image resistant for all inputs. 	[5+1+4=10]	CO4
Q 8	Explain Apache Sentry's authorization framework in detail, covering its core components, working mechanism, and integration with Hadoop ecosystem.	10	CO1
Q 9	<p>Discuss the concept of Homomorphic Encryption and its variants and analyze its significance in preserving data privacy during computation.</p> <p style="text-align: center;">OR</p> <p>Discuss various levels of encryption at data at rest. Explain the different components of HDFS level encryption.</p>	[5+5=10]	CO3
SECTION-C (2Qx20M=40 Marks)			
Q 10	Analyze the internal architecture of Kerberos with a suitable diagram and related protocols in detail.	20	CO3
Q 11	<p>a) Define and explain the core components of an access control system.</p> <p>b) Compare and contrast the following access control models with suitable real-world examples:</p> <ol style="list-style-type: none"> Discretionary Access Control (DAC) Mandatory Access Control (MAC) Role-Based Access Control (RBAC) <p>c) Analyze the advantages and limitations of each model in the context of the distributed system.</p> <p style="text-align: center;">OR</p> <p>Explain the following in brief:</p> <ol style="list-style-type: none"> CIA Triad Merkle-damaged construction Kerberos trust Security challenges in NoSQL Databases 	[5+5+5+5=20]	CO2