

DISTRIBUTED COMPUTING AND BIG DATA
2023 09:10 AM. MAX MARKS: 10.

DEADLINE: MAR {24, 25, 27},

Instructions:

- (1) Submit your assignment slides as a single pdf file on moodle. Clearly mention your roll number and name in the solution pdf.
- (2) No more than two teams can sign up for the same topic.

Deliver a 20 minute presentation (+ 5 minute Q&A) on one of the following topics:

- (1) Data Warehouse
- (2) Data Lakes
- (3) Big Data at Hotstar
[Scaling hotstar.com for 25 million concurrent viewers](#)
- (4) Big Data at Netflix
[Unreeling netflix: Understanding and improving multi-CDN movie delivery, Big Data at Netflix: Faster and Easier](#)
- (5) Big Data at Uber
[How Uber scaled its Real Time Infrastructure to Trillion events per day?](#)
- (6) Big Data in Media
[Data Science at the New York Times, The Future of Data Science in the Media](#)
- (7) Big Data & Analytics for Finance
- (8) Big Data in Sports
- (9) Big Data in Healthcare
- (10) Microsoft Azure
[Keynote: Inside Microsoft Azure Datacenter Architecture](#)
- (11) Google cloud platform (GCP)
- (12) Amazon Web Services (AWS)
- (13) Apache Spark
- (14) Kubernetes
- (15) ChatGPT

Although some suggestive resources are referenced here, feel free to explore and refer a variety of resources you could get to. Try to stay away from AI and ML wherever possible. Focus on big data tools and technologies.
