In this course we have used 4 data storage technologies and associated query languages:

1) mongoDB with its query language <https://www.tutorialspoint.com/mongodb/mongodb_query_document.htm> ( assume we used it on a cluster installed on 100 machines)

2) Hadoop distributed file system with csv files accessed with the PIG language (assume we used it on a true Hadoop cluster and the Hadoop distributed file system installed on 100 machines in a cluster)

3) RDBMS - Microsoft SQL Server using the SQL query language (assume we used it on a high performance 32 way symmetric multiprocessor.)

4) Data Cube – Microsoft Analysis Server using the MDX language (assume we used it on a high performance 32 way symmetric multiprocessor.)

Compare and contrast the advantages and disadvantages of these storage and query systems for solving the following problems.

[A] Storing real time events streaming in at very high rates (e.g. all the point of sale data from 100’s of Target department stores, twitter replies and forwards from a Taylor Swift tweet) and posting real time statistics gathered from that data.

[B] Guaranteeing the accuracy and completeness of 1000’s of transactions (bank ATM events, stock trades)

[C] Analyzing the last weeks sales from 100’s of Target department stores and deciding which items should have their prices raised or lowered.

[D] Allowing an Amazon customer to look up the details and status of any currently outstanding orders they have.

[E] Storing the responses to an online survey posted by a major league soccer team with 25 questions that the form does NOT require the user to answer every question and with the expectation that at least 500,000 fans will fill in the survey the day it is posted.