B part:

Assignment: Two topics

1. Data mining contest
Description: Check here and take part in.

Scoring: Highest rank group chose this topic gets full 50pts, and other groups chose this topic get score based on the gaps with the highest rank group.

Notice:
- All groups which take part in this contest must name their group name in the format USTCnibaba-* and provide group names to dmdw_ustc2019@163.com to prevent cheating.

- We will check the ranking of 12:00, Feb 14 2020 as final result.

2. Better GraphX performance
Description: Design a function edgeQuery(id: Int) : Edge, and something behind it. This function accepts a parameter and return the edge at position id in EdgeRDD. Use this graph with edges ordered ascending with source vid as primary key and destination vid as secondary key to organise an edgeRDD, finish our 1000000 not so random queries and you finish it.

Scoring: The group chose this topic finishing 1000000 random edge queries in a shortest time gets full 50pts, and other groups chose this topic get score based on the gaps with the fastest group.

Notice:
- More details of that dataset is shown here.

- The random query sequence for evaluating your works including 1000000 random integers will be provided by us when examining your program face to face to prevent cheating.

- We will use the time rank at Feb 14 2020 as final result.

- Also, we will have a correct result list to check if your program gives correct edge output.

- The purpose of this topic is to design some external cache mechanism for faster data access. If your works involved some modifications in Spark core, please make a clear explanation to TAs for a A part scoring standard.