CZ/CE 4123 Tutorial 1 : Big Data 5V's

Question 1

Which of the following are related to big data applications, why?

(1) A computer scientist at MIT is trying to prove NP!=P (note: a famous computer science conjecture).

(2) A manager in Walmart wants to understand the purchasing behavior of customers through the purchase records in 2021.

(3) A programmer spends 10 hours in debugging his code.

(4) Artificial Intelligence can play games.

Question 2 [Open Question]

Is it possible that an application is not considered a big data application now, but can possibly become a big data application in the future?

Question 3

For each of the following descriptions, which of the 5V's it is the most related?
(1) There are more than 60000 searches per second in the Google search engine.

(2) Suppose there is a database storing all kinds of enterprise related data, including the ratios of male/female in different companies. Ben wrote the following C program to collect the ratio of male/female for a company.

```
int cal_ratio(){
        int num_Male=getMales();
        int num_Female=getFemales();
        return num_Male/num_Female;
}
```

(3) ImageNet is a 150GB dataset that holds 1,281,167 images for training and 50,000 images for validation, organized in 1,000 categories.

(4) A website allows users to upload different forms of documents such as Excel, JPG, PDF, video.

(5) A researcher performs data mining algorithms on Amazon's purchase records, and he successfully predicts the best seller in the next month.