Final Report

Understanding of Big Data fields

Before the course, I had a deeply misunderstanding about big data. I used to suppose that big data only refers to large sets of data and the only thing I heard about it was Hadoop.

However, the importance of big data doesn't simply revolve around how much data we have. The value lies in how we use it. It is not only a term that describes large, hard-to-manage volumes of data – both structured and unstructured – that inundate businesses on a day-to-day basis, but also, big data can be analyzed for insights that improve decisions and give confidence for making strategic business moves.

The IT industry, in an attempt to quantify what is and isn't Big Data, has come up with what are known as the "V's" of Big Data. The foundational three are:

- **Volume:** The amount of data is immense. According to sources each day 2.3 trillion gigabytes of new data is being created.
- **Velocity:** The speed of data and processing (analysis of streaming data to produce near or real time results)
- **Variety:** The different types of data, structured, as well as, unstructured.

Today, big data is important and efficient mainly because it is :

Cost reduction. Big data technologies such as Hadoop and cloud-based analytics bring significant cost advantages when it comes to storing large amounts of data

Faster, better decision making. With the speed of Hadoop and in-memory analytics, combined with the ability to analyze new sources of data, businesses are able to analyze information immediately and make decisions based on what they've learned.

New products and services. With the ability to gauge customer needs and satisfaction through analytics comes the power to give customers what they want.

Impressive issue in the course

Over this course period, the most impressive issue to me is the **blockchain**.

I heard the word "blockchain" a lot these days but never got to know what it is. During this course, I got a more comprehensive understanding about it.

Blockchain is a peer-to-peer decentralized distributed ledger technology that makes the records of any digital asset transparent and unchangeable and works without involving any third-party intermediary. It is an emerging and revolutionary technology that is attracting a lot of public attention due to its capability to reduce risks and frauds in a scalable manner.

Also, it is related to another popular word -- **bitcoin**

Bitcoin and Blockchain

Although the advent of Blockchain has taken the world by storm, many people still get confused about these two terms. Thus, it is important to understand how these terms differ and how they are interrelated.

Bitcoin is a cryptocurrency, which is an application of Blockchain, whereas Blockchain is simply an underlying technology behind Bitcoin that is implemented through various channels.

Help to future career

Personally, I am determined to be a front-end engineer. After learning this course, I realize the importance of **Data Virtualization**

Under the wave of big data, one of the most obvious characteristics is the exponential growth of data. This trend can be seen from the figure above. The challenge that follows is how to display data more vividly and interactively, which is what we usually Talk about "data visualization". Nowadays, We have seen all kinds of charts. Visual charts are created from the entire process of data -> cleaning -> interaction -> vision -> development.The front end of big data projects is generally used for data visualization, so libraries related to visualization icons (such as D3.js, G2, Echarts) are best to be familiar with.

To some extends, it offers me a direction to pay efforts to, and I gained a lot about that.