

Section 1: Theory

1. Data Moat

From my point of view, the term “moat” refers to business competitive advantage, as a result; data moat is business competitive advantages derived from utilizing data. The data moat can be treated under three topics:

- The data moat is created by combining proprietary (1st-party) data sources with others. A company is able to collect a huge amount of data; moreover, the process of combining data is unique (internal and external sources). Therefore the business insight generated is also various.
- Furthermore, we could consider how a company transform its data into business practices: building, improving and innovating customer services, using a more scientific, measurable and automated decision-making approaches in different fields such as marketing – targeting customers; efficient operations; generating actionable insights and people analytics in HR, these all benefits are fueled by AI and data

To sum up, “Competitors can copy your product but they can’t copy your data”.

2. OLTP vs OLAP

While OLTP database characterized by a large number of short on-line transactions, OLAP Database characterized by a relatively low volume of transactions. Furthermore, we could observe that OLTP database is more well-organized with fewer attributes and features; while OLAP database is enriched in terms of features engineering.

Additionally, because the queries are concluded into inserting, updating and deleting; the main concentration of building OLTP systems is to put on fast query processing, maintaining data integrity in multi-access among different environments and an effectiveness measured by a number of transactions per second. As a result of a wide database, OLAP queries are often very complex and involve aggregations.

Last but not least, OLTP Records are narrow and long which are suitable for real-time business operations; OLAP records are short and wide which mainly supports the purpose of data analysis, finding actionable business insights and decision-making.

3. A modern data team - data professionals.

A modern data team consists of data engineers, data scientists and data analysts.

Firstly, Data engineer is also called an ETL "Extract, transform, load" engineer who is responsible for moving and propagating access to data. They handle the beginning step of data processing: collecting; moving and storing data. As a result, it is a combination of computer science and database experience.

Secondly, data scientists rely more on mathematics/ statistics in order to develop predictive models or automated classifications. AI and Deep Learning are typical topics of data scientists.

The third one is a data analyst, who handles business questions by applying data. The three main languages are SQL, Python and R play a stellar role in business. Furthermore, a data analyst must be proficient in data visualization for not only discovering business insights but also explaining them to audiences – other departments.

4. Data Privacy

From my personal view, although data privacy law has been mentioned for a long time and governments attempt to develop it, regulation typically lags behind innovation. Tech giants such as Google, Facebook, Baidu... are “too big to fail”. They even have close connections with governments. As a result, users demand more competitiveness for them. Due to the reality that regulations could not catch up the innovations, however; a competitive market could protect and give more benefits to users. For many users, they have to accept the data privacy agreement given by these big companies, otherwise, they have no other options to satisfy the same customer’s needs.

5. Car Owners and Cars

Car and Car Owners relationship is M: N, we could see that a person is able to own more than one car, while a car is manufactured in mass production so that it could be sold to as many customers as within the company’s production capacity. In this case, we assume that a car is defined by its produced year; model and manufacturers. If we identify a car by series number or Vehicle registration plate; and period of time we could transform the many-to-many relationship (M: N) to 1:1. The middle entity should have a vehicle registration plate and time.