Data extraction

Anny Yao

Data extraction is the process of collecting raw data from a variety of sources, aggregating it, and converting it into a usable and meaningful format (databases, Excel spreadsheets, etc.) for further search, analysis, and storage.[1](Matillion, 2024)Businesses across industries rely on data to make decisions, and organizations can more effectively access key information to make more efficient, intelligent, and accurate decisions. It is an important step in the evolution of today's digital age.

First, through data extraction, companies can access customer behavioral data to improve the accuracy of their decisions. A typical example is the e-commerce industry. Many companies in this market collect customers' browsing history on websites, media interactions, past purchases, and so on. This data can help companies analyze consumption trends, predict consumer needs, optimize products, and increase sales. Using incremental extraction methods (one of the data extraction methods used to extract only the data that has changed since the last extraction, thereby reducing the amount of data transferred and processed compared to other extractions), retailers can update customer preferences and improve product offerings.), retailers can update customer preference data in real time, eliminating the burden of dealing with redundant information and large-scale data synchronization. [2](Astera, 2024)

Second, data extraction can be performed in an automated manner, increasing the efficiency of many organizations. In the healthcare industry, healthcare organizations extract data

---

[1] https://www.matillion.com/blog/what-is-data-extraction-everything-you-need-to-know, October 20, 2024
[2] https://www.astera.com/zh-CN/type/blog/what-is-data-extraction-a-brief-guide/, October 20, 2024

from multiple systems, such as data and reports from a patient's electronic health record, and integrate it for diagnostic support, treatment plan optimization, and public health analytics. By automating data extraction tools (which can be scheduled to run at specific intervals or triggered by specific events, reducing the need for manual intervention and ensuring that data is continuously updated), hospitals can reduce manual data entry errors and improve efficiency.[3](Stitchdata, 2024) And the combination of batch processing and cloud storage ensures the security and stability of medical data during extraction and storage.

Then, data extraction is integrated with different data formats, which can be used in future development by combining streaming data processing technology (data sent in large quantities in a continuous, uninterrupted incremental manner with low-latency processing.), companies can monitor and analyze data in real time, especially in industries such as finance and e-commerce that require rapid changes in strategy. Data extraction can also be applied in the process of machine learning technology, where the system can automatically identify and optimize extraction rules by learning the patterns of historical extracted data to improve extraction efficiency and data quality.[4](AWS, 2024)

Overall, data extraction is gradually becoming the core of business. Through effective data extraction, enterprises can not only improve the accuracy of business analysis, but also reduce manual intervention through automation and improve operational efficiency. As the data era continues to advance, data extraction will bring greater potential for innovation across industries.

---

[3] https://www.stitchdata.com/resources/what-is-data-extraction/, October 20, 2024
[4] https://aws.amazon.com/cn/what-is/streaming-data/,Octorber 20 2024

References

https://www.matillion.com/blog/what-is-data-extraction-everything-you-need-to-know,

Matillion, October 20, 2024

https://www.astera.com/zh-CN/type/blog/what-is-data-extraction-a-brief-guide/, Astera,

October 20, 2024

https://www.stitchdata.com/resources/what-is-data-extraction/, Stitchdata, October 20,

2024

https://aws.amazon.com/cn/what-is/streaming-data/, AWS, October 20, 2024