

Multiple Source Data Processing and Integration

Author, Muhammad Haroon

A Data Science Foundation Blog

December 2020

www.datascience.foundation

Data Science Foundation

Data Science Foundation, Atlantic Business Centre, Atlantic Street, Altrincham, WA14 5NQ
Tel: 0161 926 3670 Email: admin@datascience.foundation Web: www.datascience.foundation
Registered in England and Wales 4th June 2015, Registered Number 9624670

Copyright 2016 - 2017 Data Science Foundation

Data blending is a recently developed approach used mainly by those working in big data analytics. It's a method of integrating data from different sources into a single structure. Blending offers a relatively fast and straightforward way to access several disparate data sources and identify correlations between them without the trouble and costs of conventional data integration. Ideally, data blending is carried out using data preprocessing software developed for non-technical data analysts who need to prepare data interactively without any programming.

Interactivity is an essential aspect of data blending. Working with visual interfaces and in real-time, the analyst mixes, visualizes, and analyses data in cycles – combines, visualizes, analyses, changes blending based on research and repeat this process. Interactive in this instance includes that the analyst is communicating with the data.

Data processing from multiple sources somehow connects with data warehousing. Data warehousing is a mechanism that physically incorporates data as either a data store hub with dependent data marts or as an alternate to the set of compliant data marts using the bus structure. Tailored data marts mitigate (but do not eradicate) the "least common denominator" impact on the joints. But warehousing causes data latency as data processing and initialization is conducted as periodic batch processing. More specifically, warehousing reduces interactivity because the function of the data combination is shifted away and isolated from data display and analysis.

Data integration is the method of integrating data from various, incompatible sources to a data warehouse destination. It is necessary to turn raw data into useful information that drives more robust, quicker decision-making.

There is no generic process for the integration of data. However, data integration strategies usually include a few standard features, such as a data source network, a master server, and clients analyzing information from a master server.

In a traditional data integration process, it requests data to the master server. The master server then gathers the necessary data from various sources. The data is collected from the sources and compiled into a single coherent data set. This data set is being offered back to the client for use.

Data Science Foundation

Data integration tools and technologies are built to transform raw data into insights. Business reports and dashboards show decision-makers what's going on in their company, and they can even shed some light on why it's going on.

Data is a valuable resource for the company. Still, its utility is restricted unless the data is up-to-date, reliable, and accessible to everyone in the company who can profit from its analysis.

Enterprise data is also segregated in organizational departments: financial data remains in the financing department, and marketing data remains in the marketing department. Data integration dumps knowledge silos by making the data accessible to more decision-makers that can use BI resources to gain insight into process improvement, improving goods, saving money, and finding hidden opportunities. By gathering and evaluating data from different sources through aggregation, business leaders may uncover trends and patterns that would have been challenging to pinpoint earlier.

Data integration is closely linked to interoperability, including technological, semantic, legal, and institutional levels. The semantic level is all about guaranteeing that the data's structure and context are maintained and understood. The technical standard includes data sharing and data integration facilities. The legal level tackles factors such as allowing cooperation in terms of diverse traditional structures and organizational policies. Concerning this degree of interoperability, the confidentiality of the research data of users is an essential consideration. Finally, organizational interoperability involves aligning procedures with common organizational priorities and communicating user preferences and specifications. Data integration and interoperability are critical complaints when dealing with Big Data and are not unique to any particular domain. In an educational environment, gathering and integrating data from different sources will provide perspectives that have implications for learning, teaching, retention, and program design.

And if an organization collects all the data it wants, the data also remains in various data sources. For example, for a standard customer 360 view usage case, the data to be compiled could include data from their CRM systems, internet traffic, marketing operations systems, customer—apps, sales, and customer satisfaction systems, and even partnership data, to highlight a few. Information from all these various sources also needs to be gathered for analytical needs or step in implementation. It can be no quick operation for data engineers or developers to put them all together.

Employees throughout every department—and even in separate physical locations—frequently need access to the company data for joint and individual initiatives. IT needs a safe solution for distributing data through self-service connectivity throughout all business lines.

Data Science Foundation

Also, almost every department is producing and developing data that the rest of the company needs. Data integration has to be collaborative and cohesive to enhance coordination and unification within the enterprise.

When an organization takes steps to integrate the data better, it dramatically reduces the time required to process and analyze data. Automation of centralized views eliminates the need for manual data collection. Workers no longer need to establish links from the start whenever they need to execute a report or construct an application.

There's a lot to keep up with when it comes to corporate data services. To collect data manually, workers should understand each location and profile they may need to discover all the appropriate software-enabled until they start—to ensure that their sets of data are true and accurate. When a data repository is introduced, and the worker is not informed of it, the data set would be insufficient.

Also, without even a data integration mechanism that synchronizes data, the reporting needs to be revised periodically to compensate for any changes. However, with automatic updates, reports can be generated conveniently in real-time, whenever they are required.

Data integration efforts enhance the value of business data over time. When data is integrated into some centralized framework, quality problems are detected. The required changes are made, resulting in more reliable data—a foundation for quality evaluation and analysis.

About the Data Science Foundation

The Data Science Foundation is a professional body representing the interests of the Data Science Industry. Its membership consists of suppliers who offer a range of big data analytical and technical services and companies and individuals with an interest in the commercial advantages that can be gained from big data. The organisation aims to raise the profile of this developing industry, to educate people about the benefits of knowledge based decision making and to encourage firms to start using big data techniques.

Contact Data Science Foundation

Email: admin@datascience.foundation

Telephone: 0161 926 3641

Atlantic Business Centre

Atlantic Street

Altrincham

WA14 5NQ

web: www.datascience.foundation

Data Science Foundation

Data Science Foundation, Atlantic Business Centre, Atlantic Street, Altrincham, WA14 5NQ

Tel: 0161 926 3670 Email: admin@datascience.foundation Web: www.datascience.foundation

Registered in England and Wales 4th June 2015, Registered Number 9624670