WHAT IS DATA NORMALIZATION?



It's no secret. We are officially living in the era of big data. Nearly every business—especially large-scale enterprises—collects, stores, and analyzes data for the benefit of growth. In most daily business operations, managing data is a norm, using tools such as:

- Databases
- Automation systems
- CRM platforms

If you have worked in any company for some time, then you've probably encountered the term Data Normalization. A best practice for handling and employing stored information, data normalization is a process that will help improve success across an entire company.

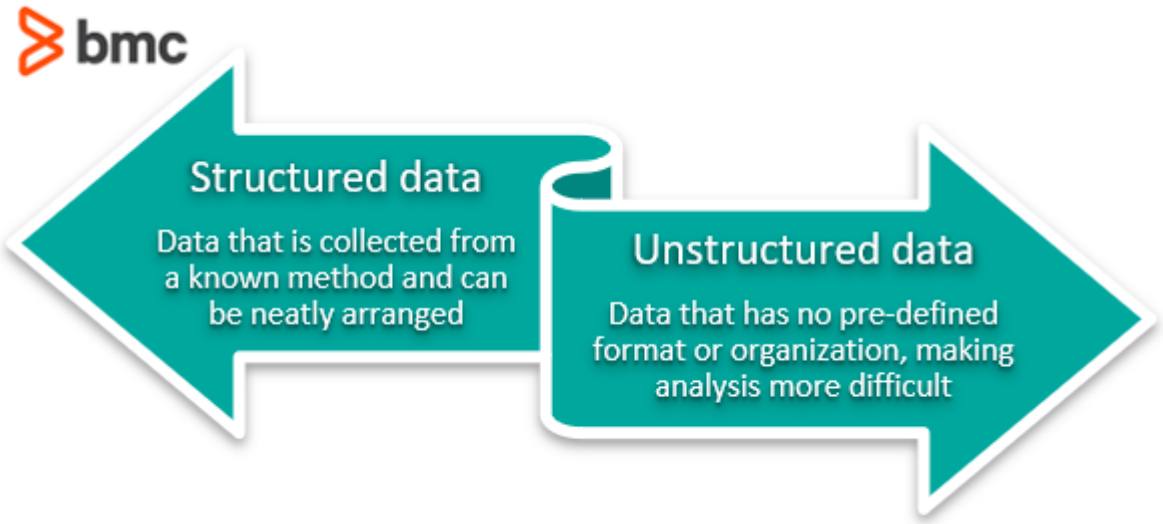
Here is everything you need to know about data normalization along with some tips on how to improve your data effectively.

What is data normalization?

Data normalization is generally considered the development of clean data. Diving deeper, however, the meaning or goal of data normalization is twofold:

1. Data normalization is the organization of data to appear similar across all records and fields.
2. It increases the cohesion of entry types leading to cleansing, lead generation, segmentation, and higher quality data.

Simply put, this process includes eliminating unstructured data and redundancy (duplicates) in order to ensure logical data storage. When data normalization is done correctly, you will end up with standardized information entry. For example, this process applies to how URLs, contact names, street addresses, phone numbers, and even codes are recorded. These standardized information fields can then be grouped and read swiftly.



Who needs data normalization?

Every business that wishes to run successfully and grow needs to regularly perform data normalization. It is one of the most important things you can do to get rid of errors that make running information analysis complicated and difficult. Such errors often sneak up when changing, adding, or removing system information. When data input error is removed, an organization will be left with a well-functioning system that is full of usable, beneficial data.

With normalization, an organization can make the most of its data as well as invest in data gathering at a greater, more efficient level. Looking at data to improve how a company is run becomes a less challenging task, especially when cross-examining. For those who regularly consolidate and query data from software-as-a-service applications as well as for those who gather data from a variety of sources like social media, digital sites, and more, data normalization becomes an invaluable process that saves time, space, and money.

Are you ready to harness the power of data? See how DataOps with BMC can transform your analytics.

How data normalization works

Now is the moment to note that, depending on your specific type of data, your normalization will look differently.

At its most basic, normalization is simply creating a standard format for all data throughout a company:

- Miss EMILY will be written in Ms. Emily
- 8023097864 will be written 802-309-7864
- 24 canillas RD will be written 24 Canillas Road
- GoogleBiz will be written Google Biz, Inc.
- VP marketing will be written Vice President of Marketing

Beyond basic formatting, experts agree that there are five general rules or “normal forms” to performing data normalization. Each rule focuses on putting entity types into number categories depending on the level of complexity. Considered to be guidelines to normalization, there are instances when variations from the form need to take place. In the case of variations, it is important to consider consequences and anomalies.

For the purposes of complexity, in this article, the first and three most common forms are discussed at a top-level and all data is considered in table format.

1. First Normal Form (1NF)

The most basic form of data normalization is 1NFm which ensures there are no repeating entries in a group. To be considered 1NF, each entry must have only one single value for each cell and each record must be unique.

For example, you are recording the name, address, gender of a person, and if they bought cookies.

2. Second Normal Form (2NF)

Again working to ensure no repeating entries, to be in the 2NF rule, the data must first apply to all the 1NF requirements. Following that, data must have only one primary key. To separate data to only have one primary key, all subsets of data that can be placed in multiple rows should be placed in separate tables. Then, relationships can be created through new foreign key labels.

For example, you are recording the name, address, gender of a person, if they bought cookies, as well as the cookie types. The cookie types are placed into a different table with a corresponding foreign key to each person's name.

3. Third Normal Form (3NF)

For data to be in this rule, it must first comply with all the 2NF requirements. Following that, data in a table must only be dependent on the primary key. If the primary key is changed, all data that is impacted must be put into a new table.

For example, you are recording the name, address, and gender of a person but go back and change the name of a person. When you do this, the gender may then change as well. To avoid this, in 3NF gender is given a foreign key and a new table to store gender.

As you begin to better understand the normalization forms, the rules will become more clear while separating your data into tables and levels will become effortless. These tables will then make it simple for anyone within an organization to gather information and ensure they collect correct data that is not duplicated.

Benefits of data normalization

As mentioned above, the most important part of data normalization is better analysis leading to growth; however, there are a few more incredible benefits of this process:

More space

With databases crammed with information, organization and elimination of duplicates frees up much-needed gigabyte and terabyte space. When a system is loaded with unnecessary things, the processing performance decreases. After cleaning digital memory, your systems will run faster and load quicker, meaning data analysis is done at a more efficient rate.

Faster question answering

Speaking of faster processes, after normalization becomes a simple task, you can organize your data without any need to further modify. This helps various teams within a company save valuable time instead of trying to translate crazy data that hasn't been stored properly.

Better segmentation

One of the best ways to grow a business is to ensure lead segmentation. With data normalization, groups can be rapidly split into categories based on titles, industries—you name it. Creating lists based on what is valuable to a specific lead is a process that no longer causes a headache.

Data normalization is not an option

As data becomes more valuable to all types of business, the way it is organized in mass qualities can not be overlooked.

From ensuring the delivery of emails to preventing misdials and improving analysis of groups without the worry of duplicates, it is easy to see that when data normalization is performed correctly it results in better overall business function. Just imagine if you leave your data in disarray and miss important growth opportunities due to a website not loading or notes not getting to a VP. None of that sounds like success or growth.

Choosing to normalize data is one of the most important things you can do for your organization today.

Additional resources

For more on this topic, explore these resources:

- BMC Machine Learning & Big Data Blog
- Big Data vs Analytics vs Data Science: What's The Difference?
- Data Management vs Data Governance: Main differences
- 4 Reasons to Automate the Ingestion of Data
- 3 Keys to Building Resilient Data Pipelines
- Data Visualization and Tableau Online, two multi-part Guides