SCSA4003- BUSINESS ANALYTICS

UNIT 2- TYPES OF DIGITAL DATA

UNIT 2 TYPES OF DIGITAL DATA 9 Hrs.

Definition, Sources, Storage and Characteristics of data - Structured, Unstructured and Semi Structured Data - Importance of data quality - Dealing with missing or incomplete data - Data Classification Data Warehouse: Definition, characteristics, framework- Data lake - Business Reporting, Visual Analytics: Definition, concepts - Different types of charts and graphs - Emergence of data visualization and visual analytics

Definition:

Digital data, in information theory and information systems, is information represented as a string of discrete symbols each of which can take on one of only a finite number of values from some alphabet, such as letters or digits. An example is a text document, which consists of a string of alphanumeric characters .

Digital data is data that represents other forms of data using specific machine language systems that can be interpreted by various technologies. The most fundamental of these systems is a binary system, which simply stores complex audio, video or text information in a series of binary characters, traditionally ones and zeros, or "on" and "off" values.

Digital data storage

What is digital data storage?

Digital data storage is mostly offline storage for backup and fail-safe data. In recent years, digital data has evolved to include cloud storage. It is a server that hosts all types of uploaded data, including media files. It generally exists to serve companies that have large amounts of data and need it protected and backed up.

How is digital data stored?

In a technical sense, data is stored as code or numbers for a computer to read and control. It's then guided based on the computer input rules and stored in different locations. Data within files can be stored offline in different drive types, on a physical location like a hard drive and online in the cloud.

Different ways to store data

There are different types of data storage and it's important to understand how they contrast from one another. RAM, or memory, is temporary data storage so the computer can quickly access it. The data stored here isn't permanent. It instead allows a computer to read data fast, opposed to the slower alternatives of storage.

The other type of storage is a device like a hard drive, which holds permanent data unlike the RAM storage. This storage is potentially mobile, as the drive can be external as well as internal. Personal data within different types of files including media generally are stored onto things like the hard drive or external drives for future use. It also helps the uploader retrieve data quickly in the future.

Alternatives to traditional storage

Digital data storage began as a computer storage technology for audio. It then transformed to include different digital files. In modern times, companies are moving away from traditional on-site locations in favor of cloud systems that optimize collaboration. When companies need to share data with outside parties, they implement a digital data center. These systems save space, money and time. They are also important because of the security they provide to backup data immediately.

Companies protect sensitive data through sufficient, around-the-clock service. Because of this, they're trusting their data to systems to enterprises whose main goal is digital data storage. These systems prevent data overload, and they also store data within **infrastructures** that allow companies to quickly access their data.

Digital data storage systems

Data storage systems are servers that host company data in a location different than the company location. **Digital asset management** (DAM) software similarly stores data but there are a few nuanced differences. DAM is suited for media files and documents – it boosts retrieval, sharing and collaboration.

DAM technically stores data, though it's more geared toward companies wishing to store and share digital files with third parties. It also organizes files efficiently. Consider a DAM system when your storage needs include a more broad range of file types, such as media and other documents.

The future of digital storage

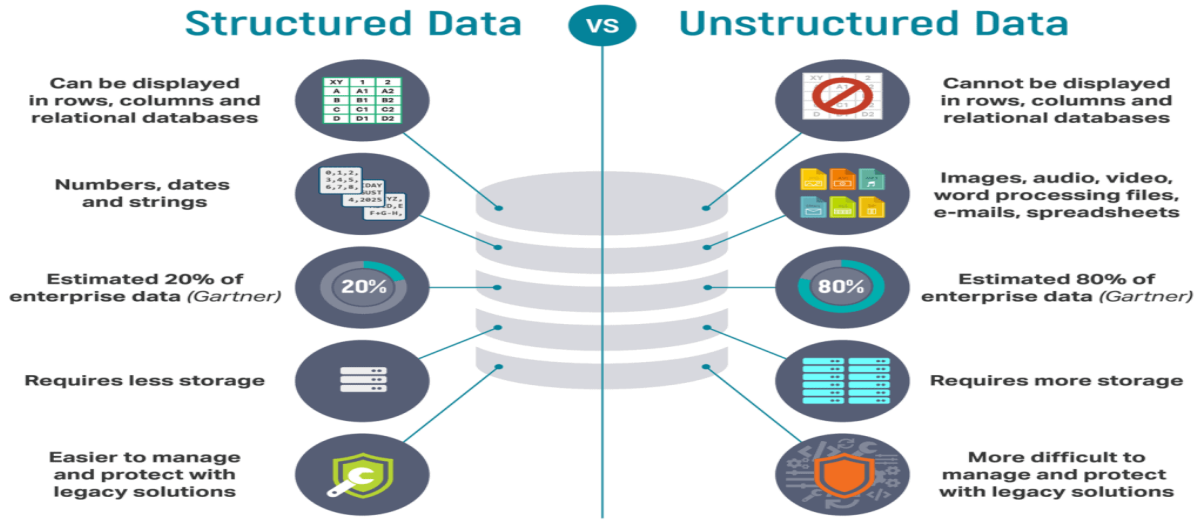
As data continues to grow and evolve, companies in turn need more ways and space to store it. Certainly, nobody can predict the future. However, it's certainly possible to give strong estimates based on what we know so far. If we follow the trajectory that we are currently in, data should continue to increase exponentially. Furthermore, companies will need to keep extensive records for legal reasons.

Will this data continue to be stored on local servers? Will it mostly take to the cloud? There's a lot of uncertainty as to what will win out in the end. There's even talk of things like magnetic tape data storage becoming a large presence. The most important thing you can do is pay attention to what you need specifically to successfully maintain your data.

Digital data storage should not only help free up storage space but also allow your company a chance to quickly retrieve files. Use the right storage system in order to get the most efficient data use.

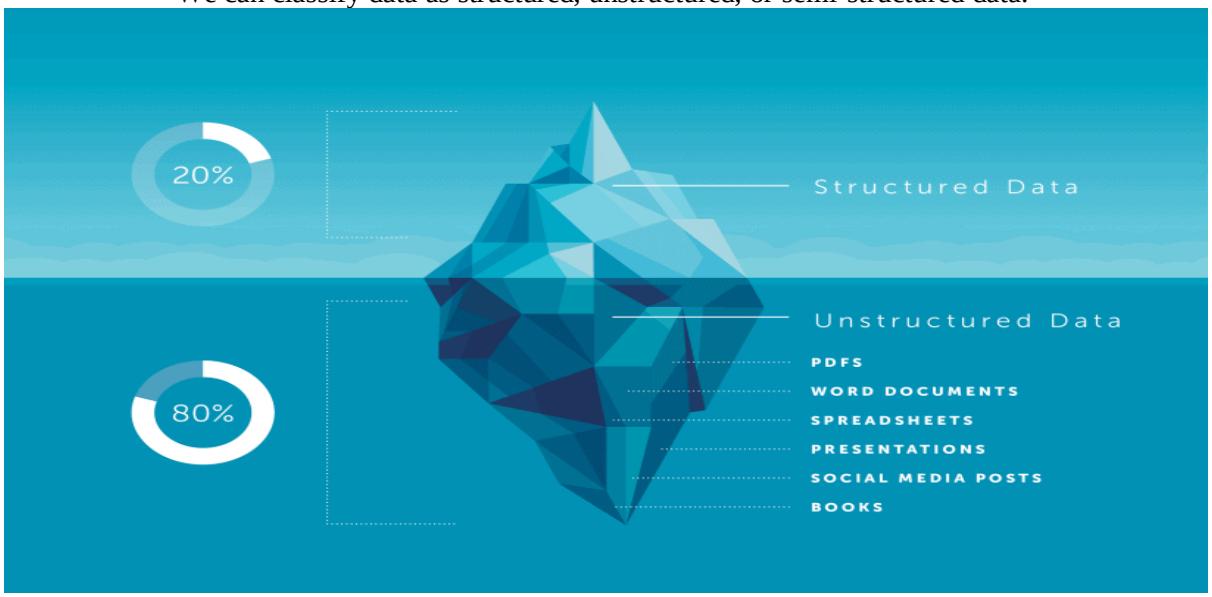
Structured Vs Unstructured Data: Structured data is generally tabular data that is represented by columns and rows in a database and Unstructured data is information that either does not organize in a pre-defined manner or not have a pre-defined data model.

Data, types of Data, and **Structured Vs Unstructured Data,** and suitable **Datastores.**



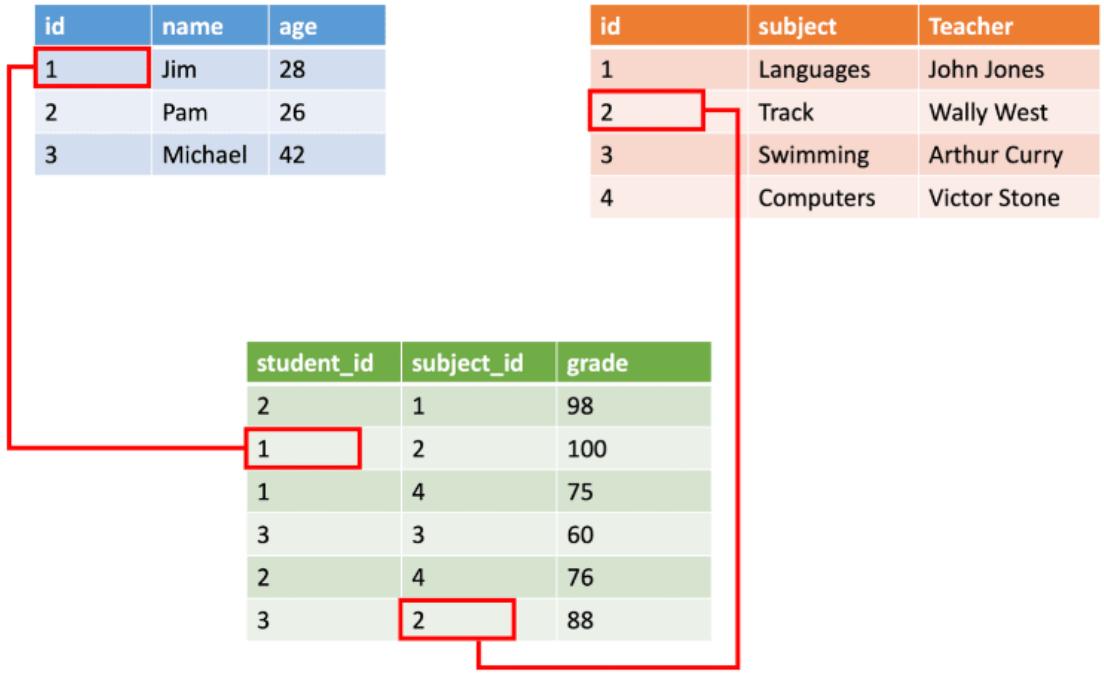
What Is Data?

- Data is a set of facts such as descriptions, observations, and numbers used in decision making.
- We can classify data as structured, unstructured, or semi-structured data.



What is structured data?

- **Structured data** is generally tabular data that is represented by columns and rows in a database.
- Databases that hold tables in this form are called ***relational databases***.
- The mathematical term "*relation"* specify to a formed set of data held as a table.
- In structured data, all row in a table has the same set of columns.
- SQL (Structured Query Language) programming language used for structured data.



2) What is Semi-structured Data

- **Semi-structured** data is information that doesn't consist of Structured data (relational database) but still has some structure to it.
- Semi-structured data consist of documents held in *JavaScript Object Notation* **(JSON) format**. It also includes *key-value* stores and *graph* databases.

```
## Document 1 ##
{
  "customerID": "103248",
  "name":
  {
    "first": "AAA",
    "last": "BBB"
  },
  "address":
  {
    "street": "Main Street",
    "number": "101",
    "city": "Acity",
    "state": "NY"
  },
  "ccOnFile": "yes",
  "firstOrder": "02/28/2003"
}
```

3) What is Unstructured Data

- **Unstructured data** is information that either does not organize in a pre-defined manner or not have a pre-defined data model.

- Unstructured information is a set of text-heavy but may contain data such as numbers, dates, and facts as well.
- **Videos, audio, and binary** data files might not have a specific structure. They're assigned to as **unstructured** data.

Structured data is stored is predefined format and is highly specific; whereas **unstructured data** is a collection of many varied data types which are stored in their native formats; while **semi structured data** that does not follow the tabular data structure models associated with relational databases or other data table forms.

Characteristics Of Structured (Relational) and Unstructured (Non-Relational) Data

Relational Data

- Relational databases provide undoubtedly the most well-understood model for holding data.
- The simplest structure of columns and tables makes them very easy to use initially, but the inflexible structure can cause some problems.
- We can communicate with relational databases using **Structured Query Language (SQL).**
- SQL allows the joining of tables using a few lines of code, with a structure most beginner employees can learn very fast.
- Examples of relational databases:
    - MySQL
    - PostgreSQL
    - Db2

Non-Relational Data

- Non-relational databases permit us to store data in a format that more closely meets the original structure.

- A *non-relational database* is a database that does not use the tabular schema of columns and rows found in most traditional database systems.
- It uses a storage model that is enhanced for the specific requirements of the type of data being stored.
- In a non-relational database the data may be stored as **JSON documents**, as simple **key/value pairs**, or as a graph consisting of edges and vertices.
- Examples of non-relational databases:
    - Redis
    - JanusGraph
    - MongoDB
    - RabbitMQ

What is Data Quality?

Data quality is defined as:
the degree to which data meets a company's expectations of accuracy, validity, completeness, and consistency.
By tracking data quality, a business can pinpoint potential issues harming quality, and ensure that shared data is fit to be used for a given purpose.
When collected data fails to meet the company expectations of accuracy, validity, completeness, and consistency, it can have massive negative impacts on customer service, employee productivity, and key strategies.

Why Is it Important to Have Data Quality?

Quality data is key to making accurate, informed decisions. And while all data has some level of "quality," a variety of characteristics and factors determines the degree of data quality (high-quality versus low-quality). Furthermore, different data quality characteristics will likely be more important to various stakeholders across the organization.
A list of popular data quality characteristics and dimensions include:

- Accuracy
- Completeness
- Consistency

- Integrity
- Reasonability
- Timeliness
- Uniqueness/Deduplication
- Validity
- Accessibility

Because data accuracy is a key attribute of high-quality data, a single inaccurate data point can wreak havoc across the entire system.

Without accuracy and reliability in data quality, executives cannot trust the data or make informed decisions. This can, in turn, increase operational costs and wreak havoc for downstream users. Analysts wind up relying on imperfect reports and making misguided conclusions based on those findings. And the productivity of end-users will diminish due to flawed guidelines and practices being in place.

Poorly maintained data can lead to a variety of other problems, too. For example, out-of-date customer information may result in missed opportunities for up- or cross-selling products and services.

Low-quality data might also cause a company to ship their products to the wrong addresses, resulting in lowered customer satisfaction ratings, decreases in repeat sales, and higher costs due to reshipments.

And in more highly regulated industries, bad data can result in the company receiving fines for improper financial or regulatory compliance reporting.

Top 3 Data Quality Challenges

Data volume presents quality challenges. Whenever large amounts of data are at play, the sheer volume of new information often becomes an essential consideration in determining whether the data is trustworthy. For this reason, forward-thinking companies have robust processes in place for the collection, storage, and processing of data.

As the technological revolution advances at a rapid pace, the **top three data quality challenges** include:

1. Privacy and protection laws

The General Data Protection Regulation (GDPR) and the California Consumer Privacy Act (CCPA), which gives people the right to access their personal data, are substantially increasing public demand for accurate customer records. Organizations must be able to locate the totality of an individual's information almost instantly and without missing even a fraction of the collected data because of inaccurate or inconsistent data.

2. Artificial Intelligence (AI) and Machine Learning (ML)

As more companies implement Artificial Intelligence and Machine Learning applications to their business intelligence strategies, data users may find it increasingly difficult to keep up with new surges of Big Data. Because these real-time data streaming platforms channel vast quantities of new information continuously, there are now even more opportunities for mistakes and data quality inaccuracies.

Furthermore, larger corporations must work diligently to manage their systems, which reside both on-premises and through cloud servers. The abundance of data systems has also made the monitoring of complicated tasks even more challenging.

3. Data governance practices

Data governance is a data management system that adheres to an internal set of standards and policies for the collection, storage, and sharing of information. By ensuring that all data is consistent, trustworthy, and free from misuse within every company department, managers can guarantee compliance with important regulations and reduce the risk of the business being fined.

Without the right data governance approach, the company may never resolve inconsistencies within different systems across the organization. For example, customer names can be listed differently

depending on the department. Sales might say “Sally.” Logistics uses “Sallie.” And customer service lists the name as “Susan.” This poor-quality data governance can result in confusion for customers that have multiple interactions with each department over time.

6 Ways on How to Determine Data’s Quality

The Data Quality Assessment Framework (DQAF) is a set of data quality dimensions, organized into six major categories: *completeness, timeliness, validity, integrity, uniqueness, and consistency.*
These dimensions are useful when evaluating the quality of a particular dataset at any point in time. Most data managers assign a score of 0-100 for each dimension, an average DQAF.

1. Completeness

Completeness is defined as a measure of the percentage of data that is missing within a dataset. For products or services, the completeness of data is crucial in helping potential customers compare, contrast, and choose between different sales items. For instance, if a product description does not include an estimated delivery date (when all the other product descriptions do), then that “data” is incomplete.

2. Timeliness

Timeliness measures how up-to-date or antiquated the data is at any given moment. For example, if you have information on your customers from 2008, and it is now 2021, then there would be an issue with the timeliness as well as the completeness of the data.
When determining data quality, the timeliness dimension can have a tremendous effect — either positive or negative — on its overall accuracy, viability, and reliability.

3. Validity

Validity refers to information that fails to follow specific company formats, rules, or processes. For example, many systems may ask for a customer’s birthdate. However, if the customer does not enter their birthdate using the proper format, the level of data quality becomes automatically compromised. Therefore, many organizations today design their systems to reject birthdate information unless it is input using the pre-assigned format.

4. Integrity

Integrity of data refers to the level at which the information is reliable and trustworthy. Is the data true and factual? For example, if your database has an email address assigned to a specific customer, and it turns out that the customer actually deleted that account years ago, then there would be an issue with data integrity as well as timeliness.

5. Uniqueness

Uniqueness is a data quality characteristic most often associated with customer profiles. A single record can be all that separates your company from winning an e-commerce sale and beating the competition.
Greater accuracy in compiling unique customer information, including each customer’s associated performance analytics related to individual company products and marketing campaigns, is often the cornerstone of long-term profitability and success.

6. Consistency

Consistency of data is most often associated with analytics. It ensures that the source of the information collection is capturing the correct data based on the unique objectives of the department or company.
For example, let’s say you have two similar pieces of information:

1. the date on file for the opening of a customer’s account vs.
2. the last time they logged into their account.

The difference in these dates may provide valuable insights into the success rates of current or future marketing campaigns.
Determining the overall quality of company data is a never-ending process. The most crucial components of effective data quality management are the identification and resolution of potential issues quickly and proactively.

Data Classification

What is Data Classification

Data classification tags data according to its type, sensitivity, and value to the organization if altered, stolen, or destroyed. It helps an organization understand the value of its data, determine whether the data is at risk, and implement controls to mitigate risks. Data classification also helps an organization comply with relevant industry-specific regulatory mandates such as SOX, HIPAA, PCI DSS, and GDPR.

Data Sensitivity Levels

Data is classified according to its sensitivity level—high, medium, or low.

- **High sensitivity data**—if compromised or destroyed in an unauthorized transaction, would have a catastrophic impact on the organization or individuals. For example, financial records, intellectual property, authentication data.
- **Medium sensitivity data**—intended for internal use only, but if compromised or destroyed, would not have a catastrophic impact on the organization or individuals. For example, emails and documents with no confidential data.
- **Low sensitivity data**—intended for public use. For example, public website content.

Types of Data Classification

Data classification can be performed based on content, context, or user selections:

- **Content-based classification**—involves reviewing files and documents, and classifying them
- **Context-based classification**—involves classifying files based on meta data like the application that created the file (for example, accounting software), the person who created the document (for example, finance staff), or the location in which files were authored or modified (for example, finance or legal department buildings).
- **User-based classification**—involves classifying files according to a manual judgement of a knowledgeable user. Individuals who work with documents can specify how sensitive they are—they can do so when they create the document, after a significant edit or review, or before the document is released.

Data States and Data Format

Two additional dimensions of data classifications are:

- **Data states**—data exists in one of three states—at rest, in process, or in transit. Regardless of state, data classified as confidential must remain confidential.
- **Data format**—data can be either structured or unstructured. Structured data are usually human readable and can be indexed. Examples of structured data are database objects and spreadsheets. Unstructured data are usually not human readable or indexable. Examples of unstructured data are source code, documents, and binaries. Classifying structured data is less complex and time-consuming than classifying unstructured data.

Data Discovery

Classifying data requires knowing the location, volume, and context of data. Most modern businesses store large volumes of data, which may be spread across multiple repositories:

- Databases deployed on-premises or in the cloud
- Big data platforms
- Collaboration systems such as Microsoft SharePoint
- Cloud storage services such as Dropbox and Google Docs
- Files such as spreadsheets, PDFs, or emails

Before you can perform data classification, you must perform accurate and comprehensive data discovery. Automated tools can help discover sensitive data at large scale. See our article on Data Discovery for more information.

The Relation Between Data Classification and Compliance

Data classification must comply with relevant regulatory and industry-specific mandates, which may require classification of different data attributes. For example, the Cloud Security Alliance (CSA) requires that data and data objects must include data type, jurisdiction of origin and domicile, context, legal constraints, sensitivity, etc. PCI DSS does not require origin or domicile tags.

Creating Your Data Classification Policy

A data classification policy defines who is responsible for data classification—typically by defining Program Area Designees (PAD) who are responsible for classifying data for different programs or organizational units.

The data classification policy should consider the following questions:

- Which person, organization or program created and/or owns the information?
- Which organizational unit has the most information about the content and context of the information?
- Who is responsible for the integrity and accuracy of the data?
- Where is the information stored?
- Is the information subject to any regulations or compliance standards, and what are the penalties associated with non-compliance?

Data classification can be the responsibility of the information creators, subject matter experts, or those responsible for the correctness of the data.

The policy also determines the data classification process: how often data classification should take place, for which data, which type of data classification is suitable for different types of data, and what technical means should be used to classify data. The data classification policy is part of the overall information security policy, which specifies how to protect sensitive data.

DATA WAREHOUSE

Data Warehouse is a relational database management system (RDBMS) construct to meet the requirement of transaction processing systems. It can be loosely described as any centralized data repository which can be queried for business benefits. It is a database that stores information oriented to satisfy decision-making requests. It is a group of decision support technologies, targets to enabling the knowledge worker (executive, manager, and analyst) to make superior and higher decisions. So, Data Warehousing support architectures and tool for business executives to systematically organize, understand and use their information to make strategic decisions.

Data Warehouse environment contains an extraction, transportation, and loading (ETL) solution, an online analytical processing (OLAP) engine, customer analysis tools, and other applications that handle the process of gathering information and delivering it to business users.

Data warehouse-Definition

A Data Warehouse (DW) is a relational database that is designed for query and analysis rather than transaction processing. It includes historical data derived from transaction data from single and multiple sources. A Data Warehouse provides integrated, enterprise-wide, historical data and focuses on providing support for decision-makers for data modeling and analysis.

A Data Warehouse is a group of data specific to the entire organization, not only to a particular group of users.

It is not used for daily operations and transaction processing but used for making *decisions.*

A Data Warehouse can be viewed as a data system with the following attributes:

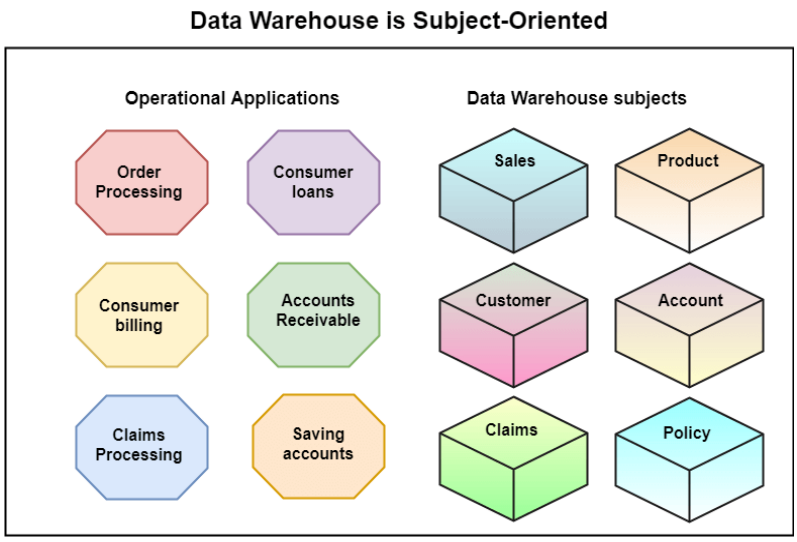
- It is a database designed for investigative tasks, using data from various applications.
- It supports a relatively small number of clients with relatively long interactions.
- It includes current and historical data to provide a historical perspective of information.
- Its usage is read-intensive.
- It contains a few large tables.

"Data Warehouse is a subject-oriented, integrated, and time-variant store of information in support of management's decisions."

Characteristics of Data Warehouse

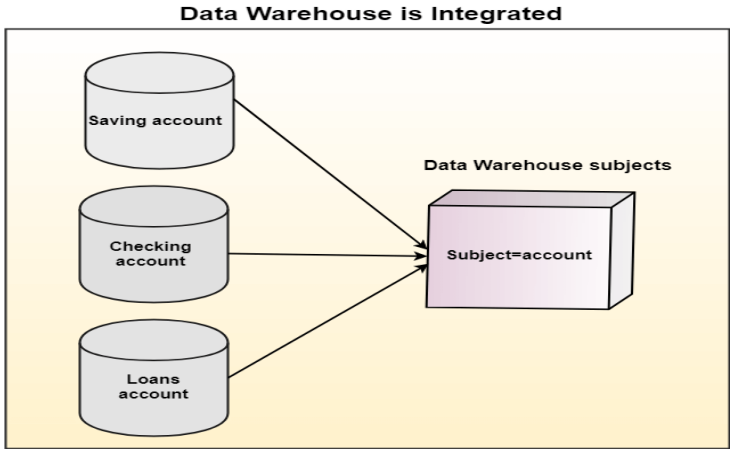
1) ***Subject-Oriented***

A data warehouse target on the modeling and analysis of data for decision-makers. Therefore, data warehouses typically provide a concise and straightforward view around a particular subject, such as customer, product, or sales, instead of the global organization's ongoing operations. This is done by excluding data that are not useful concerning the subject and including all data needed by the users to understand the subject.



2)Integrated

A data warehouse integrates various heterogeneous data sources like RDBMS, flat files, and online transaction records. It requires performing data cleaning and integration during data warehousing to ensure consistency in naming conventions, attributes types, etc., among different data sources.

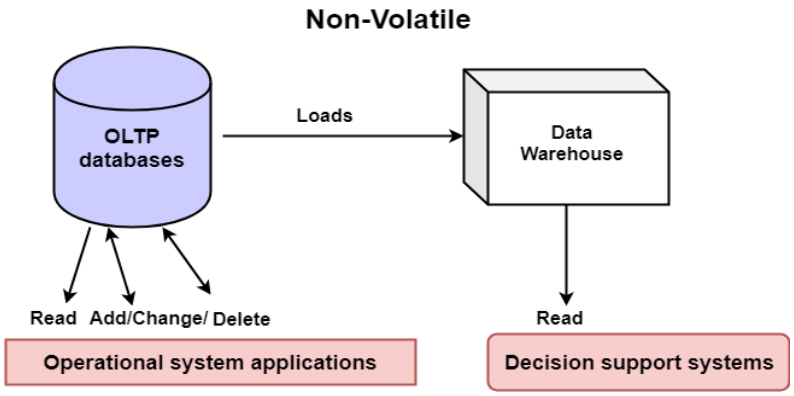


3)Time-Variant

Historical information is kept in a data warehouse. For example, one can retrieve files from 3 months, 6 months, 12 months, or even previous data from a data warehouse. These variations with a transactions system, where often only the most current file is kept.

4)Non-Volatile

The data warehouse is a physically separate data storage, which is transformed from the source operational RDBMS. The operational updates of data do not occur in the data warehouse, i.e., update, insert, and delete operations are not performed. It usually requires only two procedures in data accessing: Initial loading of data and access to data. Therefore, the DW does not require transaction processing, recovery, and concurrency capabilities, which allows for substantial speedup of data retrieval. Non-Volatile defines that once entered into the warehouse, and data should not change.



History of Data Warehouse

The idea of data warehousing came to the late 1980's when IBM researchers Barry Devlin and Paul Murphy established the "Business Data Warehouse."

In essence, the data warehousing idea was planned to support an architectural model for the flow of information from the operational system to decisional support environments. The concept attempt to address the various problems associated with the flow, mainly the high costs associated with it.

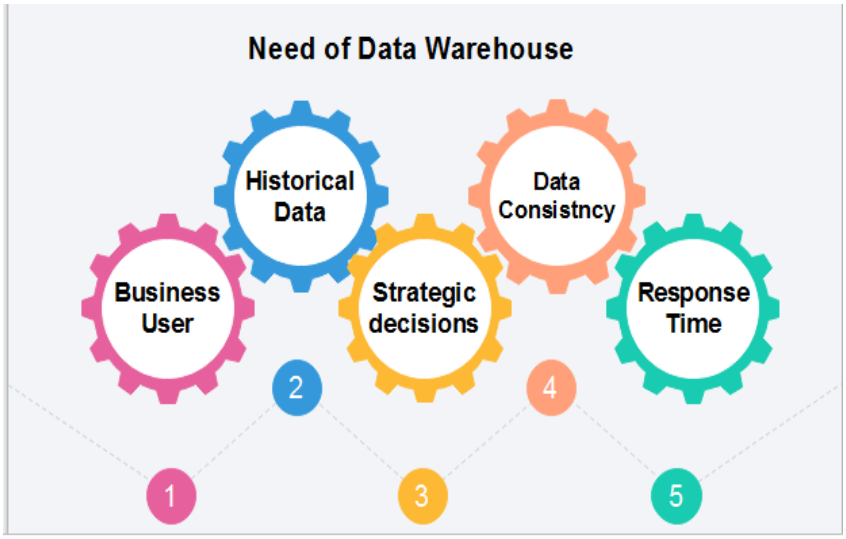
In the absence of data warehousing architecture, a vast amount of space was required to support multiple decision support environments. In large corporations, it was ordinary for various decision support environments to operate independently.

Goals of Data Warehousing

- To help reporting as well as analysis
- Maintain the organization's historical information
- Be the foundation **for decision making.**

Need for Data Warehouse

Data Warehouse is needed for the following reasons:



1. **Business User:** Business users require a data warehouse to view summarized data from the past. Since these people are non-technical, the data may be presented to them in an elementary form.
2. **Store historical data:** Data Warehouse is required to store the time variable data from the past. This input is made to be used for various purposes.
3. **Make strategic decisions:** Some strategies may be depending upon the data in the data warehouse. So, data warehouse contributes to making strategic decisions.
4. **For data consistency and quality:** Bringing the data from different sources at a commonplace, the user can effectively undertake to bring the uniformity and consistency in data.
5. **High response time:** Data warehouse has to be ready for somewhat unexpected loads and types of queries, which demands a significant degree of flexibility and quick response time.

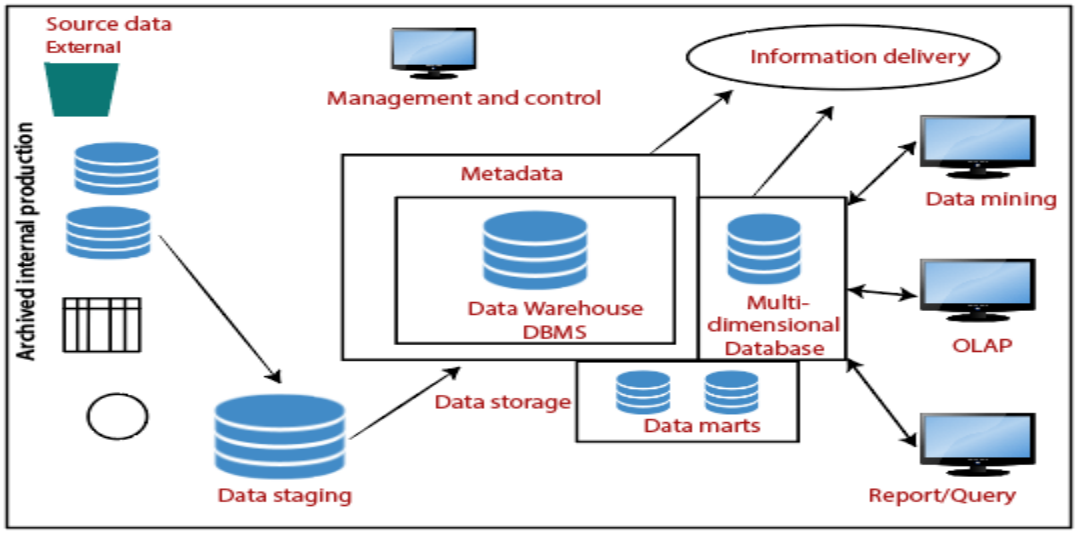
Benefits of Data Warehouse

1. Understand business trends and make better forecasting decisions.
2. Data Warehouses are designed to perform well enormous amounts of data.

3. The structure of data warehouses is more accessible for end-users to navigate, understand, and query.
4. Queries that would be complex in many normalized databases could be easier to build and maintain in data warehouses.
5. Data warehousing is an efficient method to manage demand for lots of information from lots of users.
6. Data warehousing provide the capabilities to analyze a large amount of historical data.

Components or Building Blocks of Data Warehouse

The architecture of Data warehouse contains both software and hardware components.



Components or Building Blocks of Data Warehouse

The figure shows the essential elements of a typical warehouse. Here the Source Data component shows on the left. The Data staging element serves as the next building block. In the middle, it is the Data Storage component that handles the data warehouses data. This element not only stores and manages the data; it also keeps track of data using the metadata repository. The Information Delivery component shows on the right consists of all the different ways of making the information from the data warehouses available to the users. The components of Datawarehouse are :

1)Source Data Component

Source data coming into the data warehouses may be grouped into four broad categories:

Production Data: This type of data comes from the different operating systems of the enterprise. Based on the data requirements in the data warehouse, we choose segments of the data from the various operational modes.

Internal Data: In each organization, the client keeps their "**private**" spreadsheets, reports, customer profiles, and sometimes even department databases. This is the internal data, part of which could be useful in a data warehouse.

Archived Data: Operational systems are mainly intended to run the current business. In every operational system, we periodically take the old data and store it in achieved files.

External Data: Most executives depend on information from external sources for a large percentage of the information they use. They use statistics associating to their industry produced by the external department.

2)Data Staging Component

After the extraction of data from various operational systems and external sources, a file is prepared storing in the data warehouse. The extracted data coming from several different sources need to be changed, converted, and made ready in a format that is relevant to be saved for querying and analysis.

1) **Data Extraction:** This method has to deal with numerous data sources. We have to employ the appropriate techniques for each data source.

2) Data Transformation: As we know, data for a data warehouse comes from many different sources. If data extraction for a data warehouse posture big challenges, data transformation present even significant challenges. We perform several individual tasks as part of data transformation.

First, we clean the data extracted from each source. Cleaning may be the correction of misspellings or may deal with providing default values for missing data elements, or elimination of duplicates when we bring in the same data from various source systems.

Standardization of data components forms a large part of data transformation. Data transformation contains many forms of combining pieces of data from different sources. We combine data from single source record or related data parts from many source records.

On the other hand, data transformation also contains purging source data that is not useful and separating outsource records into new combinations. Sorting and merging of data take place on a large scale in the data staging area. When the data transformation function ends, we have a collection of integrated data that is cleaned, standardized, and summarized.

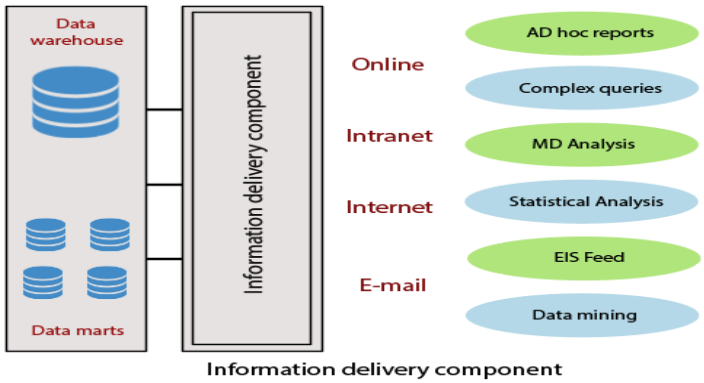
3) Data Loading: Two distinct categories of tasks form data loading functions. When we complete the structure and construction of the data warehouse and go live for the first time, we do the initial loading of the information into the data warehouse storage. The initial load moves high volumes of data using up a substantial amount of time.

4)Data Storage Components

Data storage for the data warehousing is a split repository. The data repositories for the operational systems generally include only the current data. Also, these data repositories include the data structured in highly normalized for fast and efficient processing.

5)Information Delivery Component

The information delivery element is used to enable the process of subscribing for data warehouse files and having it transferred to one or more destinations according to some customer-specified scheduling algorithm.



Information delivery component

6)Metadata Component

Metadata in a data warehouse is equal to the data dictionary or the data catalog in a database management system. In the data dictionary, we keep the data about the logical data structures, the data about the records and addresses, the information about the indexes, and so on.

7)Data Marts

It includes a subset of corporate-wide data that is of value to a specific group of users. The scope is confined to particular selected subjects. Data marts are lower than data warehouses and usually contain organization. The current trends in data warehousing are to developed a data warehouse with several smaller related data marts for particular kinds of queries and reports.

8)Management and Control Component

The management and control elements coordinate the services and functions within the data warehouse. These components control the data transformation and the data transfer into the data warehouse storage. On the other hand, it moderates the data delivery to the clients. Its work with the database management systems and authorizes data to be correctly saved in the repositories. It monitors the movement of information into the staging method and from there into the data warehouses storage itself.

Need of a Data warehouse

- Data Warehouse queries are complex because they involve the computation of large groups of data at summarized levels.
- It may require the use of distinctive data organization, access, and implementation method based on multidimensional views.
- Performing OLAP queries in operational database degrade the performance of functional tasks.
- Data Warehouse is used for analysis and decision making in which extensive database is required, including historical data, which operational database does not typically maintain.

- The separation of an operational database from data warehouses is based on the different structures and uses of data in these systems.
- Because the two systems provide different functionalities and require different kinds of data, it is necessary to maintain separate databases.

Difference between Database and Data Warehouse

Database	Data Warehouse
1. It is used for Online Transactional Processing (OLTP) but can be used for other objectives such as Data Warehousing. This records the data from the clients for history.	1. It is used for Online Analytical Processing (OLAP). This reads the historical information for the customers for business decisions.
2. The tables and joins are complicated since they are normalized for RDBMS. This is done to reduce redundant files and to save storage space.	2. The tables and joins are accessible since they are de-normalized. This is done to minimize the response time for analytical queries.
3. Data is dynamic	3. Data is largely static
4. **Entity:** Relational modeling procedures are used for RDBMS database design.	4. **Data:** Modeling approach are used for the Data Warehouse design.
5. Optimized for write operations.	5. Optimized for read operations.
6. Performance is low for analysis queries.	6. High performance for analytical queries.
7. The database is the place where the data is taken as a base and managed to get available fast and efficient access.	7. Data Warehouse is the place where the application data is handled for analysis and reporting objectives.

DATA LAKE

When it comes to storing big data, the two most popular options are data lakes and data warehouses. Data warehouses are used for analyzing archived structured data, while data lakes are used to store big data of all structures.

	Data Lake	Data Warehouse
Type of data	Unstructured and structured data from various company data sources	Historical data that has been structured to fit a relational

	Data Lake	Data Warehouse
		database schema
Purpose	Cost-effective big data storage	Analytics for business decisions
Users	Data scientists and engineers	Data analysts and business analysts
Tasks	Storing data and big data analytics, like deep learning and real-time analytics	Typically read-only queries for aggregating and summarizing data
Size	Stores all data that might be used—can take up petabytes!	Only stores data relevant to analysis

Type of data

Cleaning data is a key data skill because data naturally comes in messy and imperfect forms. Raw data that hasn't been cleaned is called *unstructured data*—which comprises most of the data in the world, like photos, chat logs, and PDF files. Unstructured data that has been cleaned to fit a schema, organized into tables and defined by data types and relationships, is called *structured data*. This is the fundamental difference between lakes and warehouses.

Data lakes store data from a wide variety of sources like IoT devices, real-time social media streams, user data, and web application transactions. Data warehouses, on the other hand, contain historical data that have been cleaned to fit a relational schema.

Data lakes are much bigger in size because they retain all data that might be relevant to a company. Data lakes are often petabytes in size—that's 1,000 terabytes! Data warehouses are much more selective on what data is stored.

Purpose of Data Lake

Data lakes are used for cost-effective storage of large amounts of data from many sources. Allowing data of any structure reduces cost because data is more flexible and scalable as the data doesn't need to fit a specific schema. However, structured data is easier to analyze because it's cleaner and has a uniform schema to query from. By restricting data to a schema, data warehouses are very efficient for analyzing historical data for specific data decisions.

Users of Data Lake

Data lakes and data warehouses are useful for different users. Data analysts and business analysts often work within data warehouses containing explicitly pertinent data that has been processed for their work. Data warehouses require a lower level of programming and data science knowledge to use.Data lakes are set up and maintained by data engineers who integrate them into data pipelines. Data scientists work more closely with data lakes as they contain data of a wider and more current scope.

Tasks performed using Data Lakes

Data engineers use data lakes to store incoming data. However, data lakes aren't only limited to storage. Remember, unstructured data is more flexible and scalable, which is oftentimes better for big data analytics. Big data analytics can be run on data lakes using services such as Apache Spark and Hadoop. This is especially true for deep learning, which requires scalability in the increasing amount of training data.Data warehouses are typically set to read-only for analyst users, who are primarily reading and aggregating data for insights. Since data is already clean and archival, there is usually no need to insert or update data.

BUSINESS REPORTING

Business reporting is used to analyze how the business is tracking all aspects of the organization. Reports through insight through which business owners and senior managers can make informed decisions based on relevant data.

Aim of Business Reporting

Through business reporting, one can critically analyze how the business is tracking in all aspects of the organization. Thorough reports provide for a deeper insight through which business owners and senior managers can make informed decisions based on relevant data. A clear overview of the current situation of the organization provides a guideline to enhance decision-making and facilitate the thinking process.

The reporting process involves:

- Compiling relevant data
- Reviewing information within specific areas (such as sales, finance, operations, etc.)
- Analysis
- Drawing conclusions
- Recommendations

The ultimate result of the report may provide an explanation for the occurrence of an issue, identify performance glitches, and recommending a pertinent course of action to address the issues in question.

Benefits of Business reporting

- Adhere to regulation and competence
- Transparency
- Facilitates decision making process and improves efficiency
- Troubleshooting
- Cost efficiency
- Customer focus
- Business reporting for business success

Adherence to Regulation and Compliance

With the right implementation set-up to cater to the organization's particular requirements, you can have your mind at rest that no salient criteria are missing through your most important reports. Human error is natural. Business reporting tools can lessen unnecessary concerns.

Transparency

Business reporting facilitates communication as members of your organization can be on the same page. Having the possibility to provide a clear picture for everyone involved can gear members of the organization to align with the organization's goals.

Facilitates decision-making process and improves efficiency

In the lack of a functional and practical business reporting system, managers and business owners tend to rely on their experience or intuition which does not necessarily quantify in a clear sense of direction to the members of the organization. Sustaining experience with the right data can help managers have all things considered to make an informed decision.

Troubleshooting

Problems are inevitable but the way they've dealt with is under our control. Identifying problems and in time to act can sufficiently decrease the likelihood of escalating problems to higher levels. Regular monitoring through reporting enables the organization to trace and address hiccups in a timely manner before it's too late.

Cost efficiency

A key characteristic for survival is the ability to adapt. In the business environment, reviewing and assessing strategies are crucial activities to amend as necessary and switch gears where needed. This is not possible without having the right data to allow managers to respond to a continuously changing environment. having the data at one's fingertips allows for proactivity which is essential for cost efficiency.

Customer focus

At the end of the day, business reporting efforts are there to improve the organization on every level and hence provide for a better customer experience. with more data about the needs of the customers, management is equipped to improve customer service. Automated reports can facilitate the generation of quarterly and annual reports for your client.

Business reporting for business success

Business reporting guides strategic decision-making to help business leaders plan ahead and work out budgets among other functions. Report data is used to sustain and back decisions while also identifying already proven successes to strengthen their strengths while addressing weaknesses. monitoring and reporting go beyond highlighting problems but more importantly, they can identify opportunities for growth and development.

VISUAL ANALYTICS

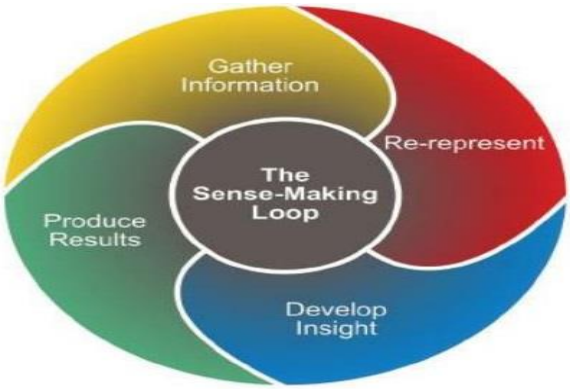
Data visualization and visual analytics are two terms that come up a lot when new and experienced analytics users alike delve into the world of data in their quest to make smarter decisions.

Data Visualization means showing data in a visual format that makes insights easier to understand for human users. Data is usually visualized in a pictorial or graphical form such as charts, graphs, lists, maps, and comprehensive dashboards that combine these multiple formats.

The *primary objective of data visualization* is *to clearly communicate what the data says, help explain trends and statistics, and show patterns that would otherwise be impossible to see*. Data visualization is used to make the consuming, interpreting, and understanding data as simple as possible, and to make it easier to derive insights from data. When BI and analytics users want to see analytics results, and learn from them quickly, they rely on data visualizations.

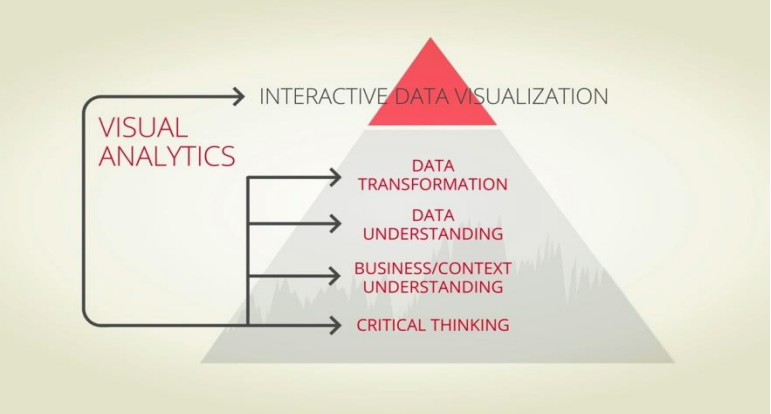
Visual analytics does the "heavy lifting" with data, by using a variety of processes — mechanical, algorithms, machine learning, natural language processing, etc — to identify and reveal patterns and trends. It prepares the data for the process of data visualization, thereby enabling users to examine data, understand what it means, interpret the patterns it highlights, and help them find meaning and gain useful insights from complex data sets.

The ***relationship between Data visualization and Visual analytics*** is symbiotic. Cook illustrates this relationship in what they describe as "The Sense-Making Loop" of the analytical reasoning process:



Role of Visualization in Analytics

Data visualization can either be **static** or **interactive**. **Static visualizations** provide users with a single view of what's in front of them. **Interactive visualizations** enable users to drill down into data and extract and examine various views of the same dataset, selecting specific data points that they want to see in a visualized format.



In this diagram, visual analytics is shown to be the foundation for interactive data, thereby demonstrating how the two are connected. Analytics acts as the source for data visualization and contributes to the health of any organization by identifying underlying models and patterns and predicting needs.

Types of Analytics (Visualization- Past, Present , Future)

Broadly, there are three types of analytics: **descriptive**, **prescriptive**, and **predictive**. The simplest type is ***Descriptive analytics***, describes something that has already happened and suggests its root causes.

Prescriptive analytics takes things a stage further: In addition to helping organizations understand causes, it helps them learn from what's happened and shape tactics and strategies that can improve their current performance and their profitability. A simple example would be the analysis of marketing campaigns.

Predictive analytics is the most beneficial, but arguably the most complex type. It helps users to identify patterns that suggest future situations and behaviors. Using predictive analytics, organizations can plan for forthcoming scenarios, anticipate new trends, and prepare for them most efficiently and cost-effectively. Predicting forthcoming trends sets the stage for optimizing the benefits your organization takes from them.

Visualization to make Smarter Decisions

The data drawn from power visualizations comes from a variety of sources: **Structured data**, in the form of relational databases such as Excel, or unstructured data, deriving from text, video, audio, photos, the

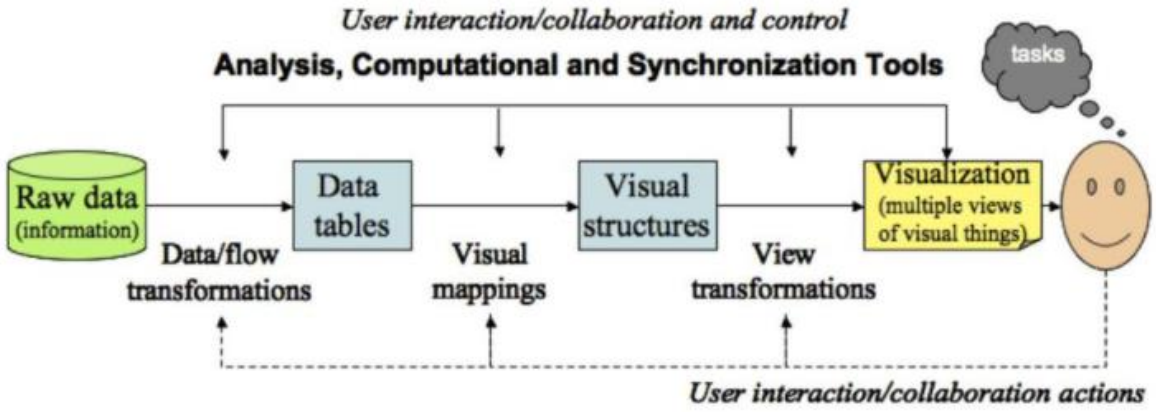
internet and smart devices. This data is gathered into either on-premises servers or increasingly into cloud data warehouses and data lakes. They are transformed into data visualizations and shared via dashboards and analytic apps so that users can make smarter, data-driven decisions.

Data teams and business and analytics teams are tasked with choosing and developing the best way to visualize data and to build well-organized dashboards in order to help end-users make smarter decisions. Dashboards need to be clear, quick to interpret, and easy to drill into to find the deeper insights when needed.

To achieve this successfully, we need *a data and analytics platform* that offers a powerful combination of visual analytics and data visualization; with the capacity to handle huge volumes of data either stored on-premises, in the Cloud, or both; with the flexibility to integrate data from any source; and with the scalability for future growth.

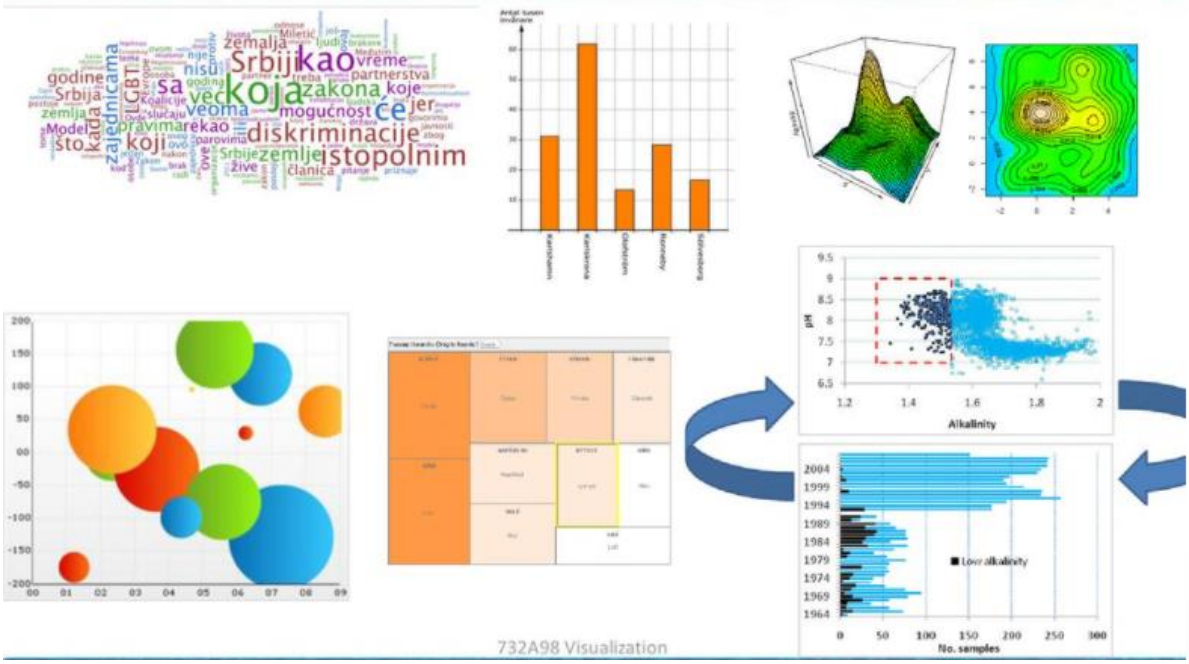
DIFFERENT TYPES OF CHARTS AND GRAPHS

- Data visualization is the process of translating large data sets and metrics into charts, graphs and other visuals.
- The resulting visual representation of data makes it easier to identify and share real-time trends, outliers, and new insights about the information represented in the data.



Key ingredient: mapping data columns to visual structures (aesthetics)

Some Examples of Data Visualization



Common general types of data visualization:

- Charts
- Tables
- Graphs
- Maps
- Infographics
- Dashboards
- **More specific examples of methods to visualize data:**
- Area Chart
- Bar Chart
- Box-and-whisker Plots
- Bubble Cloud
- BulletGraph
- Cartogram
- Circle View
- Dot Distribution Map
- Gantt Chart

- Heat Map
- Highlight Table
- Histogram
- Matrix
- Network
- Polar Area
- Radial Tree
- Scatter Plot (2D or 3D)
- Streamgraph
- Text Tables
- Timeline
- Treemap
- Wedge Stack Graph
- Word Cloud
- And any mix-and-match combination in a dashboard

Different types of graphs and charts can help you:

- Motivate your team to take action
- Impress stakeholders with goal progress
- Show your audience what you value as a business

Data visualization builds trust and can organize diverse teams around new initiatives.

How to choose right chart or graph for your data?

1. Identify your goals for presenting the data.
2. Figure out what data you need to achieve your goal.
3. Gather your data.
4. Select the right type of graph or chart.

#To compare values

Charts and graphs are perfect for comparing one or many value sets, and they can easily show the low and high values in the data sets. To create a comparison chart, use these types of graphs:

- **Column**
- **Mekko**
- **Bar**
- **Pie**
- **Line**
- **Scatter Plot**

#To show the composition of something

Use this type of chart to show how individual parts make up the whole of something, like the device type used for mobile visitors to your website or total sales broken down by sales rep.

To show composition, use these charts:

- **Pie**
- **Stacked Bar**
- **Mekko**
- **Area**
- **Waterfall**

#To understand the distribution of your data

Distribution charts help you to understand outliers, the normal tendency, and the range of information in your values.

Use these charts to show distribution:

- **Scatter Plot**
- **Mekko**
- **Line**
- **Column**
- **Bar**

#Analyzing trends in your data set

If you want to know more information about how a data set performed during a specific time period, there are specific chart types that do extremely well.

- **Line**
- **Dual-Axis Line**
- **Column**

#Better understand the relationship between value sets

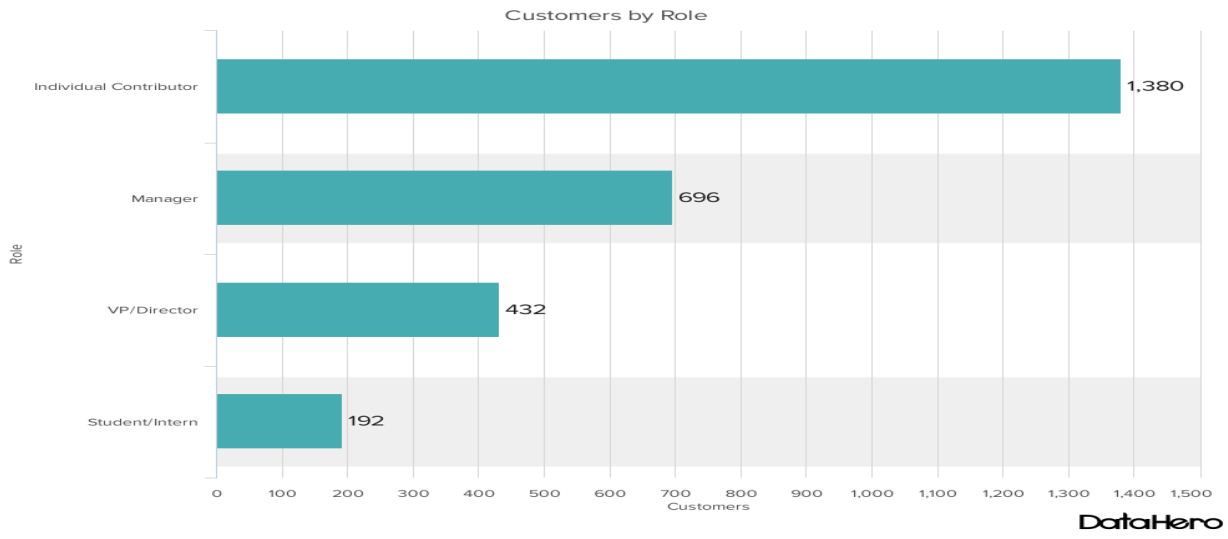
Relationship charts can show how one variable relates to one or many different variables. You could use this to show how something positively affects, has no effect, or negatively affects another variable.

When trying to establish the relationship between things, use these charts:

- **Scatter Plot**
- **Bubble**
- **Line**

1.Bar Graph

A bar graph should be used to avoid clutter when one data label is long or if you have more than 10 items to compare.



Best Use Cases for These Types of Graphs:

Bar graphs can help you compare data between different groups or to track changes over time. Bar graphs are most useful when there are big changes or to show how one group compares against other groups.

The example above compares the number of customers by business role. It makes it easy to see that there is more than twice the number of customers per role for individual contributors than any other group.

A bar graph also makes it easy to see which group of data is highest or most common.

Other use cases for bar graphs include:

- Product comparisons
- Product usage
- Category comparisons
- Marketing traffic by month or year
- Marketing conversions

Design Best Practices for Bar Graphs:

- **Use consistent colors** throughout the chart, selecting accent colors to highlight meaningful data points or changes over time.
- **Use horizontal labels** to improve readability.
- **Start the y-axis at 0** to appropriately reflect the values in your graph.

2. Column Chart

Use a column chart to show a comparison among different items, or to show a comparison of items over time. You could use this format to see the revenue per landing page or customers by close date.

Best Use Cases for This Type of Chart:

While column charts show information vertically, and bar graphs show data horizontally. While you can use both to display changes in data, column charts are best for negative data.

For example, warehouses often track the number of accidents that happen on the shop floor. When the number of incidents falls below the monthly average, a column chart can make that change easier to see in a presentation.

In the example above, this column chart measures the number of customers by close date. Column charts make it easy to see data changes over a period of time. This means that they have many use cases, including:

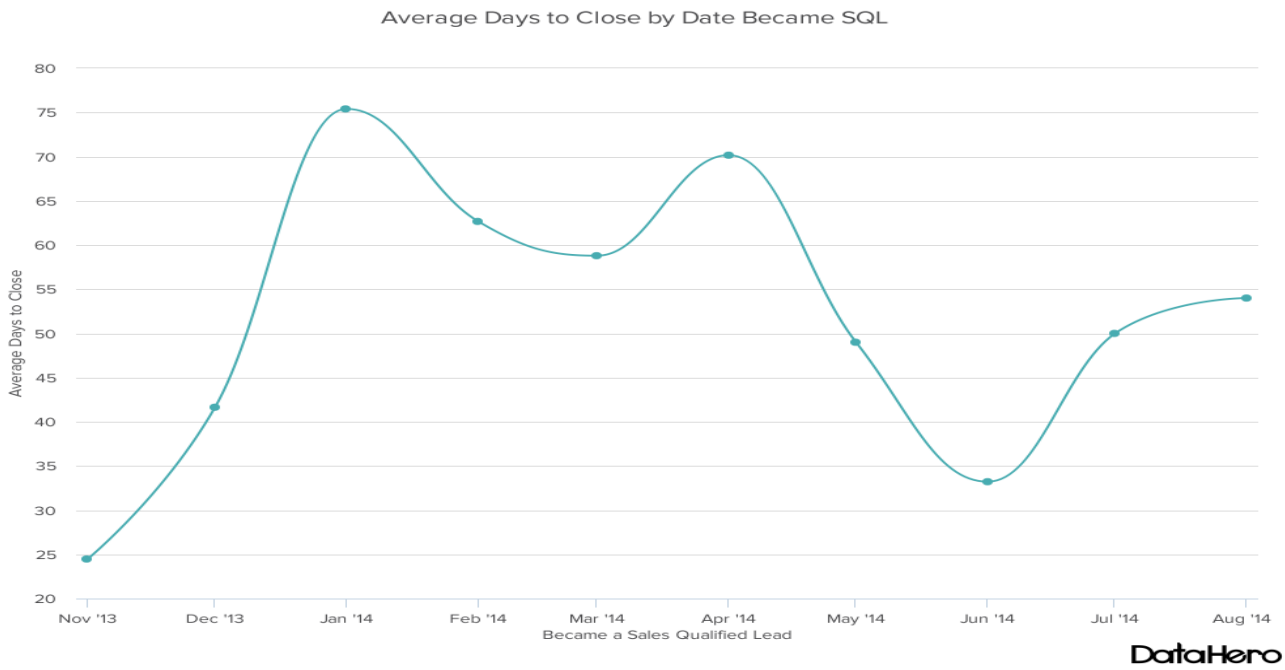
- Customer survey data, like showing how many customers prefer a specific product or how much a customer uses a product each day.
- Sales volume, like showing which services are the top sellers each month or the number of sales per week.
- Profit and loss, showing where business investments are growing or falling.

Design Best Practices for Column Charts:

- **Use consistent colors** throughout the chart, selecting accent colors to highlight meaningful data points or changes over time.
- **Use horizontal labels** to improve readability.
- **Start the y-axis at 0** to appropriately reflect the values in your graph.

3. Line Graph

A line graph reveals trends or progress over time and you can use it to show many different categories of data. You should use it when you chart a continuous data set.



Best Use Cases for These Types of Graphs:

Line graphs help users track changes over short and long periods of time. Because of this, these types of graphs are good for seeing small changes.

Line graphs can help you compare changes for more than one group over the same period. They're also helpful for measuring how different groups relate to each other.

A business might use this type of graph to compare sales rates for different products or services over time.

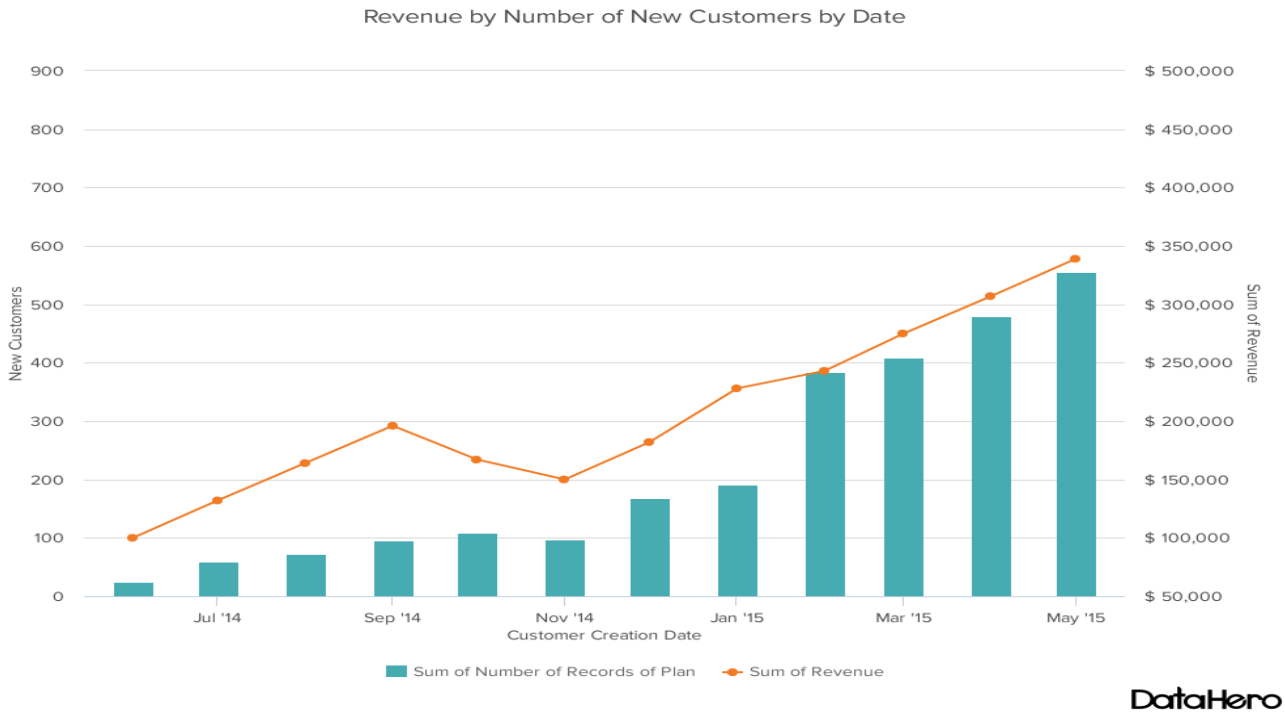
These charts are also helpful for measuring service channel performance. For example, a line graph that tracks how many chats or emails your team responds to per month.

Design Best Practices for Line Graphs:

- **Use solid lines only.**
- Don't plot more than four lines **to avoid visual distractions.**
- Use the right height so the lines take up roughly 2/3 of the y-axis' height.

4. Dual Axis Chart

A dual-axis chart allows you to plot data using two y-axes and a shared x-axis. It has three data sets. One is a continuous set of data and the other is better suited to grouping by category. Use this chart to visualize a correlation or the lack thereof between these three data sets.



Best Use Cases for This Type of Chart:

A dual-axis chart makes it easy to see relationships between different data sets. They can also help with comparing trends.

For example, the chart above shows how many new customers this company brings in each month. It also shows how much revenue those customers are bringing the company.

This makes it simple to see the connection between the number of customers and increased revenue.

You can use dual-axis charts to compare:

- Price and volume of your products
- Revenue and units sold
- Sales and profit margin
- Individual sales performance

Design Best Practices for Dual Axis Charts:

- **Use the y-axis on the left side for the primary variable** because brains are naturally inclined to look left first.
- **Use different graphing styles** to illustrate the two data sets, as illustrated above.
- **Choose contrasting colors** for the two data sets.

5. Area Chart

An area chart is basically a line chart, but the space between the x-axis and the line is filled with a color or pattern. It is useful for showing part-to-whole relations, like showing individual sales reps' contributions to total sales for a year. It helps you analyze both overall and individual trend information.

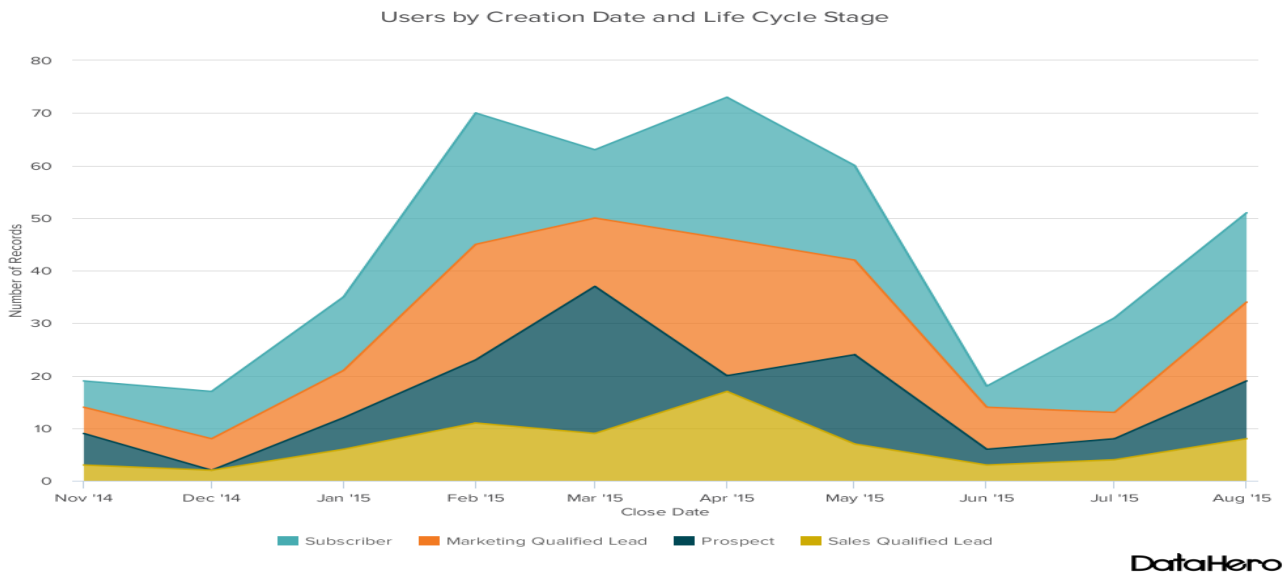
Best Use Cases for These Types of Charts:

Area charts help show changes over time. They work best for big differences between data sets and also help visualize big trends.

For example, the chart above shows users by creation date and life cycle stage.

A line chart could show that there are more subscribers than marketing qualified leads. But this area chart emphasizes how much bigger the number of subscribers is than any other group.

These types of charts and graphs make the size of a group and how groups relate to each other more visually important than data changes over time



.Area graphs can help your business to:

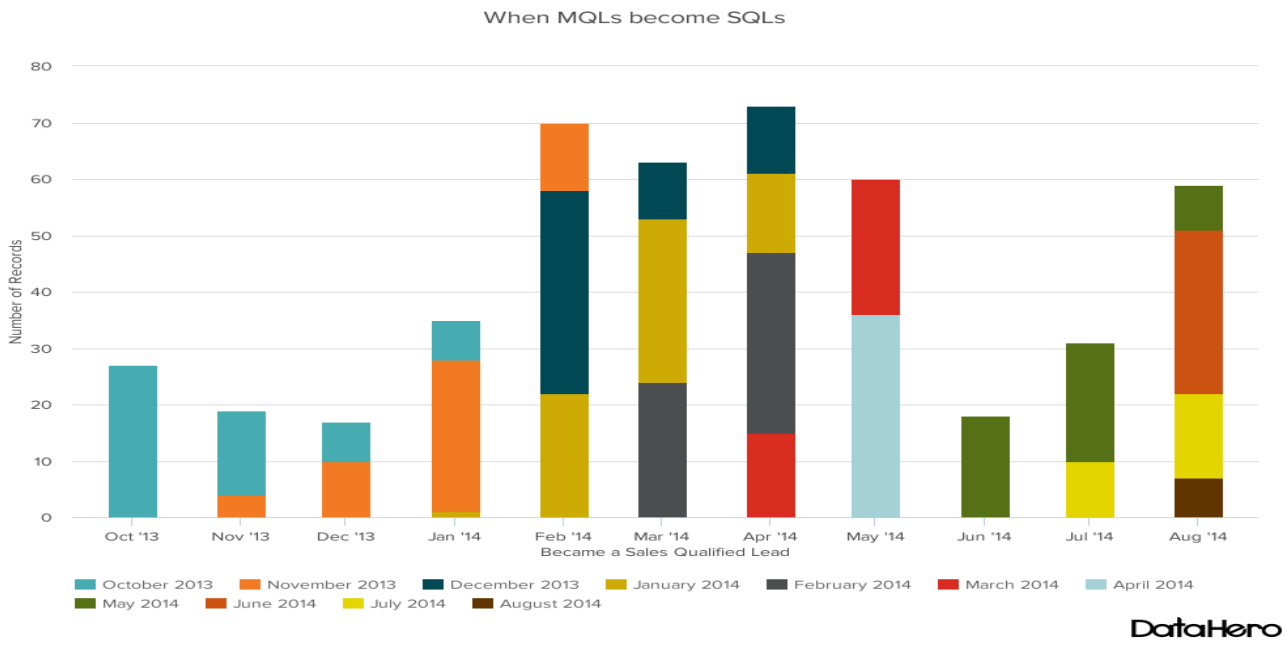
- Visualize which product categories or products within a category are most popular
- Show key performance indicator (KPI) goals vs. outcomes
- Spot and analyze industry trends

Design Best Practices for Area Charts:

- **Use transparent colors** so information isn't obscured in the background.
- **Don't display more than four categories** to avoid clutter.
- Organize highly variable data at the top of the chart to **make it easy to read.**

6. Stacked Bar Chart

This chart is used to compare many different items and show the composition of each item compared.



Best Use Cases for These Types of Graphs:

These graphs are helpful when a group starts in one column and moves to another over time.

For example, the difference between a **marketing qualified lead (MQL) and a sales qualified lead (SQL)** is sometimes hard to see. The chart above helps stakeholders see these two lead types from a single point of view– when a lead changes from MQL to SQL.

Stacked bar charts are excellent for marketing. They make it simple to add a lot of data on a single chart or to make a point with limited space.These types of graphs can show multiple takeaways, so they're also super for quarterly meetings when you have a lot to say, but not always a lot of time to say it.

Stacked bar charts are also a smart option for planning or strategy meetings. This is because these charts can show a lot of information at once, but they also make it easy to focus on one stack at a time or move data as needed.

- Show the frequency of survey responses
- Identify outliers in historical data
- Compare a part of a strategy to its performance as a whole

Design Best Practices for Stacked Bar Graphs:

- Best used to illustrate **part-to-whole relationships.**

- Use **contrasting colors** for greater clarity.
- **Make the chart scale large enough** to view group sizes in relation to one another.

7. Mekko Chart

Also known as a Mari mekko chart, this type of graph can compare values, measure each one's composition, and show data distribution across each one.

It's similar to a stacked bar, except the Mekko's x-axis can capture another dimension of your values—instead of time progression, like column charts often do. In the graphic below, the x-axis compares each city to one another.

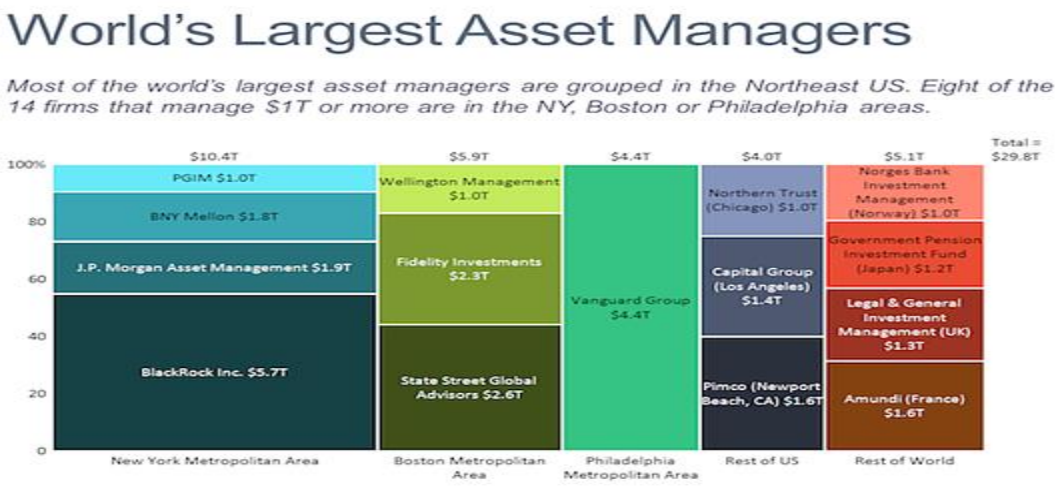


Image Source

Best Use Cases for This Type of Chart:

You can use a Mekko chart to show growth, market share, or competitor analysis.

For example, the Mekko chart above shows the market share of asset managers grouped by location and the value of their assets. This chart makes it clear which firms manage the most assets in different areas.

It's also easy to see which asset managers are largest and how they relate to each other.

Mekko charts can seem more complex than other types of charts and graphs. So, it's best to use these in situations where you want to emphasize scale or differences between groups of data.

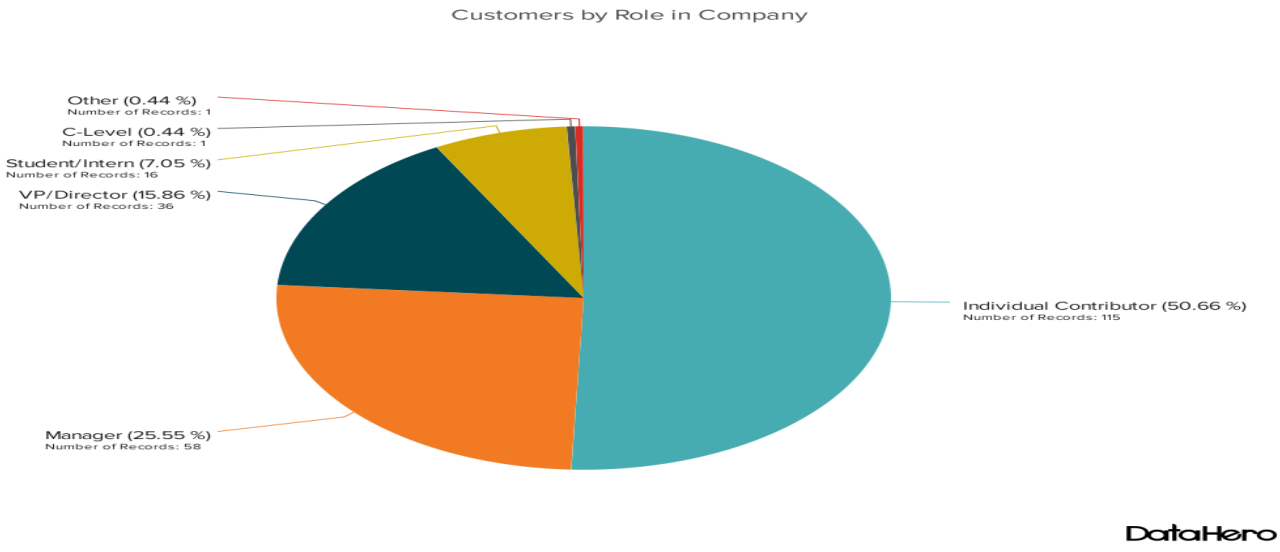
- Detailed profit and loss statements
- Revenue by brand and region
- Product profitability
- Share of voice by industry or niche

Design Best Practices for Mekko Charts:

- **Vary your bar heights** if the portion size is an important point of comparison.
- **Don't include too many composite values** within each bar. You might want to reevaluate your presentation if you have a lot of data.
- Order your bars from **left to right** in such a way that **exposes a relevant trend or message**.

8. Pie Chart

A pie chart shows a static number and how categories represent part of a whole — the composition of something. A pie chart represents numbers in percentages, and the total sum of all segments needs to equal 100%.



Best Use Cases for This Type of Chart:

The image above shows another example of customers by role in the company.

The bar graph example shows you that there are more individual contributors than any other role. But this pie chart makes it clear that they make up over 50% of customer roles.

Pie charts make it easy to see a section in relation to the whole, so they are good for showing:

- Customer personas in relation to all customers
- Revenue from your most popular products or product types in relation to all product sales
- Percent of total profit from different store locations

Design Best Practices for Pie Charts:

- **Don't illustrate too many categories** to ensure differentiation between slices.
- Ensure that the slice values **add up to 100%.**
- **Order slices** according to their size.

9. Scatter Plot Chart

A scatter plot or scattergram chart will show the relationship between two different variables or reveals distribution trends. Use this chart when there are many different data points, and you want to highlight similarities in the data set. This is useful when looking for outliers or for understanding the distribution of your data.

Best Use Cases for These Types of Charts:

Scatter plots are helpful in situations where you have too much data to quickly see a pattern. They are best when you use them to show relationships between two large data sets.

In the example above, this chart shows how customer happiness relates to the time it takes for them to get a response.

- Employment and manufacturing output

- Retail sales and inflation
- Visitor numbers and outdoor temperature
- Sales growth and tax laws

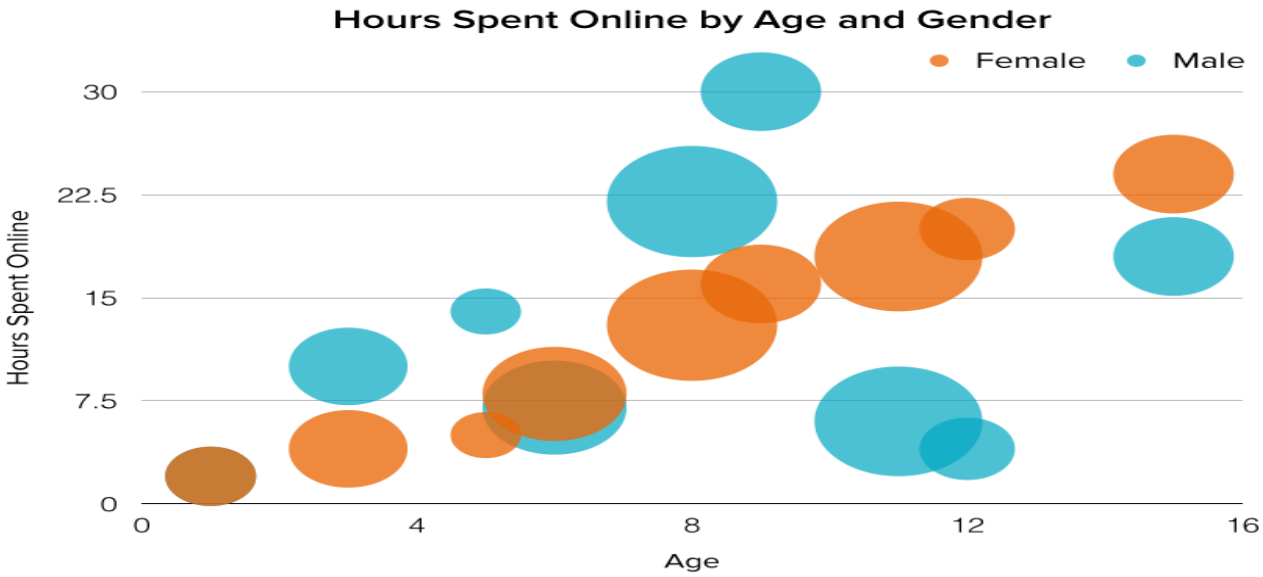
Try to choose two data sets that already have a positive or negative relationship. That said, this type of graph can also make it easier to see data that falls outside of normal patterns.

Design Best Practices for Scatter Plots:

- **Include more variables**, like different sizes, to incorporate more data.
- **Start the y-axis at 0** to represent data accurately.
- If you use **trend lines**, only use a maximum of two to make your plot easy to understand.

10. Bubble Chart

A bubble chart is similar to a scatter plot in that it can show distribution or relationship. There is a third data set shown by the size of the bubble or circle.



Best Use Cases for This Type of Chart:

In the example above, the number of hours spent online isn't just compared to the age of the user, as it would be on a scatter plot chart.

Instead, you can also see how the gender of the user impacts time spent online.

This makes bubble charts useful for seeing the rise or fall of trends over time. It also lets you add another option when you're trying to understand relationships between different segments or categories.

For example, if you want to launch a new product, this chart could help you quickly see the cost, risk, and value of your new product. This can help you focus your energies on a new product that is low risk with a high potential return.

You can also use bubble charts for:

- Top sales by month and location
- Customer satisfaction surveys
- Store performance tracking
- Marketing campaign reviews

Design Best Practices for Bubble Charts:

- **Scale bubbles according to area**, not diameter.
- Make sure **labels are clear and visible.**
- **Use circular shapes** only.

11. Waterfall Chart

Use a waterfall chart to show how an initial value changes with intermediate values — either positive or negative — and results in a final value.

Use this chart to reveal the composition of a number. An example of this would be to showcase how overall company revenue is influenced by different departments and leads to a specific profit number.

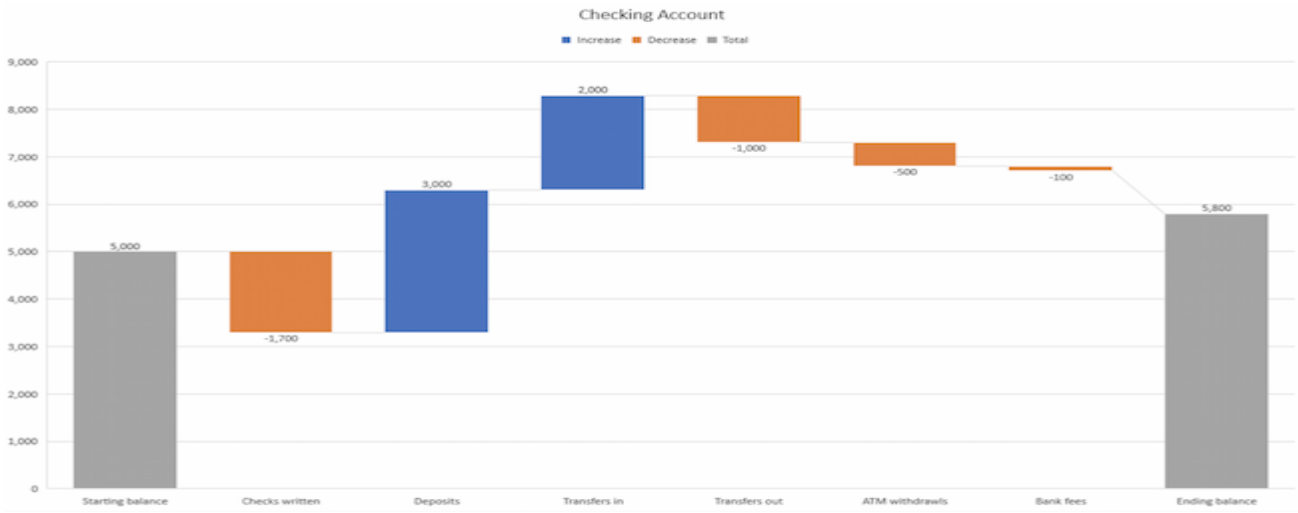


Image Source

Best Use Cases for This Type of Chart:

These types of charts and graphs make it easier to understand how internal and external factors impact a product or campaign as a whole.

In the example above the chart moves from the starting balance on the far left to the ending balance on the far right. Factors in the center include deposits, transfers in and out, and bank fees.

A waterfall chart offers a quick visual that makes complex processes and outcomes easier to see and troubleshoot. For example, SaaS companies often measure customer churn. This format can help visualize changes in new, current, and free trial users, or changes by user segment.

You may also want to try a waterfall chart to show:

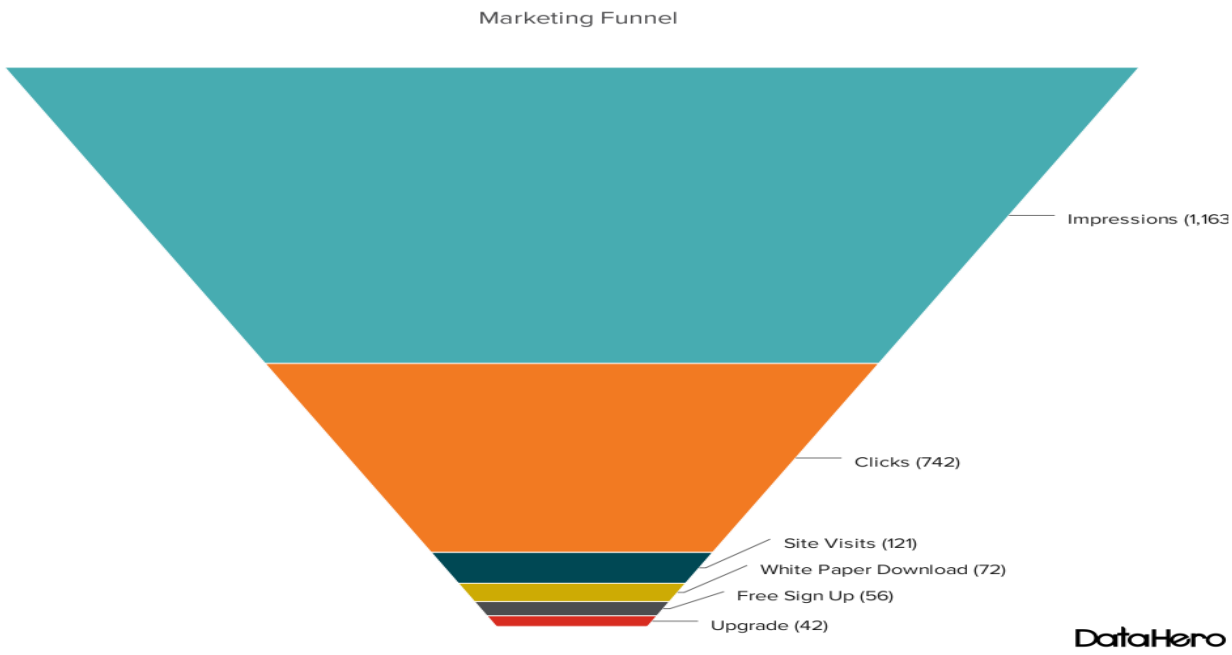
- Changes in revenue or profit over time
- Inventory audits
- Employee staffing reviews

Design Best Practices for Waterfall Charts:

- **Use contrasting colors** to highlight differences in data sets.
- Choose warm colors to indicate increases and cool colors to indicate decreases.

12. Funnel Chart

A funnel chart shows a series of steps and the completion rate for each step. Use this type of chart to track the sales process or the conversion rate across a series of pages or steps.



Best Use Cases for These Types of Charts:

The most common use case for a funnel chart is the marketing or sales funnel. But there are many other ways to use this versatile chart.

If you have at least four stages of sequential data, this chart can help you easily see what inputs or outputs impact the final results.

For example, a funnel chart can help you see how to improve your buyer journey or shopping cart workflow. This is because it can help pinpoint major drop-off points.

Other stellar options for these types of charts include:

- Deal pipelines
- Conversion and retention analysis
- Bottlenecks in manufacturing and other multi-step processes
- Marketing campaign performance
- Website conversion tracking

Design Best Practices for Funnel Charts:

- **Scale the size of each section** to accurately reflect the size of the data set.
- Use **contrasting colors** or **one color** in gradated hues, from darkest to lightest as the size of the funnel decreases.

13. Bullet Graph

A bullet graph reveals progress toward a goal, compares this to another measure, and provides context in the form of a rating or performance.

Best Use Cases for These Types of Graphs:

In the example above, this bullet graph shows the number of new customers against a set customer goal. Bullet graphs are great for comparing performance against goals like this.

These types of graphs can also help teams assess possible roadblocks because you can analyze data in a tight visual display.

For example, you could create a series of bullet graphs measuring performance against benchmarks or use a single bullet graph to visualize these KPIs against their goals:

- Revenue
- Profit
- Customer satisfaction
- Average order size
- New customers

Seeing this data at a glance and alongside each other can help teams make quick decisions.

Bullet graphs are one of the best ways to display year-over-year data analysis. You can also use bullet graphs to visualize:

- Customer satisfaction scores

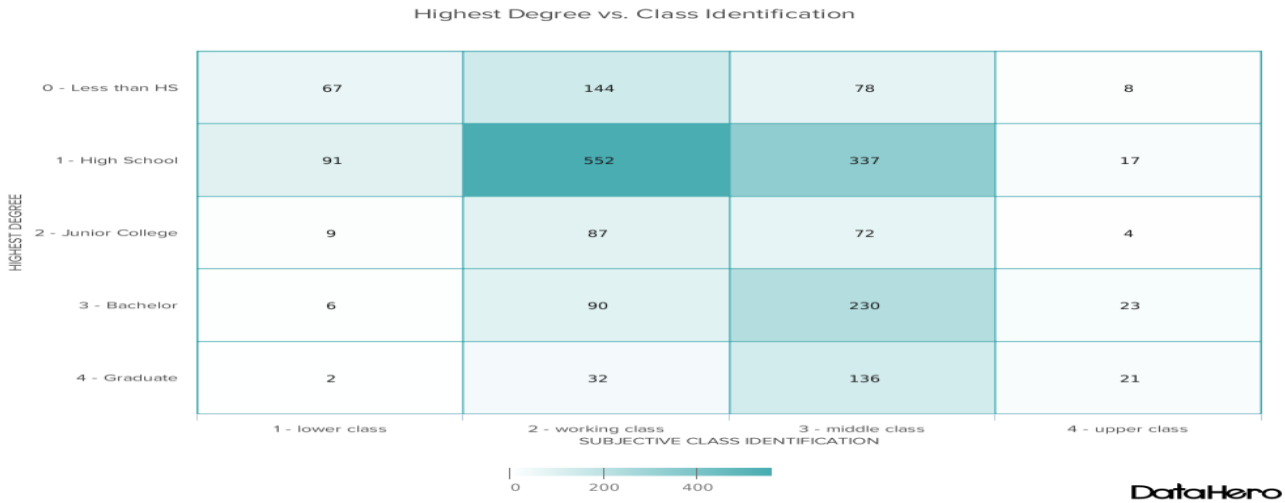
- Product usage
- Customer shopping habits
- Social media usage by platform

Design Best Practices for Bullet Graphs:

- **Use contrasting colors** to highlight how the data is progressing.
- **Use one color** in different shades to gauge progress.

14. Heat Map

A heat map shows the relationship between two items and provides rating information, such as high to low or poor to excellent. This chart displays the rating information using varying colors or saturation.



Best Use Cases for Heat Maps:

In the example above, the darker the shade of green shows where the majority of people agree.

With enough data, heat maps can make a viewpoint that might seem subjective more concrete. This makes it easier for a business to act on customer sentiment.

There are many uses for these types of charts and graphs. In fact, many tech companies use heat map tools to gauge user experience for apps, online tools, and website design.

Another common use for heat map graphs is location assessment. If you're trying to find the right location for your new store, these maps can give you an idea of what the area is like in ways that a visit can't communicate.

Heat maps can also help with spotting patterns, so they're good for analyzing trends that change quickly, like ad conversions. They can also help with:

- Competitor research
- Customer sentiment
- Sales outreach
- Campaign impact
- Customer demographics

Design Best Practices for Heat Map:

- **Use a basic and clear map outline** to avoid distracting from the data.
- **Use a single color** in varying shades to show changes in data.
- **Avoid using multiple patterns.**