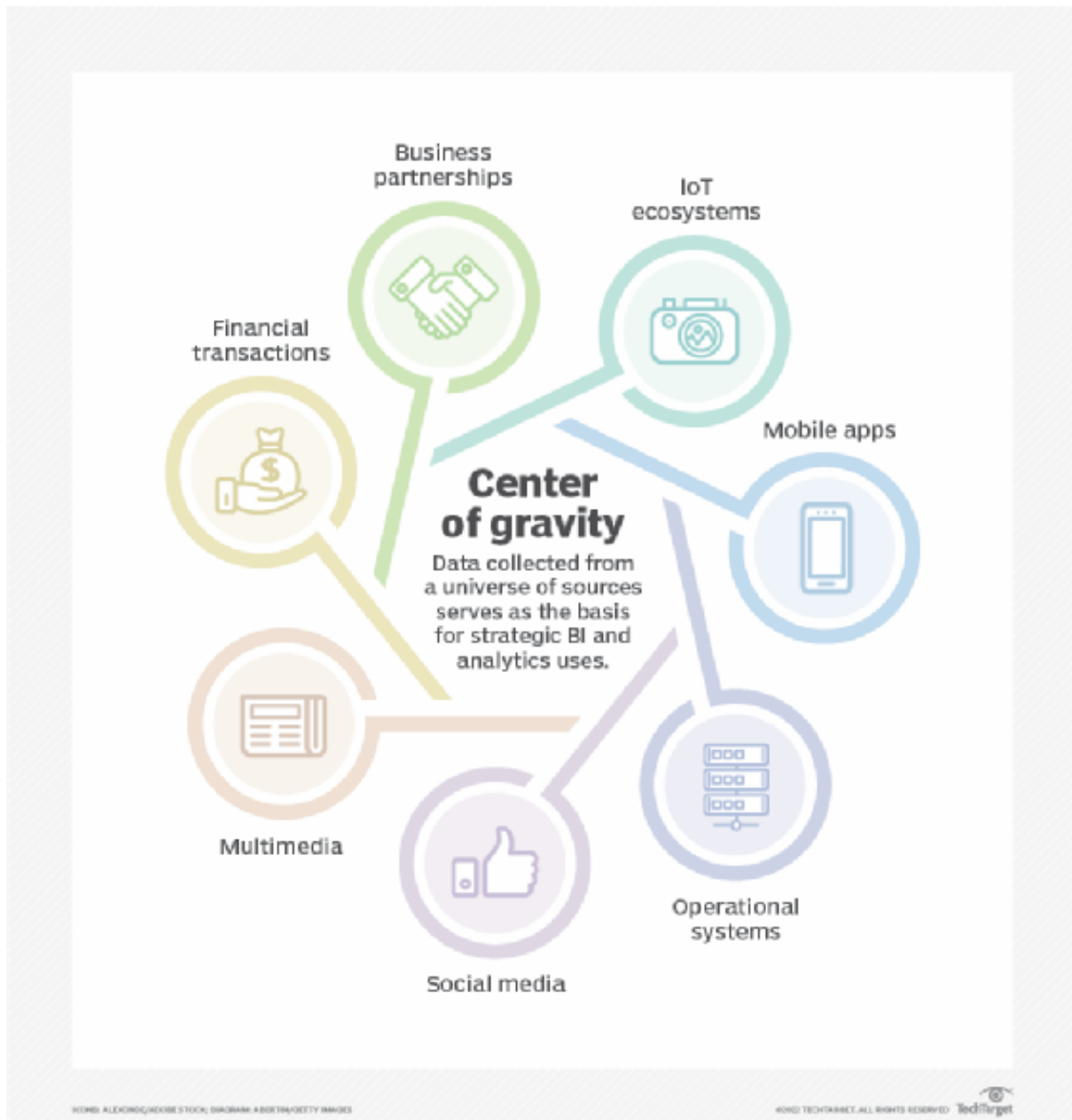Data Collection Strategy :

What is data collection?

Data collection is the process of gathering data for use in business decision-making, strategic planning, research and other purposes. It's a crucial part of data analytics applications and research projects: Effective data collection provides the information that's needed to answer questions, analyze business performance or other outcomes, and predict future trends, actions and scenarios.

In businesses, data collection happens on multiple levels. IT systems regularly collect data on customers, employees, sales and other aspects of business operations when transactions are processed and data is entered. Companies also conduct surveys and track social media to get feedback from customers. Data scientists, other analysts and business users then collect relevant data to analyze from internal systems, plus external data sources if needed. The latter task is the first step in data preparation, which involves gathering data and preparing it for use in business intelligence (BI) and analytics applications.

For research in science, medicine, higher education and other fields, data collection is often a more specialized process, in which researchers create and implement measures to collect specific sets of data. In both the business and research contexts, though, the collected data must be accurate to ensure that analytics findings and research results are valid.

Organizations collect data from a variety of systems and other data sources.

What are different methods of data collection?

Data can be collected from one or more sources as needed to provide the information that's being sought. For example, to analyze sales and the effectiveness of its marketing campaigns, a retailer might collect customer data from transaction records, website visits, mobile applications, its loyalty program and an online survey.

THIS ARTICLE IS PART OF

What is data preparation? An in-depth guide to data prep

- Which also includes:

-

DOWNLOAD 1

Download this entire guide for FREE now!

The methods used to collect data vary based on the type of application. Some involve the use of technology, while others are manual procedures. The following are some common data collection methods:

- automated data collection functions built into business applications, websites and mobile apps;

- sensors that collect operational data from industrial equipment, vehicles and other machinery;

- collection of data from information services providers and other external data sources;

- tracking social media, discussion forums, reviews sites, blogs and other online channels;

- surveys, questionnaires and forms, done online, in person or by phone, email or regular mail;

- focus groups and one-on-one interviews; and

- direct observation of participants in a research study.

These are some of the methods that organizations use to collect customer data.

What are common <mark>challenges</mark> in data collection?

Some of the challenges often faced when collecting data include the following:

- **Data quality issues.** Raw data typically includes errors, inconsistencies and other issues. Ideally, data collection measures are designed to avoid or minimize such problems. That isn't foolproof in most cases, though. As a result, collected data usually needs to be put through data profiling to identify issues and data cleansing to fix them.

- **Finding relevant data.** With a wide range of systems to navigate, gathering data to analyze can be a complicated task for data scientists and other users in an organization. The use of data curation techniques helps make it easier to find and access data. For example, that might include creating a data catalog and searchable indexes.

- **Deciding what data to collect.** This is a fundamental issue both for upfront collection of raw data and when users gather data for analytics applications. Collecting data that isn't needed adds time, cost and complexity to the process. But leaving out useful data can limit a data set's business value and affect analytics results.

- **Dealing with big data.** Big data environments typically include a combination of structured, unstructured and semistructured data, in large volumes. That makes the initial data collection and processing stages more complex. In addition, data scientists often need to filter sets of raw data stored in a data lake for specific analytics applications.

- **Low response and other research issues.** In research studies, a lack of responses or willing participants raises questions about the validity of the data that's collected. Other research challenges include training people to collect the data and creating sufficient quality assurance procedures to ensure that the data is accurate.

What are the key steps in the data collection process?

Well-designed data collection processes include the following steps:

1. Identify a business or research issue that needs to be addressed and set goals for the project.

2. Gather data requirements to answer the business question or deliver the research information.

3. Identify the data sets that can provide the desired information.

4. Set a plan for collecting the data, including the collection methods that will be used.

5. Collect the available data and begin working to prepare it for analysis.

Data collection considerations and best practices

There are two primary types of data that can be collected: quantitative data and qualitative data. The former is numerical -- for example, prices, amounts, statistics and percentages. Qualitative data is descriptive in nature -- e.g., color, smell, appearance and opinion.

Organizations also make use of secondary data from external sources to help drive business decisions. For example, manufacturers and retailers might use U.S. census data to aid in planning their marketing strategies and campaigns. Companies might also use government health statistics and outside healthcare studies to analyze and optimize their medical insurance plans.

The European Union's General Data Protection Regulation (GDPR) and other privacy laws enacted in recent years make data privacy and security bigger considerations when collecting data, particularly if it contains personal information about customers. An organization's data governance program should include policies to ensure that data collection practices comply with laws such as GDPR.

Other data collection best practices include the following:

- Make sure you collect the right data to meet business or research needs.

- Ensure that the data is accurate, either as it's collected or as part of the data preparation process.

- Don't waste time and resources collecting irrelevant data.