

Real-Time Processing: Difference & (Dis)Advantage Over Batches

July 13, 2021
by SentinelOne

Every organization has a mine of data. And when organizations realize that they're sitting on tons of valuable data, they want to put it to use. But data in its raw form isn't useful. You have to collect and process data before you can do something useful with it.

Data processing is the method of collecting data and converting it into its useful form. You can then use the processed data for analysis, analytics, intelligence, etc. Data processing comprises [multiple stages](#), namely the following:

- Collection
- Preparation
- Input
- Processing
- Output
- Storage

Of all these stages, the *processing* stage is where data is actually converted into its useful form. And that will be our topic of focus today. There are different methods of processing data. But we'll stick to the most popular methods—real-time and batch processing.

 Clock signifying real-time processing

What Is Real-Time Processing?

Real-time processing is the method where data is processed almost immediately. There's no pause or waiting in this method. These systems process data as soon as they receive input and give the processed data as output. Because of this nature, real-time processing usually requires a [continuous flow of data](#). Bank ATMs, multimedia systems in automobiles, and traffic control systems are some examples of real-time processing.

Real-time systems need to be fast. If they aren't fast enough to process data as it comes in, it will create load on the system, and the system won't be working in real time anymore.

If you think about it, nothing is really in real time. There will always be a slight delay. Even in the fastest real-time systems, there will still be a delay of 10 to the power of something seconds. So when we say "real time," we're talking about the processing that takes less time than a specified benchmark. This benchmark varies from system to system. For example, you can consider data processing of a bank ATM to be real time if it reacts in less than one-tenth of a second. But the same speed would be considered slow in a supercomputer.

Real-time processing is used when you need the output on the go, and it has its pros and cons.

Advantages of Real-Time Processing

- The delay in data processing is minimal.
- Information is up to date and can be used immediately.
- You would need fewer resources to sync systems.
- You have increased [uptime](#).
- It helps identify issues so you can take action immediately.

Disadvantages of Real-Time Processing

- It's difficult to implement with simple systems.
- It requires high-performance hardware and is expensive.
- It adds an overload of data in case of system failure.

What Is Batch Processing?

Batch processing is the method where you'd process data in batches. A large amount of data constitutes a batch. Once the batch is ready, it would be sent as input for processing. The whole batch of data is processed at once, and then the output is also produced as a batch. So everything happens in blocks of data.

Batch data processing is efficient when you need to process large volumes of data and don't need it to be in real time. Employee payslip generation and daily reporting are some examples of batch processing.

Batches can be decided based on the size of data or the period in which the data is collected. In the case of a batch based on size, you can create them based on the number of entries/records or the size of data. For example, you can create a batch when you have 1,000 records of data or when the size of the data is 1 GB. This would be helpful when your operation on processed data has a predefined size—for example, if you want to create a graph based on 1,000 entries. You will always need 1,000 entries for this operation. So you can create batches of 1,000 entries each.

On the other hand, in time-based batches, each batch comprises data collected in a particular period of time. For example, if you're analyzing data of previous working days, your batch size can be of five days. So what would happen is, data from Monday to Friday is collected, and this creates one batch. This batch is processed over the weekend, and processed data is ready for your analysis on Monday. And this repeats every week.

Advantages of Batch Processing

- It's efficient in processing large volumes of data at once.
- The effect of system failure or downtime is minimal on data processing.
- It's cost-efficient.
- There's no need for specialized hardware.
- Resources can be utilized for other tasks when data processing is not in action.

Disadvantages of Batch Processing

- It's difficult to debug.
- You'd need to wait for a complete batch to be processed to get even a part of the information.
- Employees need to be trained.

Now that we've understood what real-time processing and batch processing are and looked at their pros and cons, let's sum up and look at their differences.

Differences Between Real-Time Processing and Batch Processing

 Differences between Real-Time Processing and Batch Processing

Why Is Real-Time Processing Important?

Batch processing has its advantages, but we live in a world where time is a luxury we can't afford not to have. Organizations are spending loads of money on their business, and decision-making has become more dynamic than ever before. And to keep up with the rest of the world and stay ahead of your competition, you need real-time processing.

From an operational point of view, real-time processing gives you real-time stats and information on how your business is operating. This could be information related to processes or stats related to assets such as web applications. With real-time knowledge, you can make dynamic decisions that would be based on the most recent data, hence the most effective. You can identify issues early and work on fixing them.

From a business point of view, real-time processing can help you gain information about your customers and your business. This would help you make clever business decisions. For example, if you sell a product and notice that the demand for your product is growing, you can slightly increase the price to get more profits. Or if you notice that more people buy your product when it's on discount, you can decide on the discount value dynamically and what would be most optimal for that point in time.

I can go on and on about such cases where real-time processing can be beneficial. But the bottom line is that real-time processing is beneficial for your organization in terms of operations and business. One of the most useful implementations of real-time processing is streaming analytics that can be applied within almost every organization. So let's look at what this is.

 One of the most useful implementations of real-time processing is streaming analytics that can be applied within almost every organization.

Streaming Analytics

In closing, let's talk about streaming analytics and how it can help you work in real time.

[Streaming analytics](#) is the practice of looking at and analyzing data in real time rather than in batches. I'll explain this with an example. Web applications have become common and one of the most important assets of every enterprise today. Let's say you own an e-commerce web application. It's vital for your business that the web application is running as expected. If the website is down due to some issue, it affects the revenue. So, if something happens to your website and it goes down, then you'd want to know about it immediately. Knowing about it later in time after the batch of data is processed would result in major losses. That's why streaming analytics is important.

With streaming analytics, you realize that there's an issue and can identify and fix the issue very early and reduce losses. You can even monitor the performance of systems and identify problems. If streaming analytics is something that you're interested in or if you think it's something your organization would benefit from, you should check out [Scalyr](#).

Scalyr is a log management and analytics platform that can process data at a petabyte scale. It provides scaled event data ingestion and storage, subsecond query response, and low-cost overhead. You also have an impressive dashboard that helps you understand the information easily. If you want to experience what Scalyr offers, I suggest you [take Scalyr on a trial](#) run.

This post was written by Omkar Hiremath. Omkar is a cybersecurity analyst who is enthusiastic about cybersecurity, ethical hacking, data science, and Python. He's a part-time bug bounty hunter and is keenly interested in vulnerability and malware analysis.

Like this article? Follow us on [LinkedIn](#), [Twitter](#), [YouTube](#) or [Facebook](#) to see the content we post.

Read more about Cyber Security

- [Function as a Service \(FaaS\) Explained](#)
- [Scalyr and the Log4j Vulnerability](#)

