



Build the Future with Decodable and Stream Processing

Executive Summary

Businesses are leveraging stream processing to make smarter and faster business decisions, act on time-sensitive and mission-critical data, obtain real-time analytics and insights, and build applications with features delivered to end-users in real time. While the spectrum of use cases continues to expand across an ever-widening range of business objectives, common applications of stream processing include fraud detection, processing IoT sensor data, network monitoring, generating context-aware online advertising, cybersecurity analysis, geofencing and vehicle tracking, and many others.

Gartner Research reports that organizations are improving their decision intelligence and real-time applications by tapping the growing availability of streaming data. According to a comprehensive research report by Market Research Future (MRFR), the market size for event stream processing will reach over \$4 billion USD by 2027, growing at a compound annual growth rate of 21 percent. And an independent survey of 500+ CIOs and technology leaders by DataStax in 2022 revealed that leveraging real-time data pays off in two important ways: higher revenue growth and increased developer productivity. An impressive 71 percent of respondents agreed that they could tie their revenue growth directly to real-time data, while 78 percent of respondents agreed that real-time data is a “must-have,” not a “nice to have.”

By leveraging real-time data integration infrastructure that can power both high SLA and low-latency operational use cases, it is not necessary for businesses to maintain a separate data stack for data integration into analytical systems, as is traditionally seen with Extract-Transform-Load (ETL) and Extract-Load-Transform (ELT) systems. Currently, building real-time data apps and services involves connecting various sources and building a robust real-time pipeline. The process may sound simple at a high level, but it requires companies to build internal tooling that is pieced together using open-source parts and custom code. This requires the in-depth expertise of data platform engineers and other specialized, dedicated resources. These systems quickly become expensive to create and operate, even before a single pipeline is deployed. As a result, teams end up dedicating significant amounts of time, often months, just to get their real-time apps up and running.

Decodable’s mission is to make streaming data engineering easy. Decodable delivers the first real-time data engineering service—that anyone can run. As a platform for real-time data ingestion, integration, analysis, and event-driven service development, Decodable eliminates the need for a large data team, clusters to set up, or complex code to write. Decodable will provide its customers with:

- **The ease of use, adoption, and ecosystem of Snowflake**
- **The adoption model of DataDog, GitHub, AWS, GCP, Intercom, and Auth0**
- **The user experience sophistication of Superhuman and Intercom**
- **The developer love of dbt, pandas, Kubernetes, and Visual Studio Code**

What is Stream Processing?

Streaming data is the continuous, real-time flow of data generated by various sources. Stream processing, also referred to as event stream processing, acts on one or more data streams to analyze, transform, aggregate, and/or store those events, also in real time.

Stream processing allows applications to respond to new data events at the moment they occur. Rather than grouping data and collecting it at some predetermined interval, as with batch processing, stream processing applications collect and process data immediately as they are generated.

Data streams are generated by all types of sources, in various formats and volumes. Examples include applications, networking devices, IoT devices, server log files, website activity, banking transactions, location data, and many others.

Modern stream processing systems are parallel execution systems, meaning they can process large amounts of data because they split it into smaller chunks that are processed in parallel. Additionally, they are stateful, which means they can “remember” information across multiple records, which allows jobs to perform operations such as windowed aggregations. It has only been within the last 8 to 10 years that technology has been able to perform these kinds of computations in real-time.

Closely tied to stream processing is data integration, which is the act of getting data from one system into another, optionally performing transformation of the data. Methods for performing data integration are generally categorized in a few different ways:

- **Extract-Transform-Load (ETL)** refers to extracting (e.g., receiving, reading, querying) data from one system, transforming it, and then loading (e.g., sending, writing, inserting) it into another. The defining characteristic of ETL is that transformation primarily happens in a third system that exists between the source and the destination.
- **Extract-Load-Transform (ELT)** is equivalent to ETL except transformation occurs in the target system after load rather than in a separate third system. While this seems like a minor difference it has noteworthy implications as a consequence of electing to transform data after loading. The most important of these are that the structure of the data in the source system is mirrored in the target system, all of the data is moved to the target (no filtering, anonymizing, etc.), and transformation capabilities and properties are defined solely by the functionality of the sink system.
- **Reverse ETL** describes the process of extracting data from a data warehouse or data lake, processing it, and loading the results (primarily) into business applications. Primarily a marketing term, it still refers to extracting, transforming, and loading; there is no “reverse.” The emergence of the term comes from the assertion that ETL always ends up loading into the data warehouse, so the “reverse” is to extract from the data warehouse and send the data elsewhere. However, if the system has the necessary connectors for the source and sink systems, “reverse ETL” is simply ETL.

Benefits of Stream Processing

Gartner Research reports that organizations are improving their decision intelligence and real-time applications by tapping the growing availability of streaming data. According to a comprehensive research report by Market Research Future (MRFR), the market size for event stream processing will reach over \$4 billion USD by 2027, growing at a compound annual growth rate of 21 percent. And an independent survey of 500+ CIOs and technology leaders by DataStax in 2022 revealed that leveraging real-time data pays off in two important ways: higher revenue growth and increased developer productivity. An impressive 71 percent of respondents agreed that they could tie their revenue growth directly to real-time data, while 78 percent of respondents agreed that real-time data is a “must-have,” not a “nice to have.”

There are vast numbers of mobile devices, IoT devices, and applications capturing a wide range of human activities, all of which are streaming data into systems that are racing to keep up with exponentially increasing volumes of data. Companies of all stripes are grappling with this challenge, and even your data is part of that never-ending stream. Every time you upload to social media, complete an online transaction, or trigger the response of an IoT sensor, you are a part of countless other interactions that are also generating streams of data.

The fundamental value of streaming technologies is in being able to parse through all that data, in real time, and filter out the “noise” to find a useful “signal.” In other words, being able to gain useful and valuable insight from any number of massive and varied streams of data results in genuine competitive business advantage.

The traditional practice of transferring data into and out of static repositories and acting on them in batches will simply not suffice for many applications moving forward. Event stream processing allows for a faster reaction time and creates the opportunity for proactive measures to be taken before an opportunity is lost or a situation has passed.

Event stream processing provides an effective solution to many different challenges and gives you the ability to:

- Analyze high-velocity big data while it is still in motion
- Transform, filter, categorize, aggregate, and cleanse data before it is even stored
- Create applications and services that are able to respond in real-time
- Use fewer IT resources to process individual data points rather than large datasets
- Process data from a greater number of data and distributed event sources
- Continuously monitor data and interactions
- Scale according to data volumes
- Remain agile and handle issues as they arise
- Detect interesting relationships and patterns

Decodable is the Answer to Your Stream Processing Needs

ETL, ELT, reverse ETL, data ingestion, and stream processing for event-driven microservices will collapse into a single platform that leverages SQL for processing, operates in real-time, and connects to everything. Decodable is building that platform.

Currently, building real-time data apps and services involves connecting various sources and building a robust real-time pipeline. The process may sound simple at a high level, but it requires companies to build internal tooling that is pieced together using open-source parts and custom code. This requires the in-depth expertise of data platform engineers and other specialized, dedicated resources. These systems quickly become expensive to create and operate, even before a single pipeline is deployed. As a result, teams end up dedicating significant amounts of time, often months, just to get their real-time apps up and running. The challenges these projects must overcome can ultimately mean that projects do not come to fruition at all.

Fresher data and faster time to decisions is objectively better. The only reason the world doesn't work this way to today is because the tooling is more complicated, less mature, and less well-understood than its batch counterparts. Decodable is changing that. Once the skill gap is gone, so are the reasons people can't move their data infrastructure to real-time where and when it makes sense.

Decodable will provide it's customers with

- The ease of use, adoption, and ecosystem of Snowflake
- The adoption model of DataDog, GitHub, AWS, GCP, Intercom, and Auth0
- The user experience sophistication of Superhuman and Intercom
- The developer love of dbt, pandas, Kubernetes, and Visual Studio Code

Today, the world of data is separated into online operational data infrastructure and offline analytical data infrastructure. The two are generally connected by data integration. With real-time data integration infrastructure that can power both high SLA, low latency, operational use cases, there's no reason to maintain a separate data stack for data integration into analytical systems. By feeding both operational and analytical data infrastructure from the system and data, costs are reduced, compliance becomes simpler, data quality increases, and analytical systems better reflect reality.

Many companies are claiming the data warehouse is the center of the universe and, for many people, it is. However, if Snowflake or Databricks suddenly became inexpensive tomorrow, they wouldn't be low latency enough for operational workloads. If they were low latency, they wouldn't be transactional and strongly consistent. If they were transactional and strongly consistent, they wouldn't have application framework support.

The fact is, data warehouses are an enormous destination for data with many workloads, but not all. The operational world is made up of purpose-built systems: streaming systems, messaging systems, OLTP databases, key-value stores (DHTs), online feature stores, ultra-fast in memory caches, document databases, and a myriad of SaaS business applications, each optimized for a specific workload. While the data warehouse is trying to expand, it is simply not feasible for a single system to satisfy all of these workloads. The data warehouse has never been, nor will ever be, the center of the world. Embracing that reality, Decodable will be the platform that ties all of these systems together.