

Data Ingestion

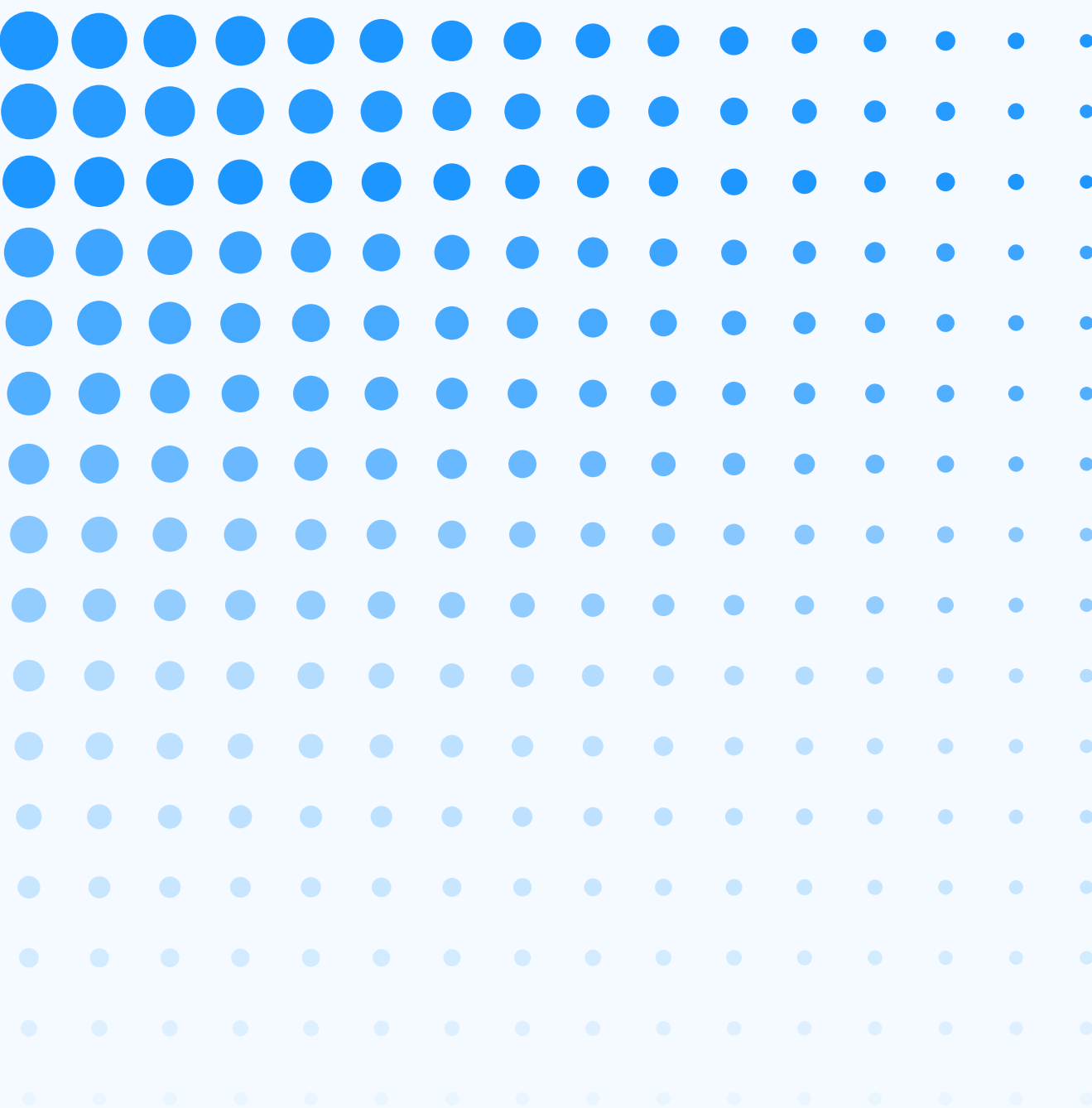
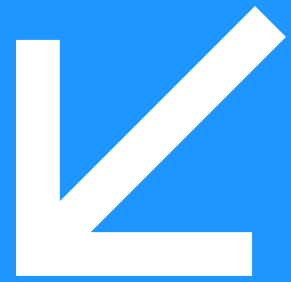
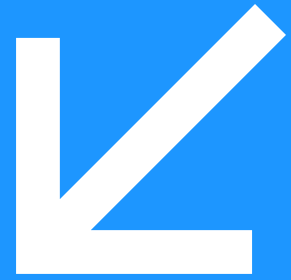


Table Of Contents



INTRODUCTION	1
WHAT IS DATA INGESTION?	2
BIG DATA INGESTION ARCHITECTURE	3
PARAMETERS OF DATA INGESTION	4
DATA INGESTION FRAMEWORK	5
TOP TIPS FOR DATA INGESTION	6
BEST PRACTICES FOR DATA INGESTION	7
COMMON TYPES OF DATA THAT NEED TO BE INGESTED	8
WHY IS IT ESSENTIAL TO HAVE AN EFFICIENT WAY OF INGESTING DATA?	9
HOW MUCH DOES DATA INGESTION AFFECT A COMPANY'S BOTTOM LINE?	10
WHAT TO LOOK FOR IN A DATA INGESTION TOOL	11
INTEGRATE.IO AS YOUR DATA SOLUTION PARTNER	12
INTEGRATE.IO'S UNLIMITED CAPABILITIES	13

Table Of Contents



INTEGRATE.IO'S SHORT LEARNING CURVE	14
INTEGRATE.IO'S DATA MIGRATIONS	15
ESTABLISH AND MAINTAIN APIS WITH INTEGRATE.IO	16
INTEGRATE.IO'S CENTRALIZED DATA SOLUTION	17
INTEGRATE.IO WILL REVOLUTIONIZE YOUR DATA INGESTION PROCESS	18

01

INTRODUCTION

Data is becoming a central aspect of conducting business in today's world.

Data ingestion, the process by which data becomes usable for analysis, can be complicated when there are many sources to consider. This whitepaper will look at how data solutions can help with data ingestion and other relevant information that you need to know about it.

02

WHAT IS DATA INGESTION?

Data ingestion is the process of collecting, transferring, and loading raw data into a storage or computing environment.

The purpose of this transfer and load is to make it usable by applications that will find some value in the data. Data can be ingested from many different sources:

- On-premise systems such as mainframes
- Servers and databases
- Cloud services like [AWS](#) SaaS products
- Mobile devices
- [IoT sensors](#)
- Websites - including social media sites like Twitter, Facebook, etc
- APIs (Application Programming Interfaces) are used to provide access to web-based software and many more.

There are usually two sub-tasks within ingesting data, depending on how the

source system deals with requests for information:

- **Push Ingestion:** This is where the source system automatically sends data as and when it is available to another location that has requested it. This can be done using a messaging queue or by directly calling an API endpoint (for example, SaaS products may provide push ingestion through their API).
- **Pull Ingestion:** Here, the application requests information from various sources - for instance, via APIs or direct database calls - and then stores this in its storage solution. Then the analysis process uses the stored data within its environment.

03

BIG DATA INGESTION ARCHITECTURE

There are several stages involved in ingesting data.

Data Ingestion Pipeline: A Data ingestion pipeline is a sequence of components that read files from a source, transform them into the desired form, and load that new content into another system for further processing or storage.

Source Systems: The source systems are the ones that provide access to raw data. These can be mainframes, servers, databases, and external sources such as social media sites and financial transaction feeds.

Transformation: This is where we change one datum into another - for example, converting a text file into CSV or JSON format. Various open-source tools are

available for this process, but these often require considerable expertise to use them effectively.

Gateway Systems: These are the first point of contact for data entering into a specific system. They can be used to filter out content that isn't required, or it may simply pass everything through unchanged and other components will do this filtering instead.

Target System: This is where the ingested information goes next - whether this is another transformation engine, an event store, some database, etc.

04

PARAMETERS OF DATA INGESTION

There are many different parameters to consider when designing a data ingestion pipeline.

Some standard parameters include:

1 DATA VELOCITY

Data velocity is the rate at which data is created and delivered. Because it's essential to ingest data quickly, this requires the ability for high throughput ingestion of information. Keep in mind that data velocity isn't just about ingesting raw data—it needs to be processed within a reasonable timeframe or stored for later use if there are problems with real-time processing.

2 DATA VOLUME

The volume of data being ingested varies depending on the number of events, files, or messages that need to be processed. High volume can cause an ingestion pipeline to become overloaded, causing delays in processing and analysis.

3 DATA FORMAT

The incoming data must conform to a specific file type to be read by other components within the pipeline - this may involve converting between formats during ingestion if they are incompatible.

4 DATA VARIETY

Different types of content require different techniques when being ingested - not only must transformation rules vary according to what kind of content they are working with (i.e., images, text, etc.), but other factors like ease-of-access come into play as well.



5 DATA QUALITY

Due to the nature of data ingestion, it's possible that a file could be malformed in some way - perhaps containing records with missing or incorrect information inside them, which must then be fixed before going onto other stages.

6 DATA STRUCTURE

The structure of the incoming file must be understood to determine where within the pipeline it belongs and what form any transformations should take.

7 LATENCY

Ingestion latency refers to how fast an event or record needs to enter into a system once it's been created. This is often dictated by application requirements which then leads back into considerations about volume and velocity above.

05

DATA INGESTION FRAMEWORK

For ingesting data to work correctly, it's necessary to have a flexible framework in place, which will allow the process to be carried out without too much difficulty.

1 DATA INGESTION WORKFLOW

This defines the order in which different components will be called upon to process information and when they should use each other's services.

2 CONNECTIVITY

Data sources must have a link with any components ingesting data from them - if the pipeline is set up correctly, it shouldn't require much additional work. It may involve setting up connections between systems, writing connectors or modifying existing ones, and so on.

3 DATASTORES

These can range from simple key-value stores to fully-fledged databases, such as [MongoDB](#). They represent an endpoint for ingested data and provide access for querying once it's been stored within them.

4 STREAM PROCESSING

These tools are intended to make use of real-time data and may be used for tasks such as aggregating information to reduce the overall load on a system.



5 LOGGING

Ingestion must be monitored closely so that any errors with files can be detected and dealt with accordingly—this will usually involve having some dashboard where metrics regarding the process can be viewed.

6 SECURITY

Ingesting large amounts of data from numerous sources increases security risks which must then be mitigated by implementing identity management systems, encryption, or other mechanisms.

7 TRACING

When events enter into a pipeline, they often pass through multiple different components before finally being written out onto datastores. Tracing follows each piece along the way so that problems can be identified and resolved more efficiently.

8 DATA INGESTION TOOLS

These are the software components that carry out data ingestions and consist of several different modules or libraries for dealing with tasks such as parsing, converting between formats, etc. There's also usually some scheduling system to ensure jobs get done at regular intervals (i.e., whenever new files become available).

9 DATA INGESTION SERVICES

Companies who develop their solutions may offer this as a managed service - essentially, they'll take responsibility for configuring and running pipeline components like those mentioned above.

06

TOP TIPS FOR DATA INGESTION

There are many different components involved in ingesting data. To make the process as painless and efficient as possible, it's essential to:

- Use a [metadata](#) system to help govern the overall ingestion process and make it more flexible.
- Make use of connectors wherever possible to reduce time spent on developing new solutions.
- Choose tools that are versatile enough to cope with different formats, volumes, etc.
- Use automation where appropriate so that data processing can be carried out quickly and easily whenever necessary or required by other components downstream.
- Monitor metrics closely during ingestions to detect errors as soon as they occur rather than after all records have been processed.
- Ensure that any security measures implemented are adequate to the level of risk involved. The more data ingested from multiple sources increases this risk - be sure you have a good system in place for detecting and dealing with breaches as soon as they happen.

07

BEST PRACTICES FOR DATA INGESTION

There are many different things companies need to take into consideration when designing a data ingestion framework, as there are many other potential factors that could cause problems if ignored.

Some of the best practices for data ingestion include:

- Ensuring that rules for transformation are both flexible and able to handle all different types of content
- Knowing what sort(s) of data will be coming at any given time so transformations can be applied accordingly
- Understanding where potential problems could arise beforehand so these can be fixed before doing anything else.

08

COMMON TYPES OF DATA THAT NEED TO BE INGESTED

There are many different types of data that need to be ingested.

Some examples include:

1 CUSTOMER INFORMATION

Ingesting customer information is necessary for companies who want to ensure they can regularly contact their customers and send them promotional offers or other new content/updates, which helps build customer loyalty and retention rates. This also includes transactional data such as payments made by credit card, bank transfer, etc. Having systems in place where all kinds of payment methods can quickly be plugged in makes tracking sales much easier over time, so everything works without any problems arising due to inflexibility.

2 CONTENT FROM SOCIAL MEDIA PLATFORMS

This is critical for companies who want insight into what consumers are saying about their brand/products to improve how they communicate with people, which helps build a stronger relationship between them and their customers while building trust. However, there would be no way of knowing any of these things even exists without data ingestion. Companies that rely heavily on social media marketing campaigns are thus in dire need of data ingestion to run their campaigns effectively.



This is especially true since all that information could quickly get buried away among other content, becoming increasingly harder to find—and then parse out what customers are saying about a company/brand or what they're interested in—without having to spend hours trawling through everything.

3 SENSORS/IOT DEVICES

If companies want to get insight into how their products are performing (i.e., where usage is highest, what components need replacing, and so on), then they'll likely want this sort of data to make improvements over time. However, without the necessary systems, it would be impossible for them to understand or improve anything about their operations.

4 ONLINE MARKETING PLATFORMS

This helps companies keep track of where their money goes online, which gives them insight into whether specific campaigns were successful and if not, why not. One example would be tracking how much traffic was generated because of particular keywords used for paid search advertisements. It makes it possible to see just how effective specific copy has been over time.

4 INTERNAL DATA

Internal data, such as sales and inventory records, are essential for making sure everything works properly. If there's no way of knowing whether or not a product in stock has been sold, then companies run the risk of having too much on hand, which could lead to them losing money and, in worse cases suffer from additional costs due to storage fees. Without knowing what's going on with sales or inventory, it would be impossible for companies to know which products are popular and which ones aren't. This is important as popularity can help determine where future budgets go to improve how a company operates over time.

09

WHY IS IT ESSENTIAL TO HAVE AN EFFICIENT WAY OF INGESTING DATA?

Having an efficient way of ingesting data is critical for companies as it ensures they'll be able to make the most of their data.

If data isn't adequately ingested, then even if a company has access to information on how customers respond to specific campaigns and products, those insights would be useless. The data would not help improve anything since there wouldn't be any system where that knowledge could quickly and effectively be acted upon.

10

HOW MUCH DOES DATA INGESTION AFFECT A COMPANY'S, BOTTOM LINE?

In today's world, it is difficult for companies to operate without having the necessary systems to effectively ingest different kinds of information.

Data is the backbone of every modern business, so companies rely on it to make decisions based on analytic insights.

This is especially true for companies that rely on social media platforms to promote their brand.

By capturing as much data as possible, companies can form a 360-degree view of

their customers, which helps them build a stronger relationship with them and keep them as loyal patrons of their brand. This, in turn, builds trust and makes it possible for companies to improve how they communicate with consumers over time while gaining valuable insights into where future budget allocations should go towards improving operations.

11

WHAT TO LOOK FOR IN A DATA INGESTION TOOL

There are many things that companies should look for when deciding which data ingestion tool would be best suited to their needs.

Some of the top considerations include:

1 WHAT KIND OF DATA DOES THE COMPANY NEED TO INGEST?

In some cases, a company may have access to big data that needs to be ingested into their existing systems. In other instances, they'll only need something simple to provide information on what's going on in real-time without having much customization required for how the tool operates.

2 HOW QUICKLY MUST THIS DATA BE PROCESSED?

Data ingestion tools have different speeds for processing and storing information, so companies should look at getting one with fast speed capabilities. If there are any delays in accessing or acting upon important insights, all efforts would ultimately go down the drain as decisions won't get made until too late due to lack of adequate time available for them (i.e., data processed long after it is ingested).



3 WHAT DOES THE COMPANY ALREADY HAVE IN PLACE?

Before choosing a tool that can ingest new information into their systems, companies should assess what infrastructure they currently have to make sure there's no duplication of efforts involved with using different tools for particular purposes when one would suffice if appropriately used.

4 WHICH INDUSTRY VERTICAL?

Some industries require special features from their ingestion tools, such as compliance capabilities for security-related issues for financial services or cloud storage options compliant with HIPAA standards in healthcare and so on.

5 SHOULD THE SOLUTION BE CLOUD-BASED?

Depending on the needs of a company, they may want to consider using a cloud-based solution. This is especially true for companies with a lot of data coming in from different parts of the world, as it would be much easier to access and share information if everything was stored on a cloud-based platform.

6 WHAT ARE THE COSTS INVOLVED?

Finally, another essential consideration to make is that of cost. While it's not the only aspect a company should consider when looking for an ingestion tool, there are many different price points on the market depending on how much information needs to be ingested and what kind of features they want, as explained above.

This could range from something affordable to enterprise-level solutions that can get very costly but provide more comprehensive capabilities for those who need them.

12

INTEGRATE.IO AS YOUR DATA SOLUTION PARTNER

Integrate.io is the company to turn towards when looking for data ingestion tools.

Integrate.io can provide a variety of different solutions which are all compatible with any system currently in use by businesses today. Integrate.io makes it easy for companies who don't have any experience ingesting

their data or working with ETL, which means that they save time and money by not hiring expensive consultants, etc. since everything works right out of the box.

13

INTEGRATE.IO'S UNLIMITED CAPABILITIES

Integrate.io's data ingestion tools are designed to work seamlessly on any platform.

Which means that no matter what kind of system a business is using, Integrate.io will be able to provide an efficient way for them to get their data into the correct format so it can be used however they need to make the most sense possible.

1 ENRICH YOUR DATA

Once data has been appropriately ingested, it's possible to enrich the information further so that companies can see even more insight than before. Integrate.io can help with this by aggregating data from multiple sources, so it's possible to get a complete picture of how customers respond and what campaigns are the most successful.

2 BUILT TO SCALE

As your company grows, it's essential to scale your data solutions to meet the demands of how much information is being processed. We at Integrate.io have built our tools so that they are always scalable, which means no matter what kind of influx there might be when it comes to customers or products/services. Businesses will never have any problem with not being able to ingest the necessary amount of data whenever needed for reporting and analysis.



3 INTEGRATIONS

Integrate.io also allows you to integrate from over 100+ different data sources, which means you'll be able to pull data from just about any place you can think of. This kind of flexibility is critical for businesses that need to work with various types of data all at once and not have to worry about compatibility issues or lose time testing out multiple systems.

4 NO MORE MISSED OPPORTUNITIES

By having access to all kinds of insights about how customers feel or specific campaigns where people engage in whatever way, companies will never again have the possibility of missing opportunities since everything is accessible immediately whenever needed. This allows business owners who know how to interpret big data to grow faster than ever before.

14

INTEGRATE.IO'S SHORT LEARNING CURVE

For companies who have never worked with Integrate.io or ETL technology before, it's essential to know that the learning curve is extremely short, which means you'll be able to start ingesting data within just a few minutes after signing up. This allows even smaller businesses without much technological expertise but needs an efficient way of getting their data into shape for reporting to do so easily and quickly, resulting in faster business growth over time.

1 RAPID IMPLEMENTATION

Integrate.io's platform is easy to implement and works right out of the box. Everything gets set up in a matter of minutes, allowing

business owners to start seeing results immediately. This means that there will never be any downtime or long periods where data doesn't get ingested.

2 SIMPLE PIPELINE SET-UP

Integrate.io's tools are designed with the end-user in mind, which means that even if you don't necessarily have much experience working with ETL or data ingestion, there will never be any complicated pipelines to set up because everything is easy. With drag and drop features, configurable pipelines, and an easy-to-navigate low code interface, Integrate.io makes it possible for any business to get the data insights needed to make better decisions.

15

INTEGRATE.IO'S DATA MIGRATIONS

As your company changes over time, it's essential to migrate data from one system to another without losing any information along the way.

Integrate.io provides a way for this, which allows you to move all kinds of different sources into whatever systems are most efficient at that moment in time easily and quickly. Hence, there is no downtime or loss of vital [business intelligence](#).

1 FAST DATA MIGRATIONS

Migrating data can take up valuable time, especially if there is a lot involved. Still, with our tools, everything gets done quickly because we know how much speed matters when making better decisions due to having access to insightful data about your customers.

2 CONNECTIONS TO LEGACY SYSTEMS

Integrate.io offers a way for companies who may be stuck using [legacy technology](#) to connect to our tools so they can enjoy the benefits that come along with using all of Integrate.io's features without having to change their entire system or getting rid of everything that currently exists.



4 CUTTING-EDGE SECURETL TECHNOLOGY

Integrate.io's SecurETL technology provides a way for companies to manage their data with the utmost security by using multiple levels of authentication and encryption, which means everything stays protected at all times.

16

ESTABLISH AND MAINTAIN APIS WITH INTEGRATE.IO

Companies who need to establish and maintain APIs for multiple systems to work together seamlessly will find this easy with our tools since they don't require much time or effort when it comes down to getting everything set up.

Maintain Other Source APIs - API

management is made more accessible with Integrate.io's connectors, so you won't have to worry about maintaining the APIs for many different sources and applications.

Integrate.io API - With our API, you're also able to easily integrate with your current systems and applications no matter how many different sources you're using.

17

INTEGRATE.IO'S CENTRALIZED DATA SOLUTION

By centralizing all your data in one place, you're able to have a better handle on all your most important assets. This means that no matter how big or small the business is, using Integrate.io for their BI needs will be well equipped to make informed decisions about what they need from their company.

18

INTEGRATE.IO WILL REVOLUTIONIZE YOUR DATA INGESTION PROCESS

Data ingestion is one of the significant challenges in big data and analytics, which has been a bottleneck for companies looking to make better decisions about future plans. Integrate.io is an all-inclusive platform that can handle any source or type of data you want to use.

Integrate.io's data centralization is a game-changer for companies looking to make better decisions based on the most accurate data possible. If your business is looking for a data ingestion solution, schedule a demo today with Integrate.io through [the link here](#).