

Engineering Team Solutions

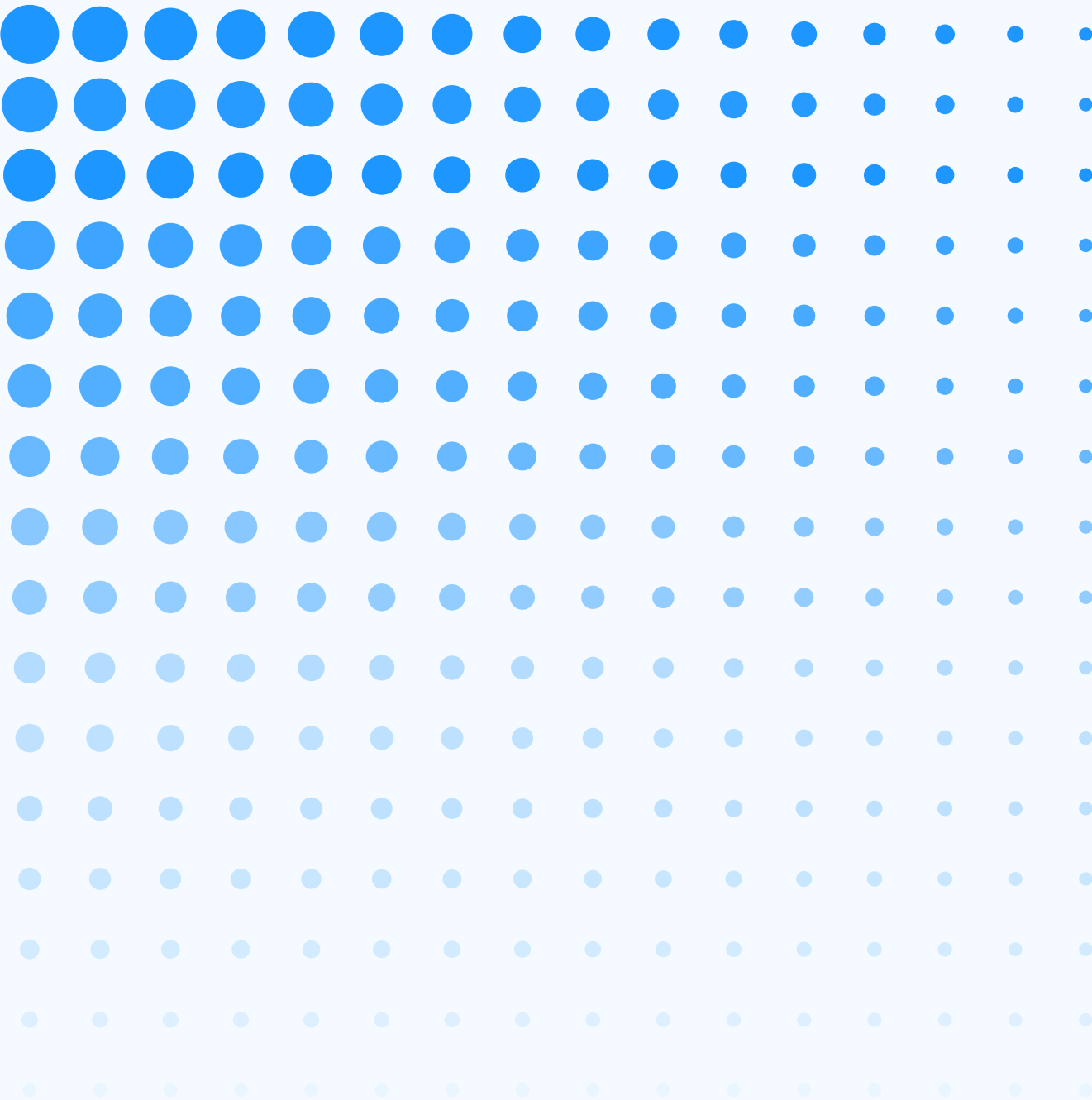


Table Of Contents



INTRODUCTION	1
WHAT IS A DATA PIPELINE?	2
COMPONENTS OF AN ELT DATA PIPELINE	3
WHY DOES MY BUSINESS NEED A DATA PIPELINE TOOL?	4
HOW DO DATA PIPELINE TOOLS WORK?	5
BENEFITS OF DATA PIPELINE TOOLS	6
WHAT ARE SOME CONSIDERATIONS WHEN FINDING A DATA PIPELINE TOOL?	7
WHAT ARE THE MOST COMMON REASONS COMPANIES NEED A DATA PIPELINE SOLUTION?	8
INTEGRATE.IO AS YOUR DATA PIPELINE SOLUTION	9
INTEGRATE.IO'S UNLIMITED CAPABILITIES	10
INTEGRATE.IO'S SHORT LEARNING CURVE	11
INTEGRATE.IO'S DATA MIGRATIONS	12
INTEGRATE.IO CAN ESTABLISH AND MAINTAIN API'S	13
INTEGRATE.IO IS YOUR ONE-STOP-SHOP FOR ALL YOUR DATA PIPELINE NEEDS	14

01

INTRODUCTION

Data Pipelines are critical to the success of most companies.

They allow for data-driven decision making automated alerting systems that save time and improve team productivity, alongside many other benefits. Many engineering teams are moving to the cloud for their software development, but they quickly realize that data pipelines can be complex. Data is generated in many different ways, and it all needs to be processed in a meaningful way for the downstream application. This document discusses how data pipeline solutions can help engineering teams with high throughput data requirements.

02

WHAT IS A DATA PIPELINE?

The pipeline is the workhorse of an organization's data operations. It can be composed in many different ways using various technologies depending on your needs for performance or cost.

With the popularity of data pipelines for both big and small software companies, many engineering teams are implementing these solutions to increase developer productivity. A pipeline is a series of steps that takes raw data from one place and transforms it into another. One type of pipeline is an Extract-Load-Transform (ELT) pipeline. Data Pipeline Solutions can help your team improve their development process by providing the tools to automate tasks like data cleansing, transformation, loading, ETL jobs scheduling, and monitoring/alerting errors.

Many teams have started using Data Pipelines, enabling them to work smarter and more efficiently when tackling complex problems involving large quantities of structured/unstructured data from disparate sources.

03

COMPONENTS OF AN ELT DATA PIPELINE

If you could create the perfect customer experience for every person who steps foot inside your store, wouldn't you want to?

Three key components make up an ELT data pipeline solution:

Extract - Ingests or extracts data from a variety of sources or applications to move it into the pipeline.

Load - Manages the storage, organization, and versioning of your datasets within Alation.

Transform - Enables you to run complex transformations on your newly ingested dataset via our powerful query language.

04

WHY DOES MY BUSINESS NEED A DATA PIPELINE TOOL?

You need a data pipeline tool when you need to:

- Integrate data from different sources.
- Handle data of various types.
- Face challenges on how to clean and integrate dirty, incomplete, or incompatible datasets (i.e., heterogeneous data).
- Timely respond to ad hoc requests for complex analyses/querying datasets with complex schemas.

- Reducing the time required for exporting results into the desired output format like .csv file or PowerPoint chart, etc.

Data Pipeline tools also help companies comply with the regulation by:

- Providing audit-ready records.
- Delivering secure data immutably.
- Tracking changes in data.
- Cleaning up messy datasets.

Data Pipeline tools provide several additional benefits, including:

- Helping people collaborate more easily.
- Making it easier to perform iterative analytics.
- Enabling teams to run exploratory analysis quickly.

05

HOW DO DATA PIPELINE TOOLS WORK?

Data pipeline tools are available in different forms depending on the needs of your organization.

They can be fully managed SaaS solutions hosted by a service provider, or they can be self-managed infrastructure that you install within an organization's network, so it is accessible only to those who have permission to access it.

These tools help engineering teams automate data transformation and move large amounts of data into their big data platform like Amazon EMR (Elastic Map Reduce), Google Cloud Platform, HD Insight, etc., without relying on ETL developers for writing this code every time new datasets need loading. Data Pipeline Tools allow organizations to develop pipelines quickly with less hassle as engineers don't require a deep knowledge about how these pipelines run under the hood.

Engineering teams can use the data pipeline tools to spin up an instance, configure it with their data sources, and ingest data.

Data pipeline tools use many use cases that help engineering teams work on their projects more efficiently by saving them time and resources. Some of these include:

- Ingesting or transforming incoming data to build analytics reports or dashboards with existing BI tools like Qlikview, etc.
- Building a data pipeline that ingests and processes logs from devices like sensors, smart appliances, etc.

- Creating an automated ETL process to consolidate multiple datasets into one schema for better analysis/reporting purposes.
- Automating the entire business intelligence (BI) life-cycle within your organization by setting up scheduled jobs to update dashboards with new reports in real-time so stakeholders can always get the latest information on top of their desks.
- Data Pipeline tools allow you to create scheduling jobs quickly using drag-and-drop features while giving users complete control over how each step is triggered and executed. The tool also allows engineers to configure alerts when certain thresholds are breached.
- Leveraging historical data for analytics, machine learning, or building real-time dashboards, etc.
- Data cleansing & transformation to make it ready for downstream applications like an ETL tool which further processes the dataset in the desired format before pushing it into the destination database schema, etc.

06

BENEFITS OF DATA PIPELINES

There are many advantages to using a data pipeline tool. Here are a few:

1 TEAM COLLABORATION & AGILITY

These tools allow teams to collaborate more quickly with better visibility, making it easier for everyone (data scientists, developers, engineers, etc.) to understand what's happening so they can move faster.

2 REPEATABILITY & AUTOMATION

The process becomes repeatable as all transformation steps are automated across multiple environments like test/dev, allowing companies to leverage CI/CD (Continuous Integration / Continuous Delivery) practices, thus increasing agility within their engineering teams while improving product quality.

3 EFFICIENCY & PRODUCTIVITY

Engineers no longer have to perform repetitive tasks when the data pipeline tool does it for them, leaving engineers with more time to build innovative products and services.

4 MANAGING DATA SECURITY REQUIREMENTS

Many organizations must maintain detailed records of all data changes or access by specific users over a given period to ensure compliance with industry regulations like GDPR (General Data Protection Regulation). Pipeline tools provide audit-ready reports on who accessed/modified specific datasets at different points in time so companies can be confident that their sensitive information is secure while still enabling teams within an organization to collaborate quickly without compromising on data security.

5 CONTROLLED ACCESS

Companies can set permissions on who has access to specific datasets and different stages of the pipeline and define triggers for each step to ensure that new dataset versions are created whenever an authorized user makes a change.

6 EASY TROUBLESHOOTING

Data transformation steps like validation, cleansing, etc., run automatically without manual intervention, creating consistent results while preserving data lineage/provenance, enabling teams to trace back issues when needed quickly and easily.

7 DATA INGESTION

Data pipeline tools enable teams not just to ingest but store large volumes of data from disparate sources into one central location, making all this information easily accessible. This means they no longer need costly ETL pipelines or complicated extractions built on top of various databases & file stores.

6 USAGE TRACKING

Data pipeline tools can track which datasets are being used and who is using them, so you know where your data assets stand from a business perspective, allowing for better decision-making for budgeting & resource allocation.

7 METADATA MANAGEMENT

In addition to tracking usage of source files, organizations can leverage metadata embedded in these file stores to automatically generate new fields that capture critical attributes about each dataset, such as author information and creation date, with no overhead required by their team members. This means developers don't need to spend time tagging every row in spreadsheets, saving them tons of time and ensuring that relevant metadata is consistently attached to each dataset.

07

WHAT ARE SOME CONSIDERATIONS FOR FINDING A DATA PIPELINE TOOL?

There are several considerations to keep in mind while choosing the right data pipeline tool for your needs:

What are the Essential Features for your Team?

Think about what you need to accomplish and how that fits into your broader data strategy. Is it just a single application, or do you want something that can scale across multiple teams in different stages of their development life-cycle with an enterprise-grade solution?

What Level of Support is Required by your Organization?

On-premise, cloud-hosted, etc.? Think through these questions so you have answers when comparing pipeline tools available in the market today.

Time & Price

How much time does each member of your engineering team spend every week performing tasks related to extracting & preparing various datasets needed within their applications? Use this information and cost estimates provided by vendors while evaluating them against one another based on capability vs. price.

Are there any Regulatory Requirements you Need to Adhere to?

Like GDPR, HIPAA, etc.? Ensure that the data pipeline tool supports all your needs and is easy enough for developers without requiring engineering resources.

Can it integrate with your Data Sources?

Does it integrate with existing applications, or do you need a separate dashboard created specifically for your team?

Data Privacy & Security

Organizations need to ensure that their data privacy and security requirements are being met by any tool they choose.

Scalability & Speed of Uploads/Updates

While there is a wide range of solutions available in the market, some tools may not handle larger datasets or have speed issues when ingesting new data from source files, for example, which could affect your ability to deploy updates quickly. In such cases, it's essential to look at what other options you have before committing to one solution over another.

Operational Complexity vs. Business Value

In many organizations, engineering teams aren't always involved in choosing technical tools, thus making it difficult for them to get up and running with the preferred solution. This is why it's essential to select a solution that your teams can get started with quickly and easily, so they don't spend valuable time on the tool instead of building new applications or services for their customers.

08

WHAT ARE SOME OF THE MOST COMMON REASONS COMPANIES NEED A DATA PIPELINE SOLUTION?

1 REGULATORY COMPLIANCE

For example, many organizations have to meet specific compliance requirements, whether it's GDPR, HIPAA, or PCI. They all require that you track user activity & usage to ensure proper controls are in place.

Automating these activities through your data pipelines can simplify this process by eliminating manual processes and allowing you more time to focus on what matters.

2 INTERNAL AUDIT CONTROLS

There's nothing worse than having an audit come back with findings because it shows there was no absolute control over data assets and time spent trying to find out who did what when is wasted money down the drain. Make sure you have data pipelines built to make auditing your business processes as easy & seamless as possible.

3 DATA QUALITY

One of the biggest challenges companies face today is maintaining quality, consistency, and integrity in their data which can be challenging when you're dealing with different sources that are all being used for other purposes. Having a solid data pipeline solution will help ensure everyone is on the same page. Hence, there aren't any gaps or surprises because, at some point downstream, it could lead to losing revenue if something isn't done about it quickly enough.

09

INTEGRATE.IO AS YOUR DATA PIPELINE SOLUTION

Integrate.io can help solve each of these challenges and more.

Our platform is flexible enough to handle any data and integrates easily with the tools your team is already using. We also provide a wide range of easy-to-use toolkits & prebuilt connectors which means you can get up and running quickly without needing engineering resources. And since we support cloud-based databases, scaling is never an issue, so if things change down the line in terms of future needs or regulations coming into play, Integrate.io will be ready for that too. Below are all the features that Integrate.io's platform can provide for your data pipeline needs.

10

INTEGRATE.IO'S UNLIMITED CAPABILITIES

Data pipelines allow your team to scale and build things faster.

They allow for data-driven decisions, automated alerting systems that save time and improve productivity, alongside many other benefits.

Built to Scale

Implementing an ETL process can be a time-consuming and expensive exercise. Data pipelines offer the flexibility to start small and scale as needed without making significant upfront investments or making changes in your current infrastructure. Integrate.io allows you to quickly build a pipeline with an easy-to-use UI and low code technology.

Focus on Your Core Business

With a data pipeline, your team can focus on getting meaningful results for business owners and executives. With Integrate.io's pre-built connectors, you can quickly and easily load your data from various sources. You'll be able to deploy your reports and dashboards to your team quickly.

Improved Productivity

Time is a critical resource when it comes to available engineering hours during crunch time or launch periods. Our platform allows you to automate manual processes so that teams can save time previously spent on tasks such as executing queries against databases and discovering insights about datasets.



Access All Your Data Sources with one Toolset

Data warehousing solutions are often limited by what they offer out-of-the-box - making building custom workflows more costly than necessary given the changing needs of business leaders who need access to new sources every day. Integrate.io allows you to easily connect all your data sources in one place so that everyone can access it on their own time and with no limitations on the number of users.

Advanced-Data Connectors

Integrate.io's out-of-the-box connectors allow easy access to over 100+ sources, including Salesforce CRM, Marketo Marketing Automation Platforms, Zendesk Customer Service Tools, GitHub Repositories, etc. With the Integrate.io ETL framework - developers can create custom connectors using simple UI tools, so they don't need any coding knowledge required.

RestAPI Integration

Our product allows for the use of RestAPI integration. This industry-standard interface can be used to build custom integrations with other popular business intelligence tools like PowerBI and SAS.

Snowflake Integration

Integrate.io is one of the few fully managed data pipeline solutions that offer native integration with Snowflake, an enterprise data warehouse service. There are many benefits of using this type of connection, including zero-footprint architecture and hardware requirements along with the ability to leverage both inbound and outbound pipelines.

Redshift Integration

Redshift is a popular data warehouse solution that allows for fast and scalable querying of datasets. Integrate.io offers native integration with Redshift to load your data quickly into this cloud-based storage system via our intuitive UI integration.

MongoDB Integration

Integrate.io offers native integration with MongoDB, a popular open-source NoSQL database. This lets you load quickly and transform your data in real-time via our intuitive UI integration with no code required.

Amazon RDS Integration

Our product also offers native integration with Amazon RDS, which is a cloud-based relational database service.



MySQL Integration

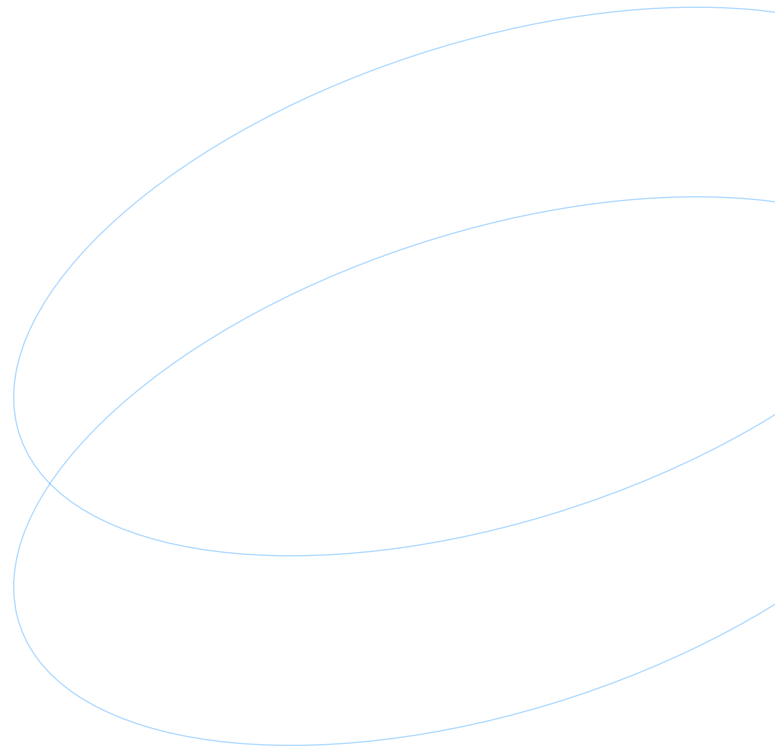
Integrate.io offers native integration with MySQL, a popular open-source relational database that entirely replaces the Oracle Database.

PostgreSQL Integration

Our product also provides native integration with PostgreSQL, an object-relational database management system (ORDBMS) available for many platforms, including Linux and UNIX and Windows operating systems. This allows you to have access to powerful features such as primary keys and foreign keys, within Integrate.io without having any expert knowledge of how databases work. Our software will handle all the complexity of loading and transforming your data seamlessly through a simple user interface.

Hadoop Integration

Our product also offers native integration with Hadoop, which is another popular open-source NoSQL database. Hadoop distributes data processing across clusters of computers using simple programming models.



11

INTEGRATE.IO'S SHORT LEARNING CURVE

The great thing about data pipeline tools is that they take a lot of the heavy lifting off your plate and allow you to focus on more important things.

Integrate.io offers an intuitive UI where users can quickly build workflows through drag-and-drop functionality without having extensive knowledge of code or scripting languages like SQL.

1 STREAMLINED WORKFLOW MANAGEMENT

Integrate.io offers streamlined workflow management, which helps make managing complex integration tasks easier than ever before - whether it's scaling out new ETL jobs automatically for increased load or reducing dependency on IT resources by building custom integrations via RestAPI connections.

2 SIMPLE PIPELINE SETUP

Integrate.io offers an easy pipeline setup, so you can build data pipelines in just minutes. Integrate.io supports low code and standard SQL syntax out of the box, so users don't have to learn complex scripting languages or new connectors for each different database they need to connect with. You'll also avoid creating errors for common mistakes like adding unnecessary quotes, which can save you time and resources.



3 FANATICAL SUPPORT

Integrate.io offers top-of-the-line support to our users, so if you run into any issues with your data pipeline solutions or have general questions about the platform - there will always be someone available to help. We can assist via unlimited phone, chat, email,

and video support - ensuring that you never have to worry about how complex your data integration tasks are.

12

INTEGRATE.IO'S DATA MIGRATIONS

Integrate.io offers data migrations so if you need to upgrade or downgrade your database - the process will happen without downtime, so there is no need for any maintenance windows.

Centralized Management Console

With our centralized management console, managing complex integration tasks has never been easier.

Data Synchronization

Integrate.io offers data synchronization, so you can seamlessly sync your cloud and on-premise databases without having to worry about platform differences, maintaining different connectors, scripts, etc. Our software will handle all the complexity of loading and transforming your data seamlessly through a simple user interface.

Automated Scheduling & Monitoring

With automated scheduling and monitoring - our platform allows you to quickly scale out new ETL jobs automatically based on updated data sources or schedules with no intervention from IT staff required. This means that scaling issues become less frequent because of peaks in traffic, making it easier for your team to meet deadlines.



Connections to Legacy Systems

Integrate.io offers connections to legacy systems so you can effortlessly connect your big data applications with all kinds of different databases without having to worry about complex coding or scripting languages. Legacy databases can be a challenge in terms of data integration, but with Integrate.io, you find that the process is much easier.

Cutting-Edge Security Technology

Integrate.io offers innovative securETL technology, which means that our platform ensures data security by locking down all sensitive information in the data center via encryption and compliance controls. Use Change Data Capture to ensure only the most recent records are being loaded and prevent overwriting essential data points.

Data Governance & Compliance

Integrate.io offers Data Governance and compliance, which means that you can enforce rules or even a full-fledged workflow to automate your ETL processes from source to destination without any downtime - ensuring all of your information is accurate, secure, and within regulatory requirements at all times.

Ease of Management & Maintenance

When it comes to ease of management and maintenance, Integrate.io is a compelling solution since there are no servers to manage, making it ideal for any business looking to boost its bottom line.

13

INTEGRATE.IO CAN ESTABLISH AND MAINTAIN APIS

Establishing and Maintaining APIs allows companies to create real-time integrations between applications, data sources, or databases without writing code. This saves time and money since Integrate.io's team of experts manages all the technical details instead of your IT department, which can take up valuable resources that could be better spent elsewhere within your business.

1 **CREATE AND MANAGE WORKFLOWS**

Create and Manage workflows allows companies to connect any two systems without writing code or maintaining complex ETL pipelines. Create custom integrations between your on-premise software, cloud applications, databases, etc., in just minutes via an intuitive interface that is easy to use and powerful enough for even the most experienced of users.

2 **PROACTIVELY MONITOR WORKFLOW PERFORMANCE**

Integrate.io offers proactively monitor workflow performance, which means you can easily schedule automated jobs within our platform while receiving alerts when something doesn't go as planned. This frees up your IT staff to handle the projects they need to and not constantly check up on things. Our proactively monitor workflow performance offering saves you valuable time and continually ensures all team members receive timely updates on ongoing situations.



3 EASILY ADAPTABLE FOR USE ACROSS ALL INDUSTRIES

The Data management solution offering from Integrate.io is highly flexible, since its core capabilities are designed to suit any industry while still providing granular control when needed. Our platform allows you to choose among different out-of-box connectors depending on your needs, with each integration working perfectly to better save you time and money.

4 MAINTAIN OTHER SOURCE APIS

Maintain other source APIs that allow companies to connect your on-premise software, cloud applications, databases, etc., via the Integrate.io platform/ With native connectors for almost every major cloud service provider like Amazon, Google, and Microsoft Azure, you will never need to worry about your data again.

5 INTEGRATE.IO API MANAGEMENT

Integrate.io API Management allows companies to fully control and manage all aspects of their APIs, such as who has access to them, what data is allowed in/out, and monitor and troubleshoot problems. This delivers peace of mind that your critical business information will always be available when you need it most.

6 POWERFUL API ACCESS CONTROLS

With Integrate.io's robust API access controls, businesses get a flexible way to manage who has access to APIs along with a full suite of authorization options such as OAuth authentication, so users have peace of mind knowing their sensitive information is safe at all times.

14

INTEGRATE.IO IS YOUR ONE-STOP SOLUTION FOR ALL OF YOUR DATA PIPELINE NEEDS

Our robust management platform allows you to easily manage all of your APIs, data sources, and databases in one place.

Integrate.io's platform provides brands with the ability to integrate and manage data from any source in real-time easily.

For more information on Integrate.io's platform, book a call with one of our professionals at the [link provided here](#). We would love to show how our platform can help increase your business operations.