DCKAP

# The ABCs
# of Data Integration

# Table of Contents

# The ABCs of Data Integration

## Introduction to Data Integration



Businesses generate tons of data every day. They manage enterprise data using a variety of on-premise systems and applications.

Now, despite the fact that businesses produce huge amounts of data every second, that data does not make any sense unless it is managed and integrated properly. As a result of enormous amounts of data accumulating, it can become impossible to manage. This is where data integration comes into play.

Imagine you're running a pizza restaurant with two ovens. One oven cooks delicious pizzas, but it only keeps track of how many you make. The other oven also cooks amazing pizzas, but it only remembers the toppings used. Wouldn't it be cool if you could see the whole picture – how many pizzas were made with each topping?

That's what data integration is like! It's like combining all the information from different ovens(different data sources) into one big cookbook (a central repository like a data warehouse). This cookbook tells you everything – how many pizzas, what toppings, even which oven cooked them!

## What is Data Integration?

Data integration is the process of merging data from different sources into a single, unified view. The resulting merged data is then stored in a designated repository, such as a data warehouse, data lake, or data lakehouse, for easy access and management.

Data integration begins with the process of ingestion and then uses the ETL process to extract, transform, and load data from various sources into a unified view.

DATA SOURCE 1

DATA SOURCE 2  →  DATA INTEGRATION  →  UNIFIED VIEW

DATA SOURCE 3
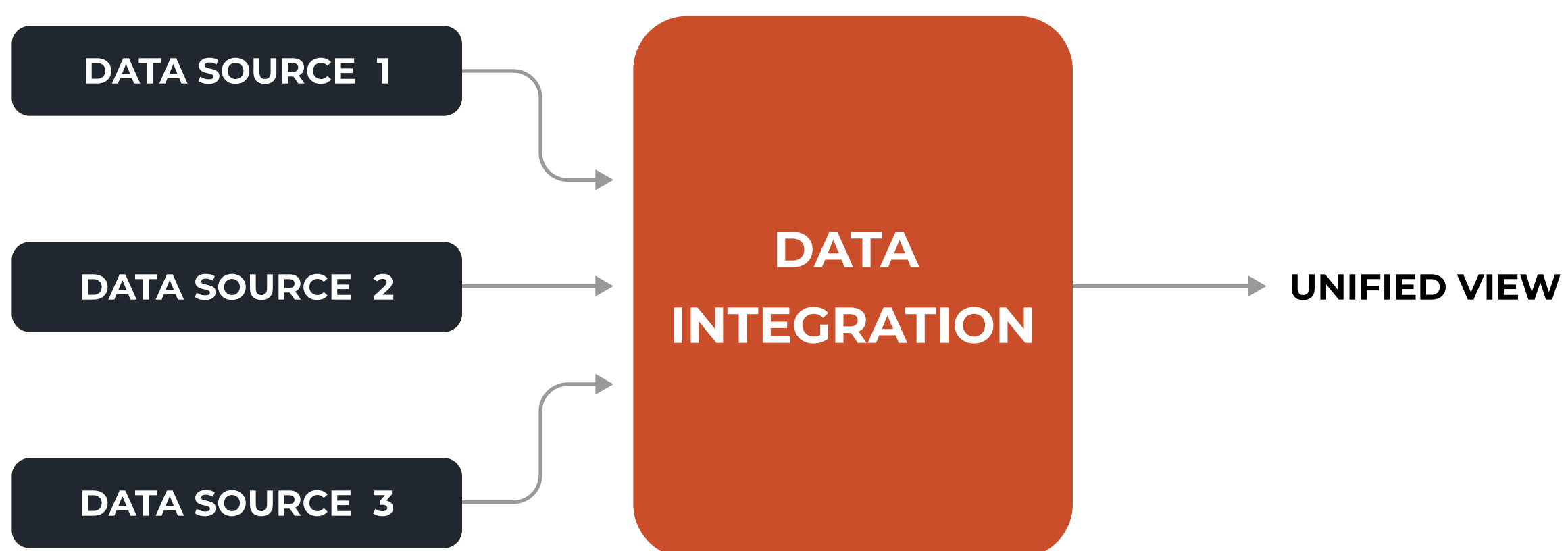
Let's say you're a distributor handling products from various suppliers. Data integration is like the process of collecting information about incoming shipments, supplier details, and sales records. In the ETL process, you extract this data, transform it into a standardized format, and load it into a unified view. Now, you have a comprehensive overview of your inventory, sales, and supplier relationships, making it easier to manage and optimize your distribution operations.

## Why is Data Integration Important?
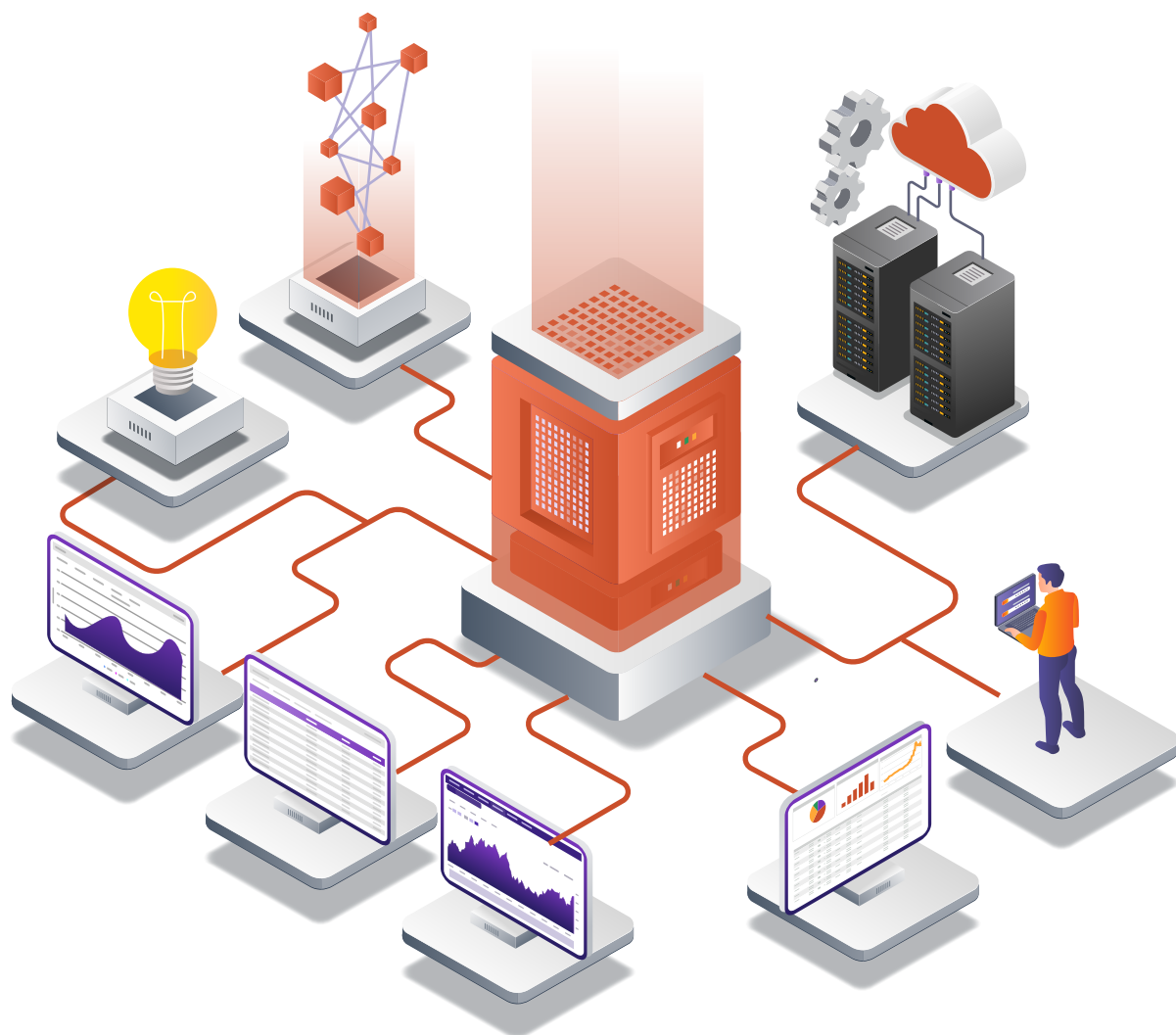


According to Global News Wire, the market for data integration is expected to expand from USD 11.6 billion in 2021 to approximately USD 19.6 billion by 2026, indicating a compound annual growth rate (CAGR) of 11%.

Whether an organization wants to merge databases between partners, or a government agency wants to eliminate data silos, data integration systems can play a crucial role in reducing tedious data manipulation.

Here are a few advantages of using data integration solutions for your business:

1. **Increases effectiveness and conserves time:** With the use of smart techniques like data integration, a company can streamline its data preparation and comprehension, resulting in a faster and easier process. No more manual labor – the system takes care of organizing everything. This frees up more time to analyze the data and make business operations more efficient and competitive.

2. **Reduces errors:** Trying to handle a company's data manually is like trying to juggle a bunch of balls in the dark — it's a difficult task. However, data integration simplifies this process, providing accurate and up-to-date information that reduces the risk of errors.

3. **Help making data-backed decisions:** In business, making informed decisions is crucial. With data integration, companies can make smarter choices by gaining a comprehensive understanding of their data. Gone are the days of relying on incomplete or outdated information – data integration ensures that businesses have all the necessary data at their fingertips. It is the key to unlocking a company's full potential. Just like flipping a switch, it illuminates the inner workings of a business and allows for clear understanding. With data at your fingertips, making critical decisions becomes quicker and more intelligent.

4. **Teamwork gets better:** Collaboration becomes seamless with data integration. Imagine every team member, regardless of their location, having access to the same information. It's like giving everyone a VIP pass to see and utilize the company's data. As a result, teamwork becomes a perfectly orchestrated dance across the entire organization. Together, everyone can achieve greatness.

5. **Delivers more valuable data:** The data integration process enhances the value of a business' data over time, delivering a wealth of valuable insights. By centralizing data into a repository system, any potential issues with its quality are promptly identified and addressed, leading to more accurate and reliable data. This serves as the backbone for in-depth analysis and drives quality results.

# Decoding Data Integration

Data integration is a step-by-step process that brings together information from different places and turns it into a single, usable format.

Here's a simple breakdown of how it typically happens:

**1. Source Identification:** Find where data comes from, like databases or spreadsheets.

**2. Data Extraction:** Pull the data using tools or direct queries.

**3. Mapping Data:** Create a map to match up different terms or codes.

**4. Validation:** Check for errors to keep data accurate.

**5. Transformation:** Organize data into a common, consistent format.

**6. Loading Data:** Put the organized data in a central place for analysis.

**7. Synchronization:** Keep integrated data updated regularly or in real-time.

**8. Security Measures:** Follow rules for handling sensitive data, and add extra security.

**9. Metadata Management:** Keep extra info about the data for context.

**10. Data Access and Analysis:** Use tools to analyze integrated data for decision-making.

# Challenges in Data Integration

## 1. Data integration platform challenges:

Data integration platforms can be tricky to use. Often, they require highly qualified professionals, who can be difficult to find and expensive. When relying on these experts, business analysts may face delays of six months to integrate data from different sources, slowing down the value of data analytics.

## 2. Difficulty in managing data at scale:

It can be challenging to deal with large amounts of data. It is hard for organizations to make high-quality data easily accessible for analytics. As data sources and silos increase, finding the right balance between transferring and copying data becomes challenging, impacting flexibility.

## 3. Issues with data semantics:

Data can become difficult to understand when it contains various versions that convey the same information but appear differently. For instance, dates may be presented as numbers or in the format of month, day, or year. Tools like ETL and master data management help sort out this challenge by changing the way the data looks.

## 4. Tight coupling of data with applications:

Previously, data was tightly connected to particular applications, posing challenges in its adaptability to other purposes. However, a notable transformation is occurring in the separation of applications and data layers, paving the way for increased versatility in harnessing data.

## 5. High costs of data integration infrastructure:

Managing the infrastructure needed for data integration initiatives demands hefty financial investment in both capital and operational aspects. However, by utilizing cloud-based data integration services with a managed-as-a-service approach, the burden of these costs can be significantly reduced.

## 6. Managing multiple delivery styles:

Customers demand a variety of data delivery options, including batch, streaming, and event, all accessible through a single platform. Consequently, businesses are seeking advanced data integration and analysis solutions to optimize results, especially as more parts of their operations generate digital traces.

## Data Integration Technologies

When it comes to transferring data from one system to another, a crucial factor is a data pipeline that fully understands the structure, significance, and path of the data within technical systems. Choosing the most effective data integration methods depends on various factors, such as

- The amount, speed, and diversity of the data being integrated.

- The attributes of the data sources and destinations.

- The timeframe and resources at hand.

Data integration commonly employs various techniques, including:

### 1. Manual Data Integration:

Engineers write code manually to move and manipulate data, tailoring it to meet specific business requirements.

### 2. Application-Driven Data Integration:

Direct links between applications enable the movement and transformation of data triggered by specific events.

### 3. Common Data Storage:

Information is gathered from diverse sources and stored in a centralized data lake or data warehouse, creating a singular data source.

### 4. Data Virtualization:

Different data sources are blended in a virtual database, providing end users with accessible information.

### 5. Middleware Data Integration:

Software is utilized to facilitate the seamless transfer of information between different systems.

# Data Integration tools

Data Integration tools are essential in the Data Integration process, as they facilitate the smooth transfer of data from one location to another. These tools are responsible for cleansing, transforming, and mapping data, making the entire process highly efficient and effective.

There are four main types of Data Integration tools, each with its own unique capabilities.

Let's take a closer look at these types and what they offer:

1. **On-premise data integration tools:** These tools are best suited for merging data from multiple local or on-premise sources. They come equipped with optimized native connectors that enable efficient batch loading from various data sources and can be deployed on a private cloud or local network. Some popular examples include Informatica PowerCenter and Oracle Data Service Integrator.

2. **Open-Source Data Integration Tools:** Offering a cost-effective alternative to proprietary tools, open-source options provide complete control over data in-house. Clover ETL and Talend Open Studio are examples of open-source data integration tools.

3. **Cloud-based data integration tools:** Also known as integration platforms as a service (iPaaS), are invaluable resources for integrating data from various sources into a cloud-based Data Warehouse. Among the notable options in this category are Hevo Data and Xplenty.

4. **Proprietary data integration tools:** tools such as Zapier, are designed for specific business needs, and sold commercially with closed-source code. Despite being ready-to-use and offering comprehensive support, they come at a higher cost and don't permit internal modifications.

To get a clearer picture and deeper understanding, let's explore the distinctions between ETL, ELT, and iPaaS. This way, we can make a well-informed comparison and pick the perfect fit for your needs.

| Aspect | ETL (Extract, Transform, Load) | ELT (Extract, Load, Transform) | iPaaS (Integration Platform as a Service) |
|---|---|---|---|
| **Data Processing Flow** | Extract data from the source, transform it, and then load it into the warehouse | Extract data from the source, load it into the warehouse, and then transform | Cloud-based platform facilitating seamless integration between applications and systems |
| **Transformation Location** | Transformation occurs before loading into the warehouse | Transformation occurs within the data warehouse | Transformation often occurs in real-time or near real-time within the platform itself |
| **Use Case Complexity** | Suited for complex transformations before storage | Suited for less complex transformations, leveraging data warehouse | Designed for connecting SaaS applications and on-premise systems, supporting a variety of integration scenarios |
| **Scalability** | Less scalable as more data requires a larger staging area | Highly scalable as the warehouse scales seamlessly | Highly scalable due to cloud-based nature |

# Data Integration vs What DCKAP Integrator Does

**Similarities and Common Traits**

DCKAP Integrator and Data Integration tools are like superhero helpers for managing data. They both focus on making sure different types of data can work together smoothly.

1  **Integration Focus:** Both DCKAP Integrator and generic Data Integration tools share a primary emphasis on seamlessly integrating data across diverse systems.

2  **Data-Centric Approach:** Both solutions are designed with a central focus on managing and integrating data, regardless of the industry or specific applications.

3  **Versatility:** DCKAP Integrator, with its industry-specific features, and Data Integrator tools, with their broad applicability, showcase versatility in catering to diverse data management needs.

4  **Flexibility:** Both solutions offer flexibility, allowing users to address a range of data integration scenarios, from simple migrations to complex data pipelines.

# Key Differences

The DCKAP Integrator has been specifically designed for distributors, catering to their unique integration needs. Its main focus is to seamlessly integrate ERP systems with vital business applications, ensuring maximum efficiency.

Key Features include:

- **ERP-centric:** Built to connect seamlessly with major distributor ERP systems like Epicor Prophet 21, Oracle NetSuite, DDI Inform, Infor, Distribution One, and more.

- **Pre-built connectors:** Offers a library of pre-built connectors for common distributor applications, reducing development time and costs.

- **Industry expertise:** Backed by DCKAP's deep understanding of distribution industry challenges and best practices.

Data Integration tools specialize in seamlessly combining data from multiple systems. Their capabilities extend across various industries and can handle an array of data management tasks.

The primary focus of these tools is to efficiently integrate and manage data from diverse sources, regardless of industry or application.

It offers a wide range of tools and platforms such as ETL, CDC, data warehousing, and data virtualization.

This ensures flexibility and makes it suitable for a diverse set of data integration scenarios, from straightforward data migrations to intricate data pipelines.

| Feature | DCKAP Integrator | Data Integration Tools |
|---|---|---|
| Industry Focus | Distributors | Generic |
| Primary Focus | ERP integration | Data integration |
| Connectors | Pre-built for distributors | Wide range of options |
| Expertise | Distribution industry | Data management |

## Choosing the Right Tool

The best choice for you depends on your specific needs and industry. Consider the following:

- **Industry:** If you're a distributor, DCKAP Integrator's industry-specific features and expertise can be invaluable.

- **Integration goals:** If your primary goal is to integrate your ERP system with other business applications, DCKAP Integrator is a strong contender.

- **Data management needs:** If you have more general data integration needs, a broader data integration tool might be more suitable.

## DCKAP Integrator

DCKAP Integrator is an iPaaS platform that can help you connect any two systems like ERP, eCommerce, CRM, and more. It is a cloud-based platform that is easy to use and affordable. DCKAP Integrator can help you to automate your workflows and improve your business efficiency.

If you are interested in learning more about how to integrate your systems securely, please Contact Us today. We would be happy to help you get started.

## References

https://appinventiv.com/blog/what-is-data-integration/
#:~:text=Helps%20in%20smarter%20business%20decisions,them%20to%20understand%20data%20
clearly.

https://streamsets.com/learn/data-integration/

https://www.workato.com/the-connector/what-is-data-integration/

DCKAP

www.dckap.com

CONNECT WITH US