(https://blog.appliedai.com/data-integration/)

DATA INTEGRATION: IN-DEPTH GUIDE [2018 UPDATE]

By Francesca Ling (Https://Blog.appliedai.com/Author/Francesca_ling/) | May 17, 2018 (Https://Blog.appliedai.com/Data-Integration/)

To achieve a **unified view of data** that is sourced from different locations and formats, it is necessary to have an established **data integration solution**. This can also include occasions when two companies are merging or in the consolidation of internal applications. Data integration can also be beneficial in the creation of **a better and more comprehensive data warehouse**; ultimately leading to a more accurate and effective analysis.

Data integration is the process of **taking data from many disparate sources and making it usable**. As the number of sources continues to grow, the need for effective data integration also continues to grow in importance. There are a number of **benefits associated with an effective data integration solution**. Some of these include:

- A single and reliable version of truth that is synced and accessible across locations
- An enhanced capacity for analysis, forecasting, and decision making based on accurate data
- A fully comprehensive view of an organization and its customers
- Data availability throughout a business and its stakeholders

Ultimately, data integration lays the foundation for effective Business Intelligence (BI) and the effective decision making it enables.

Understanding 4 Components of Data Integration

Data integration is a term that covers a range of subtopics. A few of the most important categories include:

1. **Data warehousing:** Central sources of data from many different sources.
2. **Data migration:** Moving data between locations, formats, or applications
3. **Enterprise Application Integration (EAI):** Creating interoperability between systems
4. **Master Data Management:** An effort to create one single master reference.

Data Warehousing

Image source: Dremio

Data warehousing is a technology that **aggregates structured data from one or multiple sources** in order to compare and analyze it to achieve greater business intelligence. It is effective for getting a better understanding of the overall performance of a business because it makes a wide range of data available for analysis. It varies from a traditional operational warehouse because it is designed to give a long-term view of data over time.

Owing to the focus on data aggregation instead of transaction volume, a data warehouse is essential when there are analytical needs that require actions 'against' the ongoing performance of an operational database. For example, if there is a need for a complex query on a database, it must enter a fixed state temporarily. With databases that work off transactions, such a state can be difficult to reach – **creating a need for another entity to do the analytical work**; like a data warehouse.

An additional benefit associated with a data warehouse is that it can be part of the final stage of an ETL (Extract, Transform, Load) process; which ultimately means that with the help of an ETL tool, it can analyze data from multiple sources. To learn more about ETL, make sure to see our blog post (https://blog.appliedai.com/etl/) on the topic.

Some benefits associated with data warehouses include:

- Improved Business Intelligence
- Rapid access to data
- Increased system and query performance
- Historical intelligence
- ROI

Some disadvantages associated with data warehousing include:

- Cost of scaling
- Challenges with raw, unstructured, or complex data
- Maintenance costs

Data Migration

Image source: Astera

Data migration is the process of **moving data between locations, formats, or applications**. It is often caused by the introduction of a new system or location for the data. One common cause today is the shift from on-premises to cloud-based storage and applications.

There are a few types of data migration to keep in mind:

- **Storage migration:** Moving data from existing arrays into their more modern counterparts to achieve faster performance, scaling, and data management tasks such as cloning, backup, and snapshots.
- **Cloud migration:** Moving data, applications, and other business elements to the cloud or between clouds. Often also requires a storage migration.
- **Application migration:** Moving an application from one environment or another.

It is important to consider the difference between integration and migration. Integration is the combination of processes that enable data from different sources to be turned into business insights. Data migration is a different process that involves the transfer of data between storage types, formats, architectures, and systems. Another distinction is that integration generally requires collection of data from outside sources, whereas migration often refers to internal movements of data.

Enterprise Application Integration

Image source: Systems Ltd

Enterprise application integration (EAI) is a category of approaches to **achieving interoperability between different systems** that businesses utilize. Specifically, it requires approaching problems related to the modular architecture of the organization. Some key factors that it includes are:

- **Interoperability:** Managing the different languages, operating systems, data formats of components so that they can be connected.
- **Integration:** Creation of a standard process for managing the flow of data between applications and systems to ensure consistency.
- **Robustness, stability, scalability:** Whatever solution is implemented, it needs to be able to adapt quickly and smoothly to changes within a business.

Prior to EAI approaches, integration was generally managed via point-to-point integration; where a unique connector is built for each pair of differing systems or applications that need to communicate. Today, EAI solutions include models of middleware to help with the **centralization and standardization of practices throughout an entire infrastructure**.

To meet the needs of modern businesses, a bus-based EAI, known as Enterprise Service Bus ESB software was developed. This software creates an **architecture that enables differing applications to interact**. Additionally, they set processes, protocols, and rules to enable secure data transfers, route messages between services, and other key tasks.

Master Data Management

Image source: @Infotrellis for Medium

Master data management (MDM) is a discipline that focuses on the **cooperation of business and IT to achieve uniformity, accuracy, stewardship, accountability, and semantic consistency of shared master data assets**. Master data includes the identifiers and attributes that make up the core of the business – such as customers, suppliers, sites, and more.

Continuous data improvement and a well-executed data quality strategy are key for effective ongoing MDM. To **create a single version of truth, it is necessary to harmonize and synchronize multiple data items**. In order to support these efforts and more, change management is essential to ensure the adoption of MDM practices and processes throughout an organization.

There are a few reasons why MDM is gaining momentum among businesses:

- The huge impact it can have. Master data is some of the most important data that an organization has and any errors within it will be felt throughout
- The complexity of today's environment in terms of data volume, availability, and other similar factors
- Compliance and regulatory requirements that have created a need for a deeper visibility and transparency

There are a few challenges associated with implementing a MDM strategy. They include:

- **Complexity:** Data quality can be varied, especially between legacy systems
- **Overlap:** The same data may be duplicated across many systems
- **Governance:** Difficulty in achieving stewardship, ownership, and policies
- **Standards:** Finding agreement on domain values

There are also a few practical challenges related to lack of qualified talent in the discipline, difficulty in executive buy in, and others; as are common in many IT initiatives.

Data Integration & Implementation

When it comes to implementing data integration practices, there are a few things that can be kept in mind in order to ease the process. There are three broad categories of data integration that all carry their own sets of best practices:

1. **Analytic data integration (AnDI):** Where actions are in the context of business intelligence or data warehousing
2. **Operational data integration (OpDI):** Making data available throughout applications and databases
3. **Hybrid data integration (HyDI):** Includes endeavors such as master data management and similar customer and product information management
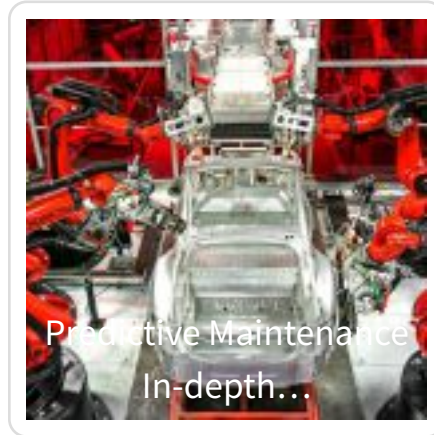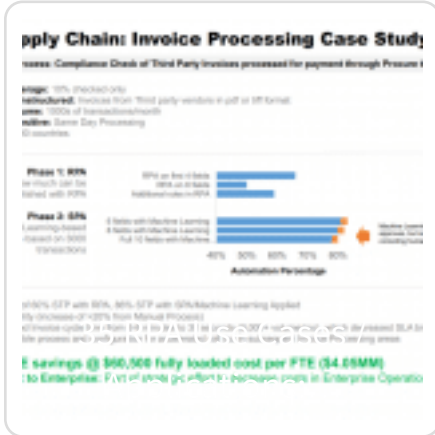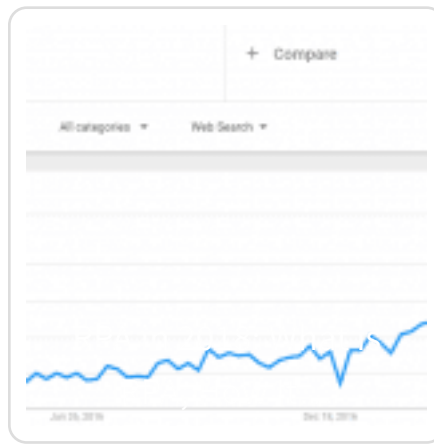
However, generally speaking, there are a few general tips and reminders to keep in mind regardless of which category the task may fall into; particularly when it comes to getting executive approval of any new efforts:

- It should be thought of as **a process that adds value** – similar to the manufacturing of a product where you begin with raw materials (data) and make them into something ultimately more valuable
- There is an **aspect of sustainability**; data integration helps to lower the carbon footprint of data centers by eliminating redundant and erroneous data and virtualizing hardware servers
- Effective data integration requires **collaboration across both technical and business operations** – resulting in a more congruent and aware team across departments
- **Consider data governance from the beginning** – and also consider how proper data integration can be supportive of effective data governance

How we integrate and manage our data is evolving every day. For more on topics like this, be sure to see our blog (https://blog.appliedai.com/), where we cover a wide range of topics related to data, AI, and more. Already know what your business needs, but need an AI vendor to fulfill it? Let our comprehensive vendor list (https://appliedai.com/vendors/1) help you to find just the partner you need.

Featured image source: Excella

Related Posts:

AI Use Cases &

Data Management Tools

Supply Chain: Invoice Processing Case Study

Data Science Consulting &...

Predictive Maintenance In-depth...

Data (Https://Blog.appliedai.com/Category/Data/)
Analytics (Https://Blog.appliedai.com/Tag/Analytics/), Business Intelligence (Https://Blog.appliedai.com/Tag/Business-Intelligence/), Data (Https://Blog.appliedai.com/Tag/Data/), Data Warehousing (Https://Blog.appliedai.com/Tag/Data-Warehousing/), Master Data Management (Https://Blog.appliedai.com/Tag/Master-Data-Management/)

EXPLORE APPLIED AI

USE CASE GUIDES (https://appliedai.com/use-case-guide)

AI STARTUPS & SOLUTION PROVIDERS (https://appliedai.com/vendors)

ABOUT

BLOG (https://blog.appliedai.com)

ABOUT US (https://appliedai.com/about-us)

CONTACT US (https://appliedai.com/contact-us)

PLAN YOUR AI JOURNEY

VENDOR SELECTION (https://appliedai.com/vendor-selection)

OTHER

PRIVACY POLICY (https://appliedai.com/privacy-policy)

(https://www.facebook.com/appliedAI)

(https://twitter.com/appliedaicom)

(https://www.linkedin.com/company/11071351)

applied AI (https://appliedai.com)