

21 Widely Used Data Integration Terms



When speaking with clients about CRM integration needs, it's easy to let some terminology slip in. To help avoid confusion we've compiled a list of commonly used Data Integration terms and provided user friendly definitions. We hope they contribute to a smoother discussion and easier agreement on opportunities and solutions.

Aggregate Data

This is data that's been combined or collated to give an overall view or summary; for instance, when you capture line totals while you're streaming data and then add them together for a final total.

Application

This is short for application program. It's a software program - such as your CRM system, word processor or accounting system - which is responsible for some sort of specific function / business management.

Application Integration

Getting information from one system (application) and sharing it with information from another is application integration. The sharing, or transfer, of data can be across many systems and go both ways, and normally involves live data and real-time transfer. An example is when marketing automation catches a logged ID looking at a report and then automatically triggers the sending of a relevant follow-up email.

Application Program Interface (API)

The API exposes functionality in a program so that external applications can use that functionality. In other words, the API is how programmers communicate with an application, whether to put information in (eg leads into a CRM system), take data out (eg transfer order details), or simply change a setting (eg flag clients for follow up).

Authorization Request

A user initiated request for access to a database that he/she does not currently have access privileges for. Often a requirement for the integration project team.

Bulk Data Transfer

This often comes up in integration projects and refers to moving large amounts of data at one time. When

looking at BDT words like data compression, buffering and blocking may come up – these are just techniques to try to speed up the data transfer and minimize any impact on system operating speeds.

Business Process

This is the idea of taking different inputs and combining them to produce output of value to customers. Business Process is important to clarify during integration projects as business needs and associated data flows can impact the timing and format of details being transferred, as well as where they are placed in the target system.

Communications Protocol

These protocols provide a set of conventions to simplify the communication between different applications. They may be mentioned during discussions, but should not be a discussion point.

Data Extraction

Data extraction is the process of getting data by reading one or more data sources. For instance, querying a third party database to get user names before you clean them and load them into the CRM database. Data extraction is normally the first step in the data integration process.

Data Integration

As with application integration, this involves sharing data; however it normally involves data at rest and moves the data in batches, not real time. An example might be pulling together all the follow up activities for clients so that they can be analyzed to determine next steps, and then update schedules in the CRM system.

Data Integrity

This is key to integration projects; it relates to the quality of data, or the extent to which data is in the correct format. If there is poor data integrity (for

instance multiple ways of spelling a company name or varying date formats) data cleansing will often be required as part of the integration process as there is no point (an potential harm) transferring “dirty” data.

Data Lake

This is a vast collection of different types of data stored in a “raw” form. Due to the relatively haphazard way that data is retained in lakes, the process of retrieving information can be challenging. Data Lakes are often tied to the concept of Big Data, where huge pools of data in varying formats are accessed as needed for analysis and business decisions.

Data Mapping

You will probably hear a lot about data mapping during integration discussions. It generally relates to the process of matching up fields in transferred data (source data) with fields in the destination (target data), including any data transformation required. Mapping is usually made with the application API.

Data Migration

This is where data from a current system is transferred to a new system (such as migration from an old CRM system to a new CRM system). It’s really just another type of integration, often performed by customized programs and mappings, and normally a one-off project. It can involve transformation of data, where data is converted from one format to another, and can be a very complex process.

Data Warehouse

The data warehouse is a collection of databases stored in particular formats to facilitate flexible access across the business. Data Warehouses tend to be more structured than a Data Lake, but also potentially lacking in some of the information that would support decision making processes.

Entity and Entity Relationship (ER) Diagram

An entity refers to a unit of something, an object or idea that does not have a specific name (often a table in the database). An ER Diagram graphically describes the overall structure of a database by describing data entities, their relationships to each other, and the ways in which they are linked.

Legacy System

This term is used to describe a computer system or application program that has been in place for many years, and which continues to be used even though it is outdated and possibly no longer supported. This can present challenges for integration as the older system may not be able to capture the information needed by a newer system, or may not have a documented API, making database access difficult.

Open Database Connectivity (ODBC)

Databases that are ODBC compliant generally have more connection options for integration purposes. ODBC is like a middle man, operating between the different database management systems of different applications; an application can issue ODBC commands and which will be understood by another database management system. ODBC is a bit like a universal translator for applications.

Real time

The speed with which information is integrated or transferred can feel immediate; when this is the case integration is said to occur in real-time.

Redundancy

When multiple copies of the same information are stored it is known as redundancy. There is a cost with duplicate data for both transmission/transfer and updating; the cost can be in time as well as in confusion and inaccuracies. Integration projects will normally try to overcome redundancy as part of the integration process. This issue is related to “Duplication” where client details are duplicated, either because they are captured in several locations at the same time or because they are represented slightly differently and therefore not automatically picked up as the same client. Duplication is also normally handled as part of the integration project.

Scalability

Scalability is the ability to get bigger or smaller in response to demand with minimal impact on cost, performance or functionality. This is will come up when looking at test sizes (are they big enough to be indicative?) or when considering whether an integration solution will be able to manage projected growth and higher. Scalability can be important for accurate cost projections, and ongoing fees.