unifi
Enterprise-class self-service data integration

PARTNER BRIEF

Data Analyst Tools for Faster Results

SIX DEGREES OF DATA INTEGRATION WITH TABLEAU

Six Degrees of Integration

Data integration has become a major feature of the big data landscape. As more and more data sources become available to business analysts and visualization tools become even more powerful, the need to integrate data into the analyst environment has become paramount.

Traditionally the task of integrating data has fallen to a dedicated team of programmers in the IT department. The task is very technical and requires a comprehensive knowledge of data structure and programming. Frequently, two or more data sets can only be integrated by writing a custom software program, often referred to as Java Map Reduction. Other tricks of the trade are used to simplify the data, identify missing attributes or delimiters and derive a structure that can be viewed by the visualization tools.

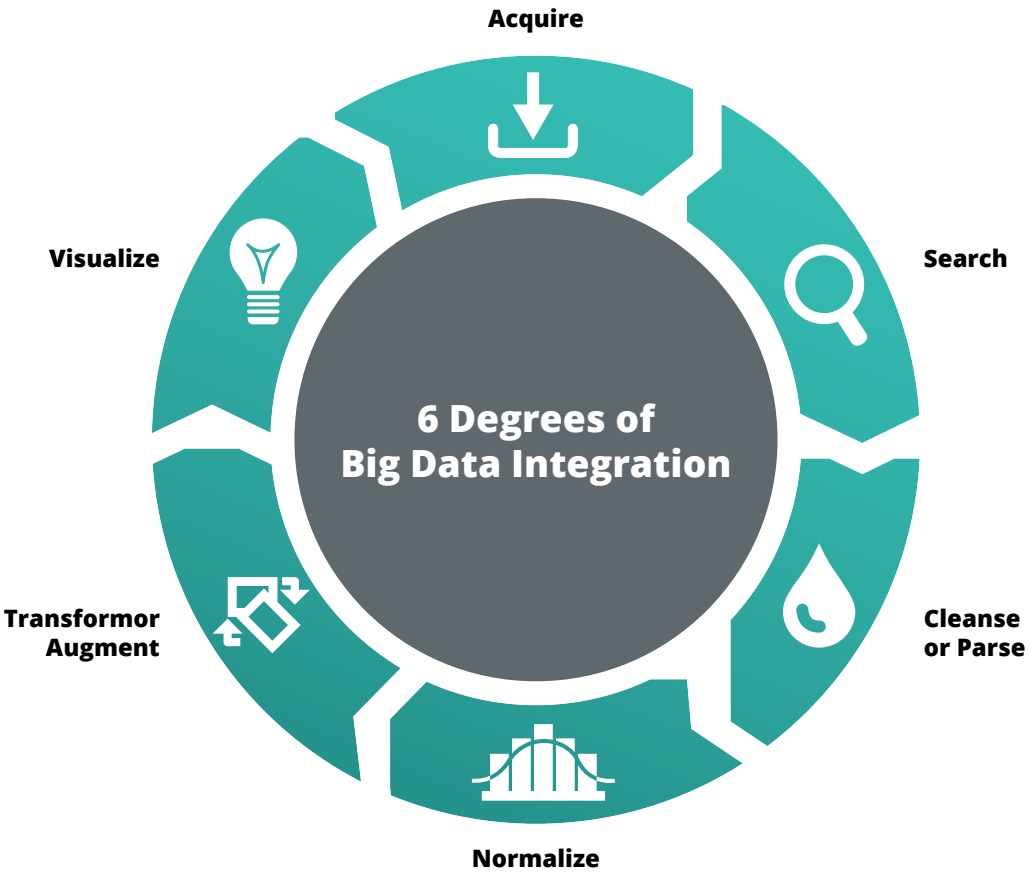
You're Not Speaking My Language!

The challenge for business analysts is the total disconnect between their desired outcome and the development effort required to deliver the result. It really is a language barrier. The business analyst speaks business and forms a hypothesis or wishes to pursue 'what if' scenarios with the data in order to derive a business insight. The programming resource, on the other hand, is looking at the problem from a purely technical perspective and sees the task as one of a structured data challenge.

This process of business case explanation being translated into technical lingo is the language barrier that often leads to frustration and delay. The nature of exploratory analysis is that the situation or hypothesis may change along the way and require further or different data integration. Plus, when the results are delivered by the programmer the analyst must determine if this data is providing the right result. Obvious errors show up in the visualization program and can be easily detected, more subtle errors may not show up and lead to critical business decisions being misinterpreted. This requires a system of checks and balances that takes even longer to deliver and costs the corporation time and money.

The Six Degrees of Integration

Data integration can be broken down in to six distinct parts. Each of these parts forms a critical piece of the end to end task from acquisition to visualization. The task of integration must be approached in a structured linear way from start to finish. Jumping in half way through will undoubtedly result in errors, delay and further frustration. The Six Degrees of Integration can be defined as:



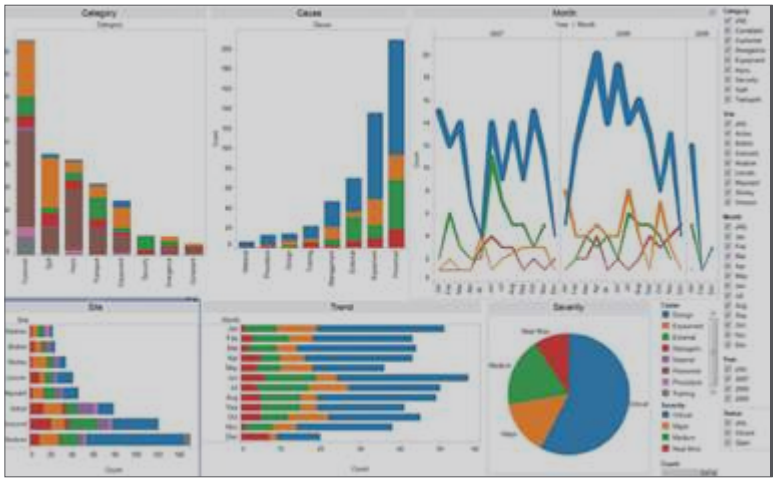
- Acquire the data sources required for analysis
- Search the data sources to find relevant information and objects of interest
- Cleanse and/or enrich the data to remove delimiters, spurious fields or values and add arguments or logic to assist in the structure of the data
- Normalize the data so it can be viewed by the visualization tool or combined with other data
- Transform the data by combining two or more data sets together
- Visualize the results by structuring the data in such a way it can be viewed by the visualization tool.

The acquisition of data falls into two parts; 1) Sourcing the data from either first party business operations such as salesforce.com, OSI, or ecommerce platforms. Or third party providers such as Twitter, Facebook or data brokers such as Datasift. 2) Importing and hosting the data in the analytics environment. This often requires firewall access or DMZ repositories to protect the organization's data integrity.

Searching data source to find objects of interest such as specific customer profiles, geographies, time of day/day of week profiles etc.

Cleansing data is critical to analysis. The challenge with many data sources is that they lack structure. That is, they do not form nicely into tables of rows and columns. This semi-structured or un-structured, data requires specialized knowledge and tools or programming skills to normalize. For data that contain wide and varying values the manual approach to cleansing is extremely time consuming and fraught with potential error.

Normalizing data is the process of providing structure so that the data can be visualized. For most visualization tools this means delivering the interface data formed as rows or columns like a spreadsheet. Without normalizing individual data sets it is practically impossible to join two data sets together based on individual objects of interest. For example: Show me all our Twitter followers who made a purchase on our website in the past month.



Transforming data is where real insights can be derived. This stage allows two or more data sources to be combined prior to visualization. A truly agile integration platform will allow analysts to pursue 'what if' scenarios with their data quickly and easily in order to derive new insights. For example, overlaying weather data on CRM data might determine the optimum weather to promote certain types of product.

Visualization is the final stage of any data analysis but this can only occur when the data has been presented to the visualization tool correctly. For many organizations different analyst or departments choose different visualization tools because of personal preference or features geared toward their day to day tasks. For the organizations that support the analysts with a common set of data sources, this means structuring the output of integration in multiple ways.

Automatic for the People

Borrowing an album title from REM, the task of data integration must be entirely automated before true data democratization can occur within the organization. If the task of integration falls to a dedicated team of programmers supporting potentially hundreds or thousands of analysts this task and the team to which it falls will always be a bottleneck and source of frustration.

Tools that allow the data analyst at whatever level to access the available source and derive insights are critical to realize the full potential of the data. Many data integration tools are available to the organization, some designed to help the IT department, others geared more toward highly technical analysts or data scientists, and still others more suited to the business analysts in a self-help operation. However, the with the exception of Unifi Software , the current set of tools and services available to the business analysts fail to address all six degrees of integration described in this document. At any point that the analyst must contact their IT support team to assist in the integration phase, delays are introduced the task of integration support just added another job ticket. The Unifi Software suite of tools addresses the task of integration end to end. Having being installed on the Hadoop environment hosting the analytics data the analyst-friendly tools allow for total self-help across any and all data types. In many cases Unifi automates the process of integration which not only simplifies the task it delivers critical insights sooner.