White Paper

BUSINESS ANALYTICS AND THE DATA COMPLEXITY MATRIX
Data environments are growing exponentially. IDC reports that compound annual growth in data through 2020 will be almost 50% per year. Not only is there more data, but there are more data sources. According to Ventana Research, 70% of organizations need to integrate more than 6 disparate data sources. At the same time, the value of unlocking that data and using it to make business decisions is also increasing. For the business user, understanding this complex data and unlocking its potential is the key to staying ahead of the competition. For IT organizations, complex data can be the bane of many programs, causing all kinds of trouble in data management and hindering system performance.

What exactly makes data complex? In the context of business analytics, there are two key drivers of data complexity:

- The size of the data
  (is it millions of rows, 100’s of millions, or billion+)?
- The number of disparate data sources
  (or data tables).

These elements drive complexity because the bigger the data, the more effort (cost) needed to query and store it. The more data sources (data tables) the more effort (cost) that is needed to prepare the data for analysis. The data complexity matrix describes data from both of these standpoints. Your data may be Simple, Diversified, Big, or Complex. When considering a Business Analytics program, different approaches are better suited for each data state.

### Simple Data

Simple data consists of smaller data sets that come from a limited number of sources. This data is simple because it does not need data model optimization or significant massaging in order to prepare it for analysis. Small data sets can be queried directly without the intermediate step of creating indexes or aggregations. With only one or two data sources, properly modeling the data relationships is straightforward, lending itself to drag and drop modeling tools. Simple Data also affords one the option of directly querying a live database, rather than an intermediate data analytics store. The characteristics of Simple Data, make it ideal for self-service data visualization tools.
Big Data

In the context of our Data Complexity Matrix, Big data consists of larger data sets (100’s of millions of rows +) that come from a limited number of data sources (tables). This data needs special preparation because of its size. You may need DBAs to work their magic, creating indexes, aggregation tables, clusters, etc. in order to query this data with reasonable performance.

For Big data, this manual data preparation step requires investment of time and resources from an IT team before any analytic output can be generated. In addition the size of the data may require some specialized tools as part of the Business Analytics solution, such as a data warehouse for analytic queries (e.g., Vertica), or a specialized data warehouse appliance (e.g., Netezza). Other options include running business analytics on Big Data using in memory solutions. However, these solutions have an inherent limitations in both size of data and number of concurrent users. Additionally as size of data grows other challenges arise. Either the system costs increase significantly because of the memory required to store the complete data set or one has to give up on data granularity to fit the data set in memory.

Creating a proper Business Analytics solution for Big Data, presents some additional challenges. Third party data warehouse and/or appliance tools create another step in the Business Analytics process. Introducing another moving part has both direct costs for the specific solution and also indirect costs in the overhead to configure, maintain, and support both the new tool and its integration with other pieces of the Business Analytics puzzle. Even with these additional tools in place, the Business Analytics solution is still likely to require data preparation and aggregation. With each
Complex Data

In the context of our Data Complexity Matrix, Complex data consists of larger data sets that come from multiple, disparate data sources. Complex data sets require special attention in both the ETL process and in managing the size of the data. Complex data combines the challenges of both Big and Diversified data. It takes a long time to prepare because of both the modeling challenges and the challenges in indexing and aggregation. Specialized skills and resources are needed throughout the Business Analytics process, turning any project into a cross department, lengthy effort. This manpower cost is amplified each time a change in the data prep or ETL is necessary in order to investigate new analytic paths. Often additional third party tools are required as well (as detailed above).

As a result Complex data often comes with a very high total cost of ownership. License fees for the Business Analytics tool are typically just the tip of the iceberg. License fees for additional data warehouse and data prep tools are additional hard costs in such cases. There are also additional costs created by the overhead and specialized skills needed to integrate and maintain multiple tools from different vendors that may or may not work effectively together. With all of these challenges, agility for the business analyst is diminished and time to insight is longer.

Diversified Data

Diversified data consists of smaller data sets that are derived from multiple data sources (tables). Diversified data requires special attention in the ETL step of the process, in order to ensure correct relational structure between the various data tables. As the number of data sources (and data tables) grows this ETL process becomes more and more complicated, requiring the skills of a DBA to remodel data, creating new schemas or many different views of the data. Another consideration is the frequency with which new data sources will be introduced. In order to keep analysis agile and current, the Business Analytics solution may need to absorb new data sources, forcing iterative reviews of the data model and the ETL. For Diversified data, this ETL step requires investment of time and resources from an IT team before any analytic output can be generated. The ETL process can become so cumbersome that it may necessitate the purchase and use of a specialized ETL tool for automating this part of the data preparation (e.g. Informatica).

Creating a proper Business Analytics solution for Diversified data, presents some additional challenges. Third party data preparation tools create another step in the Business Analytics process. Introducing another moving part has both direct costs for the specific solution and also indirect costs in the overhead to configure, maintain, and support both the new tool and its integration with other pieces of the Business Analytics puzzle.

Such aggregation, business users lose data granularity and risk missing insights derived from that granularity (a traditional OLAP problem). In addition, data aggregation limits the agility of the business analyst, forcing iterative cycles with IT, each time the analysts wants to change the queries or data sets under review.
Choosing a Business Analytics Solution

Each quadrant in the Data Complexity Matrix has its own unique aspects and challenges. When evaluating different business analytics solutions, one must consider the quadrant you reside in today and the one you will reside in tomorrow.

Sisense In-Chip™ technology solves problems for Big Data with its unparalleled data processing speed. Eliminating the need for massaging data and eliminating the need for additional data warehouse tools.

Sisense Single Stack™ technology solves problems for Diversified Data by unifying the business analytics process into a single software solution eliminating much of the ETL process and greatly simplifying what is left.

The combination of In-Chip™ and Single Stack™ provides the 1-2 punch that solves both of the problems associated with Complex Data. Faster processing of larger data sets and unlimited mashing of multiple data sources (tables).

Sisense simplifies business analytics by reducing reliance on scarce, specialized IT skills and empowering individual business users. Business analysts can go from nothing to a fully functional business analytics program using production data in 90 minutes, delivering the fastest time to insight in the market.
READY FOR A TEST DRIVE?

We offer a full Proof-of-Concept on your real data. See how Sisense lets you easily prepare and analyze both big and disparate datasets, from start to finish in 90 minutes.

GET STARTED

http://www.sisense.com/get/test-drive