BRAD ©

Sailing the Data Lake with BRAD©

Motivated by related activities around Cloud, Reporting and Analytics in its other solution verticals and industry trend The Business Agility Group (The BAG), both directly and indirectly through partnerships, has implemented a new division, **BRAD © (Big Rules, Analytics and Data)**, that focuses on consulting and implementation of Big Data-related technology and solutions.

What we do

Firstly, what do we mean by "the Data Lake"? It is a repository for non-traditional sources of data, for example data from:

- Server Logs
- Sentiment (Facebook, Twitter, Linkedin, etc)
- Sensors
- Geolocation
- Clickstreams
- Unstructured documents.

This "Big Data" delivers deep, broad insight by correlating previously separate data sources into a cohesive, comprehensive (new) picture of the subject, allowing lines-of-business to "sail the lake" looking for data previously "hidden" in traditional structures using new analytics applications, called "Data Apps". These Data Apps are what enable iterative analytics -

Current Reality
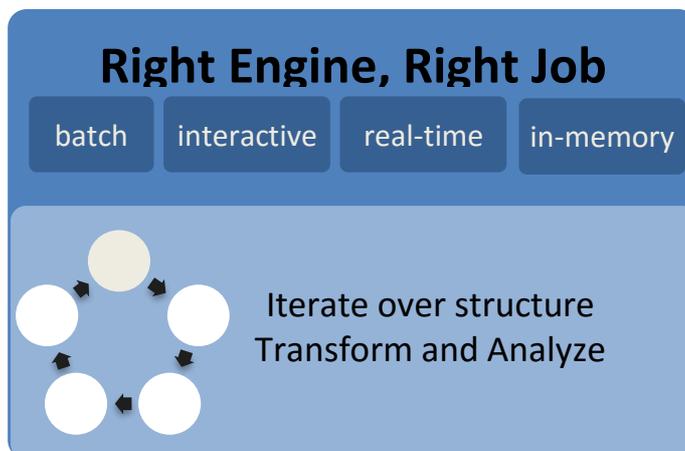
Apply "Schema on Write"
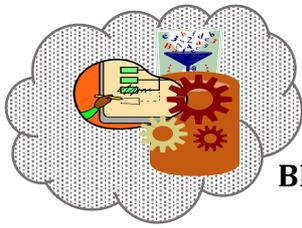
Dependent on IT

Repeatable Process: SQL Only

- Determine list of questions
- Design solutions
- Collect structured data
- Ask questions from list
- Detect additional questions

Augment with a Data Engine

Apply "Schema on Read"

Support range of access patterns to large amounts of data: polymorphic access

Right Engine, Right Job

batch | interactive | real-time | in-memory

Iterate over structure
Transform and Analyze

Concatenation of all of this data in the past would have required organisations to have a large physical storage presence, which typically restricted the function to large, internal, expensive, structured Enterprise Data Warehouses. Now with secure Cloud services available onshore, allowing cost-effective data storage as required, Big Data on-demand is much more accessible, along with the processing power to drive Data Apps and Decision Systems, involving (Big) Rules, Analytics and Visualisation functions.

Decision Systems and Big Data commonly execute in excess of 10M rules and 50M facts and are processed via 2 major paradigms –
- Pre-score populations in batch
- Scalable rules engines for real-time and streaming.

These new real-time functions consume large scale models and data for pre-scoring or real-time decision services and therefore access to more data means better, more accurate decisions, e.g. credit decisions, or fraud detection.

Decision Systems can now scale and generate Big Analytics -
- Models that were previously impractical to build (eg. 'deep models') can now be created on-demand;
- Models with training sets in the tera-/peta-byte range can be created in The Cloud; and
- New models, combined with Big Rules form the basis for business-driven Data Apps.

The resulting Analytical data can then be turned into real-time Visualisations/Views to enable more rapid decision-making based on "live" (concatenated) data. Decisions can then be based on up-to-the-minute data and not on history.
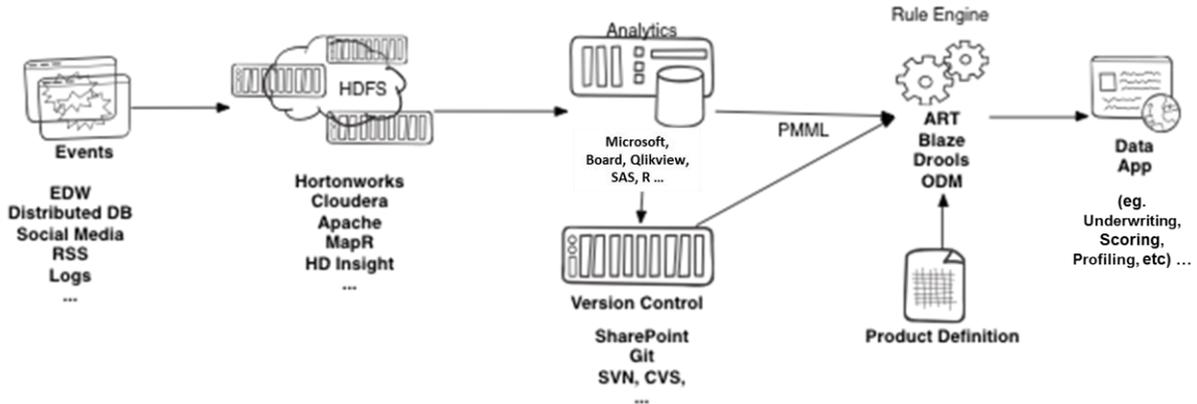
Examples of tools for **BRAD ©** that can be used to create and manage the Data Lake as well as Sail in the Lake are:
- Hadoop (HDFS, YARN);
- Cloudera;
- Hortonworks;
- MapR;
- MS HDInsight;
- Spark;
- Hive;
- MLlib;
- Teradata;
- SAS;
- Yellowfin;
- Tableau;
- Zeppelin;
- Kibana; and
- Many more.
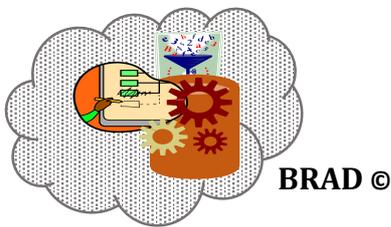
So from the helicopter what does **BRAD ©** look like?:

Note. HDFS = Hadoop Distributed File System; PMML = Predicitive Model Markup Language.

Our background

The BAG is constantly working at creating and implementing innovative methods to add value to its Client relationships and the market in general. It is constantly looking at effective ways to partner with innovation and expand its talent, skills and professional services capabilities. It does all it can to attract the best people and solutions to add value to its clients and prospects alike. This innovation has led to the formation of its **BRAD ©** division, leveraging skills and experience with such platforms as **Apache ™Hadoop®, Cloudera, Hortonworks, MapR, Microsoft HDInsight** and (local) partnerships such as **Vectis Analytics** and **Toustone** in the delivery of an end-to-end capability.

If this is something that is of interest to you, please contact us at – info@tbagility.com, including BRAD in the subject line for any specific additional information.