



Faster, Better, More Economical Data Extraction and Loading

Slaying the Big Data Dragon for Enterprise Analytics

by Sass Babayan, CEO and President WynTec LLC




The old adage of “don’t boil the ocean but boil one cup at a time” is also true for developing Big Data initiatives. The drive behind mining Big Data is to achieve more confident decision making; and better decisions can result in greater operational efficiency, cost reduction and reduced risk.

Currently, many of the big data initiatives are moving giga/tera/peta-bytes of data to massive parallel processing engines and not managing to a cohesive strategy. To just dump all of your unstructured, semi-structured and operational data to a data repository does not lend well for analytics or gathering insight. The Big Data initiative offers a wealth of knowledge in mining the vast real-estate of data. However, building a methodical approach to drain the data lake “one cup at a time” is instrumental to success.

First, understand the make-up of the organization who seeks insight in their data. Many organizations do not have the infrastructure, resources or knowledge to develop a robust data analytics framework. The strategic, tactical and business goals of the organization must be described before designing the business intelligence environment. The overarching analytical engine requires the necessary governance, training, resources and tools to complement the maturity curve of the organization. The customer base can only absorb the data asset, in the Big Data initiative, if the necessary metadata, ontology, definitions, training, and taxonomies are available.

The governance process should be established early in the process to project a road-map and program charter that allows for iterative deliverables of achievable goals. In most cases building your enterprise framework should be done one measure at a time. The success quotient is strengthened as each deliverable is met in small time slots. An enterprise framework can only be built incrementally; the big bang approach has shown little success.

Finding a vendor agnostic partner will help step through the complexities ahead. However, at times, it is challenging to find the right team as the market is growing at a rapid pace and inundated with many vendor specific products. Navigating through this molasses of overlapping technologies is daunting at best.

QUICK	EFFICIENT	POWERFUL	SECURE	INNOVATIVE
<p>More Info:</p> <p>A2B Data™ WWW.A2BDATA.COM 888-814-5554 Sales@a2bdata.com</p> 	 A Product of WynTec			 WynTec, LLC 6565 N. MacArthur Blvd, Suite 225 Irving, TX 75039 info@a2bdata.com

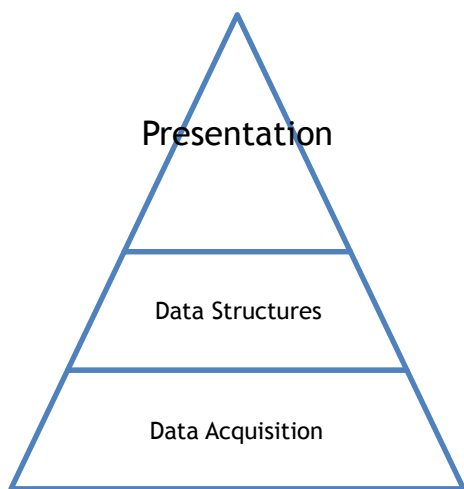
©Copyright 2016 WynTec Consulting, LLC, A2B Data™



Faster, Better, More Economical Data Extraction and Loading

The recommendation is to describe your analytical framework into simple and logical components. Once these components are understood then seek solutions to match the desired need for each component. You may find the right architecture includes a combination of varied products and designs. This is perfect if it meets the business needs.

Building an effective Big Data framework involves building data layers in three steps. Also, every data asset in each of these layers is to be accessible for analytics and reporting. To be successful, the analytical framework must allow for multiple reporting tools to coexist in a plug-and-play mode. The following diagram illustrates the three steps in “slaying the Big Data Dragon”:



The Beginning Layer (Data Acquisition)

To develop Big Data initiatives the first step is to establish a process to collect the data (master data, transaction data and metadata) for the initiative. This can be latent data (run at particular time intervals) or near real-time data.

The data acquisition process must include methods to collect heterogeneous data extracted from any source environment; whether it is RDBMS, flat files, Cloud, Web, Messages, binary or large unstructured data sets. The data acquisition process of managing changed data,



More Info:

A2B Data™
WWW.A2BDATA.COM | 888-814-5554
Sales@a2bdata.com



A Product of WynTec



WynTec, LLC
6565 N. MacArthur Blvd, Suite 225
Irving, TX 75039
info@a2bdata.com

©Copyright 2016 WynTec Consulting, LLC, A2B Data™



Faster, Better, More Economical Data Extraction and Loading

preserving history and transporting the data to a centralized repository is a science that requires great discipline. Progressive organizations are moving to automate the data extraction and load process and free their expert resources to perform more complicated data management and transformation roles. Replacing the mundane and repetitive manual coding practices is necessary to achieve success in a Big Data Initiative. The push to move to automate the data acquisition function is now a viable option. Putting automation in the hands of expert and knowledgeable resources creates an optimum paradigm.

In the Beginning Layer the data should be persisted to ensure it is non-volatile, historic and of sufficient quality. As it is only a complete process if the data is landed, history is preserved and quality is enforced.

The Middle Layer (Data Structures)

The Middle Layer is a collection of information delivered by the data acquisition process, is reorganized and transformed to make that data conformed and suitable for storage and reporting. The Middle Layer includes models and transformation utilities to conform the operational data so it is suitable for analytics.

The Middle Layer can range from structures that represent a central data repository, operational data store, an enterprise data warehouse, departmental data marts, data integration structures or reporting databases. These frameworks include the data structures and necessary data processes to collect and transform the operational data (transactional) to suitable enterprise data. The Middle Layer should be architected to allow for different models to answer different questions. One model can never answer the holistic questions.

There is a strong and obvious push from the industry to move to an agile development and deployment methodology: one that takes the historic, non-volatile and quality staged data and develops SQL based transformations. The agile lifecycle is an excellent method of prototyping the information into useful structures while leveraging the processing power of the Middle Layer. However, the agile lifecycle can only be achieved if the Beginning Layer includes the comprehensive discipline of staging all your data, history of changed data is preserved and the data is of quality.

Many of the Extraction, Transformation and Load (ETL) tools adopt a server configuration of moving the data off the Middle Layer to a separate server to perform the business transformation and then return it back to the Middle Layer. This process is seldom necessary



More Info:

A2B Data™
WWW.A2BDATA.COM | 888-814-5554
Sales@a2bdata.com



A Product of WynTec



WynTec, LLC
6565 N. MacArthur Blvd, Suite 225
Irving, TX 75039
info@a2bdata.com

©Copyright 2016 WynTec Consulting, LLC, A2B Data™



Faster, Better, More Economical Data Extraction and Loading

and costly. Since all the history is already found in the Beginning Layer and the advancement in database technologies permits SQL based transformations. Leveraging the procedural language and leaving the data in the database offers maximum performance and control of managing the data structures. Many organizations are diverting away from the overhead of ETL tools. In many cases these tools are replaced by robust scheduling tools that wrap SQL; leverage procedure logic native to the database or manipulate the ETL tool as an object management tool that executes SQL overrides.

True, that complex transformations and object behavior should be managed by robust transformation processes, but it is also true that most objects in the Middle Layer don't require such complexities. Many enterprise frameworks in the Big Data space observe the 80/20 rule: where 80% requires minimal transformation and only 20% requires complex transformations. The coexistence of SQL based agile transformations (80%) working in parallel with structured approaches (20%) to develop the Middle Layer allows for rapid prototyping and faster delivery.

Agile deployment of the data structures coupled with the disciplines and automation of the staged data enhances the delivery time for analytics. The agile framework does not compromise the architecture of the analytics framework; whether it is hub-and-spoke, federated, dimensional, data mart, MOLAP, ROLAP, federated, data vault or any other analytical framework. The process for agile analytics allows the data to be available and useful to the customer sooner in the SDLC process.

The End Layer (Presentation)

The presentation layer serves the functions of reporting, dashboards, natural language processing, scorecard, self-service, predictive and data mining. This layer is flexible to allow for plug-and-play functionality to the commercially available reporting tools. As often seen, each reporting tool has a preference on the organization of the information. Thus the structures and the meta-model at this layer are developed from both the Middle and Beginning layer but made suitable for consumption of these tools.

The closed loop process of the beginning, middle and end layers completes the active analytical architecture. The goal should be a data acquisition process that lends well in directing suitable data to suitable target architectures and creating a bi-directional handshake between the operational, analytical and unstructured databases. This allows each



More Info:

A2B Data™
WWW.A2BDATA.COM | 888-814-5554
Sales@a2bdata.com



A Product of WynTec



WynTec, LLC
6565 N. MacArthur Blvd, Suite 225
Irving, TX 75039
info@a2bdata.com



Faster, Better, More Economical Data Extraction and Loading

of the data layers to serve their intended primary purpose but distributes the necessary information across the databases:

- The operational system performs the necessary OLTP function and provides ODS data,
- The analytical structure collects the non-volatile, historic and integrated data from the numerous sources and includes heavy number crunching for analytics, and
- The unstructured database performs the text based analytics and transmits the analytical findings to the other necessary layers.

This is a key distinguishing architecture of a solid enterprise framework; build your data layers in three steps but include a handshake and data sharing between each of the layers.

Foot Note

WynTec's focus over the last two decades is centered on establishing enterprise frameworks for organizations that seek to mine their information as a corporate asset. Following the above three steps has led to success in every instance as it offers a proven foundation. As part of this methodology we have invested and partnered with formidable companies in the industry to build the products and services around A2B Data™. This product automates and manages the data acquisition process as described in The Beginning Layer. The data acquisition process is expedited and coordinated by A2B Data™ to deliver value by automating the mundane and repetitive tasks. By separating the source to target knowledge, developing best-practice design patterns, offering variable change data capture methods and metadata driven rules has created a game changing product.

Please visit our website (www.a2bdata.com) to see how our tools and services can help you slay the Big Data Dragon. I welcome you to forward this information to interested parties or contact us for discussion on how this framework can help you with your enterprise analytics.

QUICK

EFFICIENT

POWERFUL

SECURE

INNOVATIVE

More Info:

A2B Data™
WWW.A2BDATA.COM | 888-814-5554
Sales@a2bdata.com



A Product of WynTec



WynTec, LLC
6565 N. MacArthur Blvd, Suite 225
Irving, TX 75039
info@a2bdata.com

©Copyright 2016 WynTec Consulting, LLC, A2B Data™