

# **Emberly Clark Big Data Solutions**

#### **Table of Contents**

1.0 What Is Big Data and Why Is It Important?	1
2.0 Mitigating the Challenges of Big Data Solutions	2
3.0 Product Details	2
Conclusion	4

## 1.0 What Is Big Data and Why Is It Important?

According to Gartner\*, a company has crossed the Big Data threshold when high Volume, Velocity, and/or the Variety of information assets demand new, innovative forms of processing for enhanced decision making, business insights or process optimization.

Based on the above definition, scenarios describing a Big Data problem include Data Volumes that exceed a repository's capacity, excessive Data Processing Times, or a need to simultaneously interpret both Structured and Unstructured data (e.g. Structured data is typically found in relational databases while Unstructured data is comprised of emails, contracts, information related to social media, etc.)

A key competitive advantage garnered from the effective interpretation of Big Data is that, unlike most data warehouse/business intelligence projects, big data benefits are not limited solely to direct decision making.

\*In fact, fewer than half of big data projects focus on direct decision making. Rather, these projects are geared to generate deeper business insights and optimize, automate or even design new processes. The range of uses spans:

• Marketing and sales growth

1

- Operational and financial performance improvement
- Risk and compliance management
- New product and service innovation
- Direct/indirect data monetization

Given the complexity and the potential benefits, it's no surprise that many large organizations seek third party expertise to turn the challenges of Big Data into competitive advantages.

\*Gartner: Answering Big Data's 10 Biggest Vision and Strategy Questions

#### 2.0 Mitigating the Challenges of Big Data Solutions

In order to garner the benefits Big Data, many, if not most companies, elect to move data into Big Data repositories such as Hadoop, Data Lakes, Cloud Analytical Services (e.g. Azure Analysis Services), etc.

While this works, there are a number of drawbacks:

- It takes time, expense, and planning.
- New skills need to be acquired.
- Enterprises may be siloed, where autonomous silos may not elect to move data.
- Large companies may have sister brands, also autonomous.

Taking account of the above challenges, we've developed a "Hub of Hubs" product that integrates all platforms and their repositories without having to move data or learn new technologies.

The benefits are lower costs and greater agility.

The following section describes the product in detail.

### 3.0 Product Details

As described previously, the product integrates all platforms and repositories without having to move data or learn new technologies.

In addition to Hadoop, Data Lakes and Cloud Analytical Services, the product also integrates structured relational repositories (e.g. SQL Server, Oracle, DB2, etc.), NoSQL Databases (e.g. Mongo DB, Cosmos DB, Casandra, etc.), and unstructured repositories (e.g. SharePoint, Documentum, FileNet, Blob Storage, Social Media, Etc.).

In terms of the "3 V's," the solution handles the Variety property of Big Data by federating structured and unstructured data repositories into a "Logical Data Warehouse\*." Once federated, Parallel Computing Principles are applied to improve the Velocity of the data. Finally, in certain cases, optimized scale out strategies may be applied to satisfy the Volume property of the data.

A high-level illustration and a process flow description appear below:



Figure 3.1

3

#### **Process Flow:**

1) The Application supplies a Query Object to the Product.

2) The Product analyzes the Query Object and gets data from the relevant repositories.

3) The Product aggregates the results into a reportable format and returns the aggregate object to the requesting applications.

#### Conclusion

Big Data projects typically involve the movement of data into specialized repositories, which requires time, planning, and expense. Further, it may not be possible to move data due to the autonomy of the units within an enterprise or its sister brands.

Taking account of these challenges, we've developed a "Hub of Hubs" product that integrates all platforms and their repositories without having to move data or learn new technologies.

The net benefits are lower costs and greater agility.

Call us for additional information at 800-314-9503 (select options 1 then 3), or contact us by email at https://www.emberlyclark.com/contact.