Holistic Approach to Big Data #3: Governance for the Big Data Platform

1

We continue to look at the nature and requirements of a big data platform and we will also look at the Data Governance required for this platform.

2

The key aspects of a Big Data platform are:

1. Integration — the point is to have one platform to manage all of the data. There's no point in having separate silos of data, each creating separate silos of insight. From the customer point-of-view (a solution POV) big data has to be bigger than just one technology
2. Analytics — a very important point. We see big data as a viable place to analyze and store data. New technology is not just a pre-processor to get data into a structured data warehouse (DW) for analysis. Deep analytics is a significant area of value add by IBM – and the game has changed – unlike DBs/SQL, the market is asking who gets the better answer and therefore sophistication and accuracy of the analytics matters
3. Visualization — need to bring big data to the users — for example, a spreadsheet metaphor is well known technological approach understood by business users
4. Development — need sophisticated development tools for the engines and across them to enable the market to develop analytic applications
5. Workload optimization — improvements upon open source for efficient processing and storage
6. Security and Governance — many are rushing into big data like the wild west. But there is sensitive data that needs to be protected, retention policies need to be determined — all of the maturity of governance for the structured world can benefit the big data world

3

This slide illustrates the recurring pattern we are starting to see in big data implementations.

There is a classic data warehouse zone,

*

with its reporting and analytics that you are familiar with already.

*

Organizations are expanding their information infrastructure and analytics capabilities to address big data that includes streaming data with real-time connectors to external data sources. This includes ingestion of data from the external sources and real-time analytic processing (RTAP).

*

Data is also delivered to an Hadoop landing environment for longer term storage in the Hadoop Distributed File System (HDFS).

*

Data can be delivered to the data warehouse and to systems for data discovery and visualization.

*

All of this constitutes a *complete* platform for Big Data. It is "Not Only Hadoop."

4

In the new era of computing, the growing complexity of new sources of data and a variety of data types creates challenges. Left unchecked, data is becoming more uncertain and directly impacts the initiatives it supports such as analytics. The role of the IBM Information Integration and Governance platform is to create trusted information and deliver it to consuming analytical, mobile, and enterprise applications.

Because the veracity of data is decreasing as the sources and variety of data grows, 1 in 3 business leaders don't trust the information they use to make decisions. In a big data environment, organizations must improve the level of trust users have in information, ensure consistency of data, and establish safeguards over information. When information is trusted, businesses can optimize outcomes. Developing this level of trust requires an information integration and governance platform that supports a big data strategy.

Many organizations pursue big data analytics to find breakthrough insights that give them a competitive advantage. They are competing on analytics. But organizations can't really compete on analytics alone. The reality is organizations are competing on information and analytics – and in order for the information to be acted upon, it must be trusted. In the new era of computing, companies must ensure that their analytics are driven by trusted data.

5

Let's look more deeply into information governance.

This slide highlights the results of a study conducted by Aberdeen Group. The title of that study is "The Big Data Imperative: Why Information Governance Must Be Addressed Now." Here are some results of that study that highlight the need and the value in information integration and governance for big data.

"Companies with integration tools were able to streamline operational tasks by an average of 15% in the last year."

Seventy-nine percent (79%) of companies with high quality data rated their decision making as a 7 or higher on a scale of 1 to ten."

Organizations with information governance tools (including MDM) not only have more accurate data to start with, but they are improving at almost 3 times the rate of their competitors.

6

Different types of data require different types of minimal governance.

The slide here illustrates the approaches taken by groups of users for different types of data, both structured and unstructured.

7

The platform that helps customers with those use cases you saw on the previous charts is the IBM InfoSphere Information Integration and Governance (IIG) Platform

The job of the InfoSphere platform is to make information more trustworthy and to protect it. We highlight six specific areas of capability:

- Metadata, business glossary, and policy management
- Information integration
- Data quality
- Master data management (MDM)
- Data lifecycle management
- Data privacy and security

IBM's platform is differentiated because

- It's the broadest and deepest integration and governance platform on the market
- With IBM you partner with a vendor that is low risk, understands your industry and has proven success before with clients like you
- IBM is focused on agility and simplifying the portfolio for faster deployments
- IBM is known for its ability to scale and perform to meet the most complex demands. The Information Integration and Governance (IIG) platform is no different and is designed to meet the demands of big data.
- We've taken our knowledge with our thousands of customers and pre-built logic and rules for governance into the platform

Almost two thirds (64%) of organizations that used data lifecycle management tools were able to reduce cost through archiving such applications while still maintaining live access to the application data.

8

True insight into Big Data, as any data, requires a foundation of information confidence.

The fourth "V" in the characteristics of Big Data was Veracity — truthfulness & confidence in the quality of the data.

—-

You have now completed this video.