Opportunities and challenges of driving value from data analytics

ALLWYN
CORPORATION
BUILD. ADAPT. ITERATE.

AUGUST 2020

datascience@allwyncorp.com

www.allwyncorp.com

Executive Summary

From Today to Tomorrow

Data analytics has dominated almost all the industries of the world and data collection has become an important part for any organization. These days every click or scroll you do and every time you open an app, huge amounts of data are being generated and stored for business intelligence and data mining.

Various industries like finance, banking, transportation, manufacturing, e-commerce, and healthcare use this data to make smarter decisions, gain meaningful insights, and predict outcomes. Businesses today are increasingly using data science to uncover patterns, build predictions using data, and employing machine learning and AI techniques.

For example, the Banking industry uses data analytics in credit risk modeling, fraud detection, and evaluate customer lifetime value. Erica, the virtual assistant of Bank of America gets smarter with every transaction made by studying customers' banking habits and suggests relevant financial advice. Finance industries use machine learning algorithms to segment their customers, personalize relationships with them, and increase the profitability of their businesses.

Predictive analytics is another aspect of data science that has become a necessity for transportation and logistics industry. Public and private transportation providers use statistical data analysis to map customer journeys and provide people with personalized experiences during normal and unexpected circumstances. Logistics companies use artificial intelligence to optimize their operations in distribution networks, anticipate demand, and allocate resources accordingly.

Datascience and AI in biomedical and healthcare data is modernizing the healthcare industry by providing solutions to public health. From medical image analysis and drug discovery to personalized medicine, data analytics is revolutionizing patient outcomes. The use of data science and machine learning has revealed that there are solutions to the most difficult problems in different industries and future success of companies rely on their adoption of data centric approaches to discover actionable insights. By automating the analytic process, the time value of unlocking insights can be accelerated to provide rapid forecasting and decision making.

"By 2020, 50% of analytic queries will be generated using search, natural-language processing or voice, or will be auto-generated." – Gartner Analytics Magic Quadrant, 2019"

Top Four Challenges

Removing barriers for better insights

The lifecycle of data, data analytics and data science starts with collecting data from relevant data sources, performing ETL (Extract, load and transform) functions, cleaning and enabling data in a machine-readable format. Once the data is ready, statistical analysis or machine learning algorithms are uses to find patterns, predict outcomes or perform functions using natural language. Since data is at the core of data analytics, it is important to understand the challenges. Here we present the top four data challenges :

Complexity: Data spread across various sources

Merging data from multiple sources has become a challenge for many organizations. According to McAfee, an enterprise with an average of 500 employees can deploy more than 20 applications. Larger enterprises with more than 50,000 employees run more than 700 applications. Unifying the data from these applications is a complicated task that can lead to duplication, inconsistency, discrepancies, and error. With the help of data integration and profiling, accuracy, completeness, and validity of the data can be determined.

Quality: Quality of incoming Data

One of the common data quality issues in the merging process is the duplicate records. Multiple copies of the same record can lead to inaccurate insights as well as computation and storage overuse.

What if the collected data is missing, inconsistent, and not updated? At each collection point, data verification and matching methods needs to be implemented to prevent flawed insights and biased results.

Volume: Volume of data available

To find relationships and correlations, a successful machine learning algorithm depends on large volumes of data. Data collected from multiple sources and through multiple time frames are essential to create machine learning models during training, validation and deployment phases. More data does not necessarily means gathering more records but can mean adding more features to the existing data from different sources that can improve the algorithm.

### Algorithm: Conscious effort to remove confirmation bias from approach

The major advantage of AI over humans is the ability to garner insights into an algorithm's decision-making process (using explainable AI). Further more, algorithms can be analyzed for biases and its outcomes verified for unfair advantages to protected classes. Although, AI on the onset can be viewed as perpetuating human biases, it offers better insights into the data and decision-making process.

Solution Approach

Taking it to the next level

Advances in AI and machine learning has given rise to the growing importance of data analytics and therefore, data itself. Unless you have established the pre-requisite steps of data collection, data storage and data preparation, it is impossible to move to data science process.

At Allwyn, we believe that the journey to improved operations and decision-making starts with establishing a good data strategy and establishing the tools and processes required to easily analyze your enterprise data. This involves starting with your data discovery and data collection, organizing the data in a data warehouse or a data lake, and finally using machine learning to perform deep data analytics to enhance productivity, launch new business models or establish a strong competitive advantage. We have an established data life cycle process that starts with the data discovery and ends with reaching business outcomes through data analysis, machine learning and AI. We employ a two-phased approach of data transformation and operational transformation as shown below.
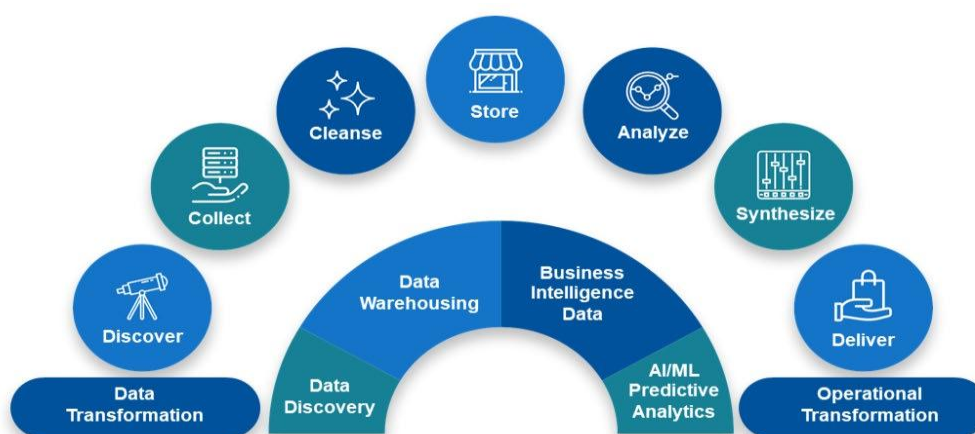
The first phase of data transformation, our goal would be to design, build and maintain an enterprise data warehouse or a data lake. This would help in making the most of an organization's valuable data assets, break down data silos, and create a data maturity model that helps accelerate providing accurate and near-real-time data for the next phase. During this phase, we also establish data governance that focuses on the privacy and security of the data.

The second phase focuses on data analytics – predictive, prescriptive or diagnostic analytics that can help the various departments of your business with actionable insights. In this phase we also help with rapid prototyping and experimenting with advanced analytics such as machine learning and AI. We help you adopt machine learning into your data analytics to help with your product innovation, and offering you a competitive edge in the marketplace.

Our data management strategy provides an enterprise with quick and complete access to the data and the analytics it needs through four steps.

Collect/
Ingest

Store

Process/
Analyze

Predict/
Visualize
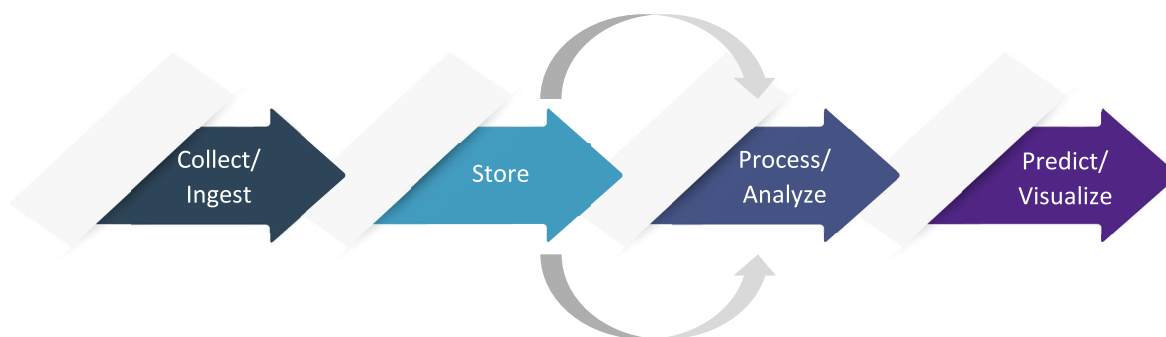
Our four-step solution for Enterprise Data Management is elaborated below.

1. ### Collect: Ingestion/Data Prep/Data Quality/Transformation

In this step we access and analyze both real-time and stationary data to determine reliably, data quality and extract, transform and blend data from multiple sources. We then map and prepare the data to load into a target Data Lake. It is important to identify all your data sources and data streams to determine your acquisition of data and establish the frequency of your batch process. This also involves establishing your infrastructure to help with the high volume of data streams and supporting a distributed environment.

2. ### Store

We use a scalable, reliable (a Cloud Based Data Lake) comprised of a variety of data repositories for both structured and unstructured formats to ensure reliable data storage. In this step, you cleanse, categorize, and store the data as per your business functions. For example, you can establish separate functional areas for sales, marketing, finance and procurement related data. This will help you with establishing functional unit while identifying the need for data integrators across functions.

3. ### Process/Analyze

Once the data is identified, organized and stored, your data is ready for data analysis, building machine learning models or statistical analysis. Data analysts or data scientists can run multiple queries or develop algorithms to analyze trends, discover business intelligence, and present outcomes to make smart decisions.

4. ### Visualize

It is essential that the output of the data analaysis be presented in a visual dashboard to provide meaningful answers to key questions driving business decisions. Here, we not only provide insightful visual dashboards, but provide search driven "Google like" products with Natural language processing capabilities to provide answers to easy-to-understand presentations for all levels of data users and the public. With products like Thought Spot, users can type a simple Google-like search in natural language to instantly analyze billions of rows of data. Users can also easily converse with data using casual, everyday language and get precise answers to questions in an instant.

Real Insights

Getting your data strategy in place is the first step to start with data analytics, data science and the AI journey. As the marketplace continues to rattle business models, adopting newer data analytics tools such as machine learning can help you not only stay ahead of competition but also continue to operate your business successfully in uncertain times. This can lead to a data-driven value cycle that can help pave the way for accomplishing the transformational change that is essential to become an AI-enabled organization.

About Allwyn Corporation

Allwyn Corporation (www.allwyncorp.com) is a forward thinking, innovative software solutions company, headquartered in the Metropolitan Washington DC area. Allwyn was founded in 2003 with a mission to help organizations address complex technology challenges by providing industry-leading tools, technologies, seasoned professionals, and proven methodologies. We are proud to be certified for ISO 9001 (Quality), ISO 27001 (Security), and ISO 20000 (Service Delivery).
With a team of ~200 professionals, Allwyn delivers high-quality services to a wide range of clients in the public and private sector.

Allwyn has been providing leading-edge IT professional services to various government agencies through the GSA MAS Schedule. We are also on the FAA eFAST, GSA OASIS+, and GSA STARS III contract vehicles.

Allwyn has experience with implementing Artificial Intelligence and Machine Learning solutions and Modernizing Applications using Low Code Technologies. Our relationships with AWS, Appian, ServiceNow, Microsoft, Databricks, Informatica, Salesforce, etc. strengthen our ability to support our customers in their Digital Transformation journey. We are already supporting several of our customers in the public as well as commercial sector with their cloud adoption strategies and Artificial Intelligence and Machine Learning implementations. For additional information on Allwyn's full range of services, please visit our website at www.allwyncorp.com.