Proposal to Enhance Big Data Infrastructure for

Mohammed Arslaan Kola

Contents

About and Business Context

Coca-Cola, the principal figurehead of the soft drink market rose to popularity with it's signature range of non-alcoholic beverages like Coca-Cola, Fanta and Sprite, to name a few. From pouring and cherishing their first drink on the 8th of May, 1886 by Dr John Pemberton at Jacobs' Pharmacy in downtown Atlanta (Coca-Cola n.d.), the company has been motivated to spread and partake in moments of joy, positivity and rejuvenation. With a market value of 179.3 Billion Dollars (Prafull, 2018) their vision is to design brands and manufacture an array of drinks that people feel fresh upon consumption, and to make a difference in their communities (Cuofano, 2023). At the core of Coca-Cola, lie leadership, collaboration, integrity, accountability, passion, diversity, and quality.

The decentralised organisation operates in over 200 countries and territories, allowing them to respond and adapt to the local markets and consumer trends locally.

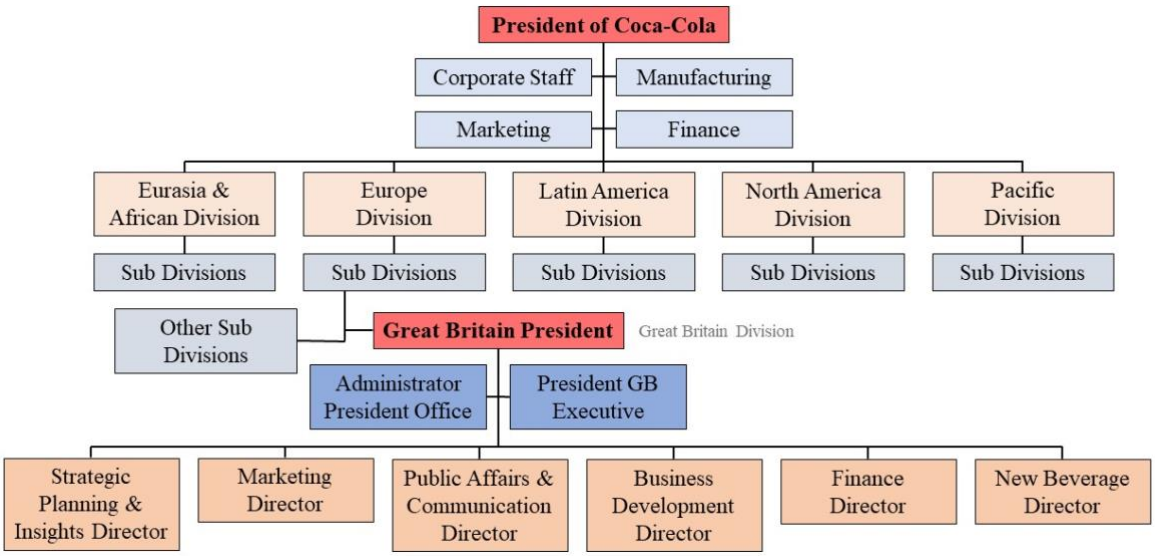

Figure 1: Simplified Organisational Structure of Coca-Cola

Figure 1 shows the simplified structure of the organisation in its hierarchy. The structure, though drilled down to the Great Britain Division, is implemented similarly at all division levels. Aforesaid hierarchy provides for an effortless functioning of the company with the operational heads in direct correspondence with the country president. This outlook enables the company to cater to their target customers with seamless communication, track the efficacy of marketing, mounting to brand recognition and loyalty. Appendix 1 shows the SWOT analysis on the company to understand their position in the market. With the long-term outcome of sustainable growth, the organisation continues to venture into new terrains of marketing; by investing in its existing brands, people and resources while simultaneously boosting efficiency in terms of outcomes as well as finances. In order to meet these milestones, Coca-Cola is entering into digital transformation and leveraging enormous data to perceive consumer behaviour and preferences.

To this effect, Coca-Cola has invested in digital marketing, consumer engagement across social media platforms, the employment of data analysis in launching personalised marketing campaigns and an e-commerce strategy to expand its

online presence thereby increasing sales. The use of big data has enabled Coca-Cola to track and analyse customer preferences and purchase trends, permitting the organisation to launch targeted campaigns that are more meaningful to the intended audience. By the power vested in big data, Coca-Cola has managed to remain at the forefront of the exceedingly competitive beverage industry, as the organisation is constantly evolving to the trends and demands of their customers.

The rapid proliferation of digital devices and new methods for quickly collecting, analysing, and using massive amounts of data, or "Big Data," have completely changed marketing in every industry (Montogomery, 2017). As Joe Kaeser (Joe, 2018) rightly said, “Data is the 21st Century oil”. For a beverage industry, big data is significant in analysing customer behaviour, enhancing the demand supply chains and conceiving new strategies and products. Data analysis from a range of platforms such as social media, online sales, point-of-sales systems gives an understanding of the consumer demographics, purchasing trends and product predilection. These insights are employed in the formulation of effective marketing, advertising as well as product development. Big data also empowers decision making furnishing the organisation with a trove of information regarding purchases, trends and the influencing factors. Appendix 2 shows the different types of data collected by the company for its analysis. By analyzing large amounts of data, Coca-Cola can identify patterns and insights that can help them improve their marketing and sales efforts, optimize their supply chain and logistics, and make more informed decisions about product development and innovation. Coca-Cola's distinct approach to gathering and utilising data expands beyond organisational level, well into the departmental levels with notable investments into the groundwork and technology to simplify acquisition, storage and analysis of big data. This employs the usage of advanced analytics tools, cloud-based storage, and data lakes.

Various units of Coca-Cola have individually stationed big data gathering and value-adding action plans at the departmental levels. The market division, for instance, uses big data to track consumer preferences, which enhances the advertising campaigns and initiatives are specifically targeted. The sales departments may utilise big data to improve sales strategies including the prediction of prime sales prospects and demands, thus meeting the consumer demands successfully by the efficacious inventory management. The supply chain department may deploy big data to better the logistics and product delivery. Data is analysed from transporters, warehouses and retail outlets, gridlocks may be identified and a lane for optimal efficiency may be devised. The research and development division tracks trends and the opportunities for the production of new goods that suit the growing customer needs.

A data administrator at Coca-Cola plays a pivotal role in monitoring the storage and analysis of the organisation’s big data. The team is responsible for accurate collection, storage and security of data while enabling hassle free access to authorised users. This involves the designing and execution of databases, data warehouses that are equipped to handle the plethora of incoming structured or unstructured data. Additionally, the data administrator works in close coordination with the data analysts and scientists ensuring the impeccable processing and furnishing of quality data. They are also in charge of identifying, handling and resolving issues that may arise. In coordination with the IT team, they would

ensure the availability of tools and infrastructure required for data storage and analysis. As a data administrator at Coca-Cola, the role encompasses collaboration with sub-teams within the company to assess the data needs and provide acceptable solutions. For instance, to coordinate with the marketing team to launch personalised campaigns, with the supply chain team to hone the inventory and logistics, or with the finance team to boost financial outcomes through data powered strategies.

Overall, the role as a data administrator is crucial in empowering Coca-Cola to leverage its big data, enabling the decision makers to take informed decisions thus gaining a competitive advantage in the market.

Current Data and Metrics used by Coca-Cola

Widely accessible and cost-effective technologies are available for the storage and analysis of data. However, for companies like Coca-Cola that are utilising data in innovative ways, leveraging it to obtain precise, reliable business strategies that enable decision-makers and to analyse outputs, business models and customer satisfactions. New trends that evoke prompt decisions may serve as blueprints for a fundamental change in research, invention, and company marketing (Alsghaier, 2017). However, according to a report from the Harvard Business Review, only 3% of the data quality scores in the study were assessed as "acceptable," showing that data quality is much lower than most firms believe (Panoho, 2019).

The concept of big data has varied definitions courtesy of various scholars and organisations. However, going by Oracle, big data is characterised in terms of the four V's: Volume, Velocity, Variety, and Veracity (Cackett, 2013). Additionally, there are "Fourteen Vs and a C Defined" features at the centre of every big data research in order to successfully manage and exploit large data as shown in Appendix 3. These qualities provide researchers and practitioners a research arena allowing them to manage big data seamlessly. Appendix 4 provides a brief overview and definitions of the V's in context of Big data to understand the terminologies used better.

Coca-Cola collects a range of structured, semi-structured and unstructured data from various sources including sales report, financial reports, social media, surveys, news, reviews etc to obtain values to strengthen the company and meet the set goals. Here is a closer look on the data collected by Coca-Cola:

- **Structured Data:** This includes organised data in a suggestible pattern, such as financial, sales, inventory data, etc. For example, Coca-Cola collects data on the sales, consumer demographics, purchasing habits, and product usage. This information is stored in the databases and can are evaluated using traditional data processing techniques.

- Semi-Structured Data: This encompasses data that has some structure but is not as suggestible as structured data. For instance, the customer feedback, customer surveys or reviews. This information is typically analysed deploying specialised tools that are able to extract insights from the data, despite the unpredictable format.

- Unstructured Data: This entails data that is dismally predictable in format and necessitates sophisticated techniques for insight extraction. Unstructured data

includes images, news articles, customer reviews, etc. This data requires sophisticated tools like natural language processing (NLP) and image recognition, in order to extract insights.Table 1 as shown in the next page depicts important data stored and analysed by the organisation. The table considers the data for 1,00,000 customers and 10,000 suppliers. Not all data gathered by the company are presented in the report. In light of the four V's of big data discussed previously, Coca-Cola is employing big data to gather insights into consumer preferences, behaviours, and buying patterns, for which the organisation gathers an immense amount of data from different sources such as customer data, sales data, and inventory data. Due to the vastness of data collected, a strong infrastructure is mandated.

To maximise the potential of big data, Coca-Cola is capturing a wide variety of data types, including structured data, semi-structured data, and unstructured data. This provides an absolute, holistic view of the consumers and operations, while allowing the identification of market trends and new opportunities. A more diverse range of data enables more reliable decision making and thus enabling the organisation to stay ahead of competition. Coca-Cola ensures the collection of accurate, veracious data which is fundamental to decision making, based on reliable insights. The data generation and collection happen simultaneously in real-time at a high velocity. This mandates the company to be equipped to process and analyse the data simultaneously, in near real-time. This coordination allows the company to take swift actions to respond to the dynamic customer demands and trends.

A critical evaluation of in-depth analysis of the data being collected and analysed by Coca-Cola's department will conclude if it is sufficient to meet its strategic goals. The evaluation aims to identify gaps or shortcomings in the current data capturing and analysis processes and to propose possible solutions to overcome these challenges. Marr's SMART strategy dashboard has been used to carry out this review. Bernand Marr (Marr, 2015) proposed the Marr's SMART strategy which is a strategic framework used to assess and evaluate an organization's ability to meet its objectives. The framework consists of five elements, comprising the acronym- SMART: Strategy, Measurable, Analytics, Reporting, and Transform. The SMART dashboard is a useful tool for evaluating a company’s gathering and analysis processes of data, against its strategic objectives.

It is imperative to grasp the strategy to determine the outcomes of the data metrics that Coca-Cola is gathering and analysing. This determines if the metrics need improvement with respect to adequacy of measurements, depending on its applicability and necessity of the organisational plan. Appendix 5 shows the SMART strategy board compiled for Coca-Cola. For a smooth running of SMART business, it is imperative that the semi-structured and unstructured data be combined from both external and internal sources. To enhance credibility of the process, it is important to analyse both internal and external data jointly, either by storing the data together or connecting via APIs based on data source's reliability, consistency, response time, and internal usage. This analysis answers the questions listed in the SMART strategy panel. The data listed can be further categorised based on the source of generation and structure. For instance, financial data is regarded as internal structured data whereas the customer purchasing trends are internal unstructured data.

Data	Structure	Storage	Velocity	Priority	Volume/Customer	Total Data Volume
Sales and Distribution Data						
Sales Volume	Structured	SQL Table	Daily	High	1GB	100 TB
Distribution Channels	Structured	SQL Table	Weekly	Medium	500 MB	50 TB
Pricing	Structured	SQL Table	Daily	High	100 MB	10 TB
Sales Patterns	Semi-Structured	NoSQL	Weekly	High	500 MB	50 TB
Customer Data						
Age	Structured	SQL Table	Weekly	Medium	100 MB	10 TB
Gender	Structured	SQL Table	Monthly	Low	100 MB	10 TB
Location	Structured	SQL Table	Daily	High	500 MB	50 TB
Purchasing Pattern	Unstructured	Data Lake	Hourly	High	1 GB	100 TB
Occupation	Structured	SQL Table	Monthly	Low	100 MB	10 TB
Reviews	Semi-Structured	NoSQL	Hourly	High	100 MB	50 TB
Complaints	Unstructured	Data Lake	Hourly	High	500 MB	50 TB
Supply Chain Data						
Supplier Details	Semi-Structured	NoSQL	Monthly	Low	1 GB	10 TB
Quality of Materials	Structured	SQL Table	Monthly	Low	100 MB	1 TB
Raw Material Cost	Structured	SQL Table	Monthly	Low	100 MB	1 TB
Raw Material Availability	Structured	SQL Table	Monthly	Low	100 MB	1 TB
Inventory Availability	Semi-Structured	NoSQL	Daily	Medium	500 MB	5 TB
Logistic Information	Unstructured	Data Lake	Daily	Medium	1 GB	10 TB
Logistic cost	Structured	SQL Table	Weekly	Medium	100 MB	1 TB
Shipping Time	Structured	SQL Table	Monthly	High	100 MB	1 TB
Distributor Details	Unstructured	Data Lake	Monthly	Medium	1 GB	10 TB
Distributor Price	Structured	SQL Table	Monthly	High	500 MB	5 TB
Marketing Data						
Advertising Cost	Structured	SQL Table	Weekly	High	500 MB	5 TB
Brand Reach	Structured	SQL Table	Hourly	High	500 MB	5 TB
Social Media likes	Structured	SQL Table	Hourly	Medium	100 MB	1 TB
Social Media followers	Structured	SQL Table	Daily	Medium	100 MB	1 TB
Customer Loyalty	Semi-Structured	NoSQL	Monthly	Medium	1 GB	100 TB
Coupons	Unstructured	Data Lake	Monthly	Low	500 MB	50 TB
Social Media Surveys	Unstructured	Data Lake	Monthly	Low	500 MB	50 TB
Financial Data						
Revenue	Structured	SQL Table	Daily	High	1 GB	100 TB
Profits	Structured	SQL Table	Monthly	High	1 GB	100 TB
Expenditure	Structured	SQL Table	Daily	High	1 GB	100 TB
Operational Cost	Structured	SQL Table	Weekly	Medium	500 MB	50 TB
Cash Flow	Structured	SQL Table	Weekly	Medium	1 GB	100 TB

Table 1: Big Data within Coca-Cola

The purchasing trends may also be categorised as internal semi-structure data. Data from external sources like the demographic details of the consumer is regarded as external structured data, while information like social-media surveys, comments, news are regarded as external unstructured data.

Data gathered is phenomenal in determining the necessary metrics to analyse the performance of various departments, and the company by large. For instance, consider the data curated from customer feedback and reviews as shown in Table 1. This data can be utilised in the calculation of metrics of customer satisfaction with the product. The data collected across social media assists the marketing team in creating brand awareness metrics, which build into strategizing the best action plan to increase reach. This data can then be used by various departments to create required metrics and derive significant results.

As Bernand Marr (Marr, 2015) said "Data and analytics go hand in hand". Coca-Cola has made significant investments in R&D, particularly in big data, to put large amounts of collected data worldwide into better use. Through this endeavour, the company has gained a deeper understanding of regional customer preferences for healthier products as well as flavour trends among consumers (Ulunma ,2020). This venture has provided the company with a profound knowledge of local customer demands in terms of health or trends. For example, millions of people leave reviews, rate products, make recommendations, and express their opinions about businesses, products, and services. These expressions guide the decision-making process backed up with huge data.

For a brand like Coca-Cola, an appropriate analysis is not the end point of the process. As the largest beverage company on the global scale, reporting the precise insights to the stakeholders is crucial. To achieve this, Coca-Cola utilises the generation of interactive dashboards to dispense such information and insights gained through the analysis. Figure 2 shows one such example of a sentimental analysis done by the marketing team to present to its stakeholders (Meg, 2019).

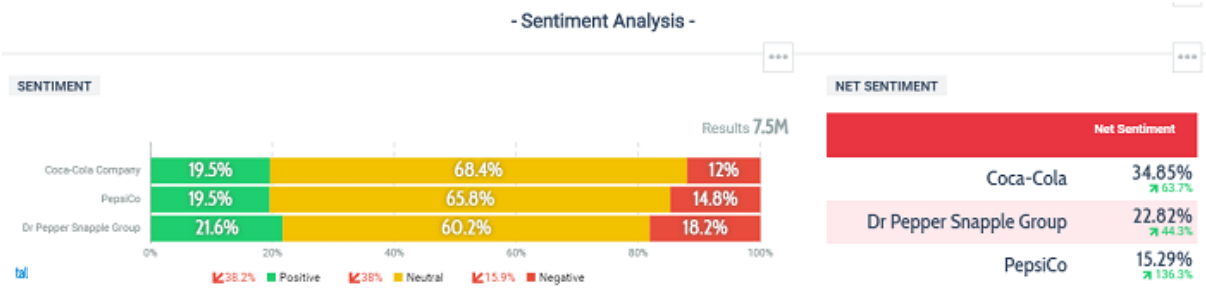

Figure 2: Sentimental Analyses Dashboard

As Daniel (Bumblauskas, 2017) stated, "Knowledge is the process of connecting the stuff of the mind and the stuff of the world. It is not a recorded thing (data, information), or at least, it is not just that. Knowledge is a form of action". Companies rarely benefit from data without analysis or knowledge without application. Figure 3 provides a process flow used by Coca-Cola for achieving the targets set using Big-Data. Insights are drawn through data gathered which is utilised in the generation of strategies which translates to organisational goals being met. The application of insights captured from the big data into actionable steps and plans makes Coca-Cola wholly utilise the big data analysis.

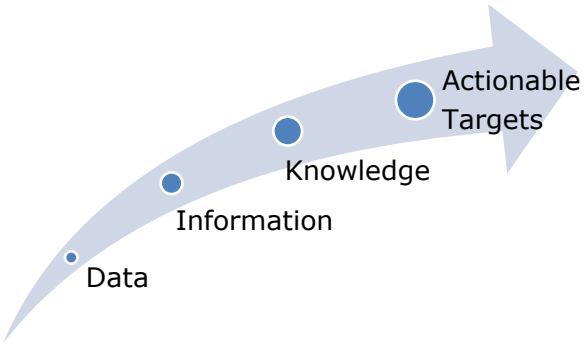

Figure 3: Process flow to convert data to achievable targets

Overall, Coca-Cola's current big data collection and analysis efforts align well with Marr's SMART model, and it can be inferred that the company is on track to meet its strategic objectives. However, continuous monitoring and assessment of big data is essential to ensure relevance and efficacy with time. There is room for improvement in terms of data quality and processing speed.

Current Infrastructure and Technology for Big Data

Coca-Cola's current infrastructure meets its big data requirements and is appropriate for the business's current scale. However, because of its long-term strategies of extensive global reach and expanding business, the current infrastructure's scalability may cause limitations. Concerns have also been expressed about the potential loss of value as a result of insufficient data collection, a lack of real-time analytical processing, and rising on-premises server and technology maintenance costs.

Coca-Cola's big data management process incorporates a number of business processes as well as IT/data warehousing technologies. These processes and technologies work together to collect, store, process, and analyse massive amounts of data generated from various sources, such as sales transactions, customer interactions, supply chain operations, marketing campaigns, and financial data.

The first step in managing big data at Coca-Cola is data collection. Internal systems such as enterprise resource planning (ERP) systems, customer relationship management (CRM) systems, and supply chain management (SCM) systems are used to collect data. Outside sources of data include social media platforms, market research firms, and government data agencies.

Data is collected and stored in a centralised repository for further analysis. To ingest data into the big data infrastructure, the data ingestion layer employs tools such as Apache Kafka and Apache NiFi. The data processing and analytics layer then processes and analyses the data in real-time and batch processing, using tools such as Apache Spark and Apache Storm. Coca-Cola stores and manages this data at multiple data centres in Georgia using data warehousing technologies (Schwartz, 2018). For such large amounts of data, relational databases, NoSQL databases, and Hadoop-based systems are used, which provide scalable and flexible storage solutions. Figure 4 shows the current infrastructure implemented by the organisation and the software’s used at different layers.

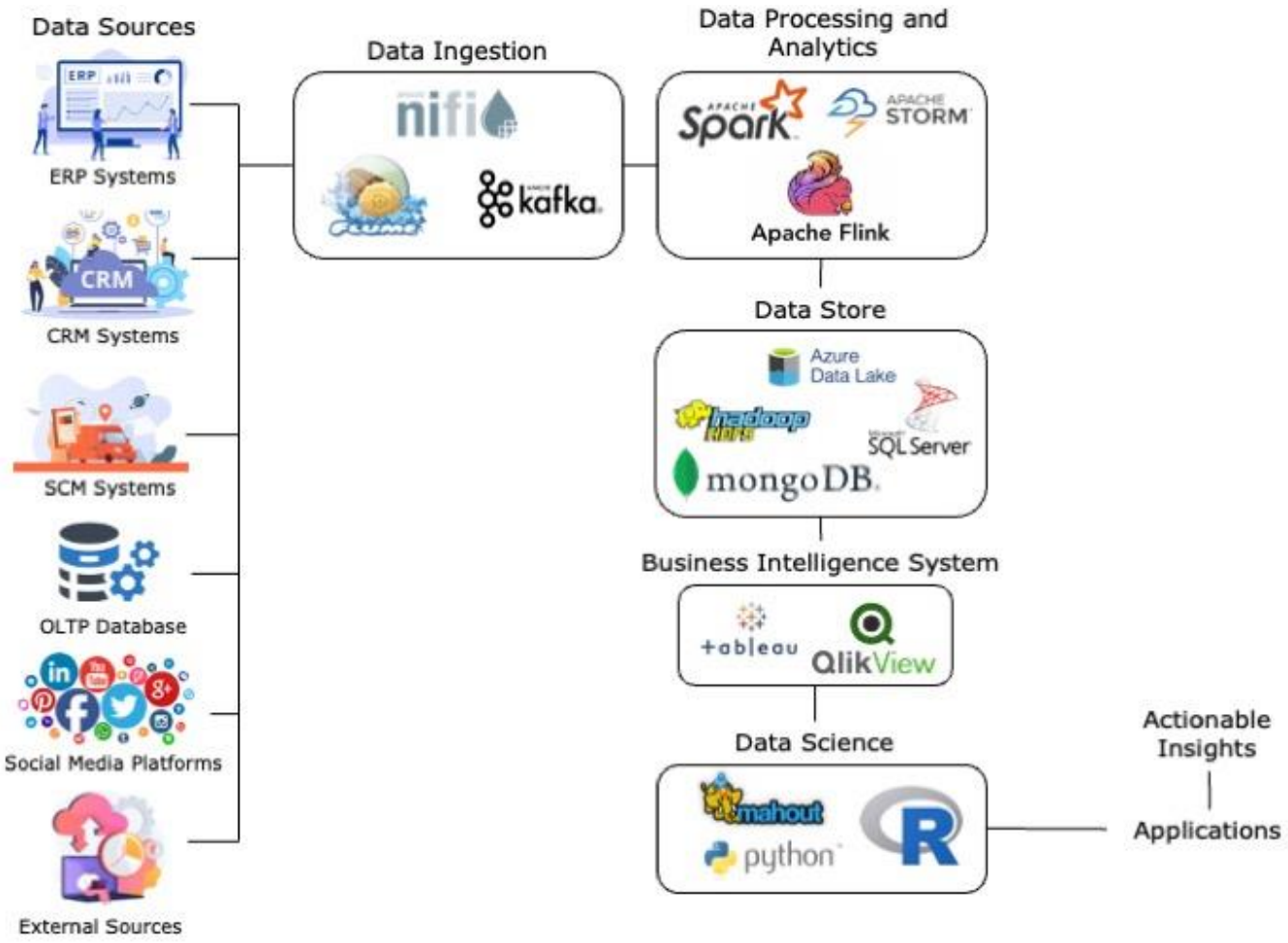

Figure 4: Coca-Cola's current Big Data Infrastructure

The processed and analysed data is stored in the data store layer, which consists of both on-premise data stores like Hadoop Distributed File System (HDFS), and Microsoft Azure Blob Storage. Business intelligence tools like Tableau, QlikView, and Power BI provide dashboards, reports, and ad hoc query capabilities to visualize and report the data in the data store layer.

Advanced analytics are performed on the data stored in the data store layer by the data science layer, which includes tools such as R, Python, and Apache Mahout. In the applications layer, custom-built applications or integrations with existing business applications such as CRM and ERP systems are created to use data insights to drive business decisions and improve business performance. Coca-big Cola's data infrastructure aims to provide actionable insights that drive business decisions and improve performance.

Coca-Cola stores various types of data using a combination of data storage technologies, as illustrated in figure 4. A relational database system, such as Microsoft SQL Server, is used for structured data such as sales volume, pricing, and customer information. These databases provide high levels of data consistency and reliability, which is critical for online transaction processing (OLTP) applications. A content management system (CMS) such as Amazon S3 is used for unstructured data such as customer reviews, social media likes, and social media followers. CMSs are built to deal with unstructured data and include features like content versioning, access control, and workflow management. For online analytical processing (OLAP) applications, Coca-Cola also employs data warehousing technologies such as Google BigQuery. Data warehousing systems are designed to store and analyse large amounts of data, and they can handle

complex queries and aggregations (Chaudhuri, 1997). Storing data with third-party service providers increases scalability and flexibility while decreasing operational costs, but it also introduces potential security risks. Coca-Cola uses a supply chain management system, such as SAP Supply Chain Management or Oracle Supply Chain Management, to obtain logistical information such as shipping time and logistics costs. These systems can aid in inventory management, route optimisation, and supplier performance monitoring.

Coca-Cola's current data access, analytics, and reporting processes involve a combination of technologies and services. Some of the potential processes and technologies being used include:

- Data Storage Technologies: As previously discussed, Coca-Cola uses a combination of data storage technologies such as relational databases, CMSs, data warehousing systems, and supply chain management systems to store and manage different types of data.
- Data Integration Technologies: To access and analyse data from multiple sources, Coca-Cola could use data integration technologies such as Apache Kafka or Microsoft Azure Data Factory. These technologies extract, transform, and load data from different sources into a unified data platform for analysis.
- Analytics and Reporting Tools: To analyse and visualize data, Coca-Cola could use analytics and reporting tools such as Tableau, Power BI, or Google Data Studio. These tools can help identify trends and patterns in data and provide insights for decision-making.
- Service-Oriented Architecture: Coca-Cola uses a service-oriented architecture (SOA) to organize their data and application infrastructure into discrete, reusable services. This can facilitate data access and integration across different applications and systems.

Coca-Cola's current data access, analytics, and reporting processes use a mix of technologies and services, such as data storage, integration, analytics, and reporting tools, as well as potentially service-oriented architecture. Coca-Cola gains insights from their data, makes informed decisions, and remains competitive in the market by leveraging these technologies. Table 2 provides the SWOT analysis done on the current infrastructure done on the organisation.

Strengths	Weakness
1. Integration of cloud-based services such as AWS and Azure provides scalability and cost-effectiveness. 2. Utilising a service-oriented architecture allows for modular and reusable services, increasing flexibility and agility. 3. Use of Hadoop Distributed File System and S3 storage enables scalable and reliable data storage.	1. Reliance on third-party providers could lead to data security risks and legal concerns regarding data ownership. 2. Technical complexity could lead to maintenance and compatibility issues. 3. Integration of data from multiple sources could lead to data quality issues and duplication. 4. Inability to handle real-time data processing and analysis.

4. Implementation of data warehousing technologies enables effective analysis of large datasets.	
Opportunities	Threats
1. Using analytics tools to identify new market trends and opportunities. 2. Leveraging big data analytics to improve supply chain management and optimize production. 3. Utilising social media data to understand consumer preferences and improve marketing strategies. 4. Leveraging machine learning and AI to improve product development and customer experience.	1. Cybersecurity threats and data breaches could compromise sensitive data and damage the company's reputation. 2. Compliance with data protection regulations such as GDPR and CCPA could increase costs and limit data usage. 3. Use of third-party providers could lead to vendor lock-in and potential issues with data portability. 4. Technical glitches and system failures could lead to data loss and impact business operations.

Table 2: SWOT analysis the current Big Data Infrastructure

As with any data infrastructure, there are important legal, security, and ethical concerns faced by Coca-Cola. Below are some key issues that are relevant to the infrastructure outlined above:

- Data security and ownership: The use of third-party providers such as AWS and Azure raises concerns about data security and ownership. Coca-Cola needs to ensure that they have appropriate contracts in place to protect their data and comply with legal requirements regarding data ownership and protection.
- Data protection regulations: Compliance with data protection regulations such as GDPR and CCPA is critical to avoid legal and reputational risks. Coca-Cola needs to ensure that they are transparent about their data practices and have appropriate policies and procedures in place to protect the privacy of personal information.
- Ethical issues: The use of big data analytics and AI raises ethical concerns about privacy, discrimination, and bias. Coca-Cola needs to ensure that their use of these technologies is transparent, fair, and unbiased.
- Technical issues: The complexity of the infrastructure could lead to technical issues such as data quality, duplication, and compatibility. Coca-Cola needs to ensure that they have appropriate governance, policies, and procedures in place to manage these issues effectively.

Overall, while the current infrastructure has several strengths, weaknesses, opportunities, and threats, Coca-Cola needs to be aware of the legal, security, and ethical issues that could impact their operations and reputation. By addressing these issues effectively, Coca-Cola can continue to leverage their infrastructure to gain insights from their data and stay competitive in the market.

Based on the analysis of the current infrastructure and pertinent issues, it is clear that the infrastructure has some limitations. Because of the increasing amount of data stored in the cloud and the complexity of the infrastructure, one major limitation is the potential for data breaches and cyber attacks. Another limitation is the legal and ethical issues that come with storing and processing sensitive customer data, which can pose legal and reputational risks if not handled properly. Furthermore, the current infrastructure may not be scalable enough to accommodate the growing volume of data and the organisation's changing needs.

There is also a risk of vendor lock-in because the infrastructure is heavily reliant on third-party service providers, limiting the organisation's flexibility and autonomy (Opara, 2016). Finally, a lack of standardisation in data storage and processing can result in data inconsistency, making effective analytics and reporting difficult. As a result, it is critical for Coca-Cola to address these limitations and work towards a more robust and secure infrastructure capable of meeting the organisation's evolving needs while adhering to legal and ethical standards.

Proposal for Enhancement of Current Infrastructure

This section proposes a variety of technologies and tools that could be implemented to improve Coca-Cola's current infrastructure. Based on the previous section's analysis, the following are the key requirements for a solution that will overcome data collection and storage issues, as well as infrastructure and technology issues:

- Scalability: The solution should be able to scale to accommodate the organization's growing volumes of data, as well as support the addition of new data sources and types. This can be accomplished by utilising cloud-based solutions that provide elastic storage and processing capabilities, as well as distributed computing frameworks such as Hadoop and Spark.

- Data Security: Given the sensitive nature of some of the data being collected, such as customer information and supplier details, it is crucial that the solution is designed with robust security mechanisms. This includes measures such as encryption, access controls, and auditing to ensure that data is protected both at rest and in transit.

- Data Governance: The solution should include strong data governance mechanisms to ensure compliance with relevant legal and ethical requirements. This includes features like data lineage tracking, data quality monitoring, and data cataloguing to enable effective data asset management and control.

Appendix 6 contains a detailed explanation of the solutions mentioned above. By prioritising these requirements, the organisation can create a solution that not only solves the current data collection and storage problems, but also lays the groundwork for future data-driven innovation and growth. Figure 6 depicts the proposed big-data infrastructure for addressing the issues listed in the preceding section.

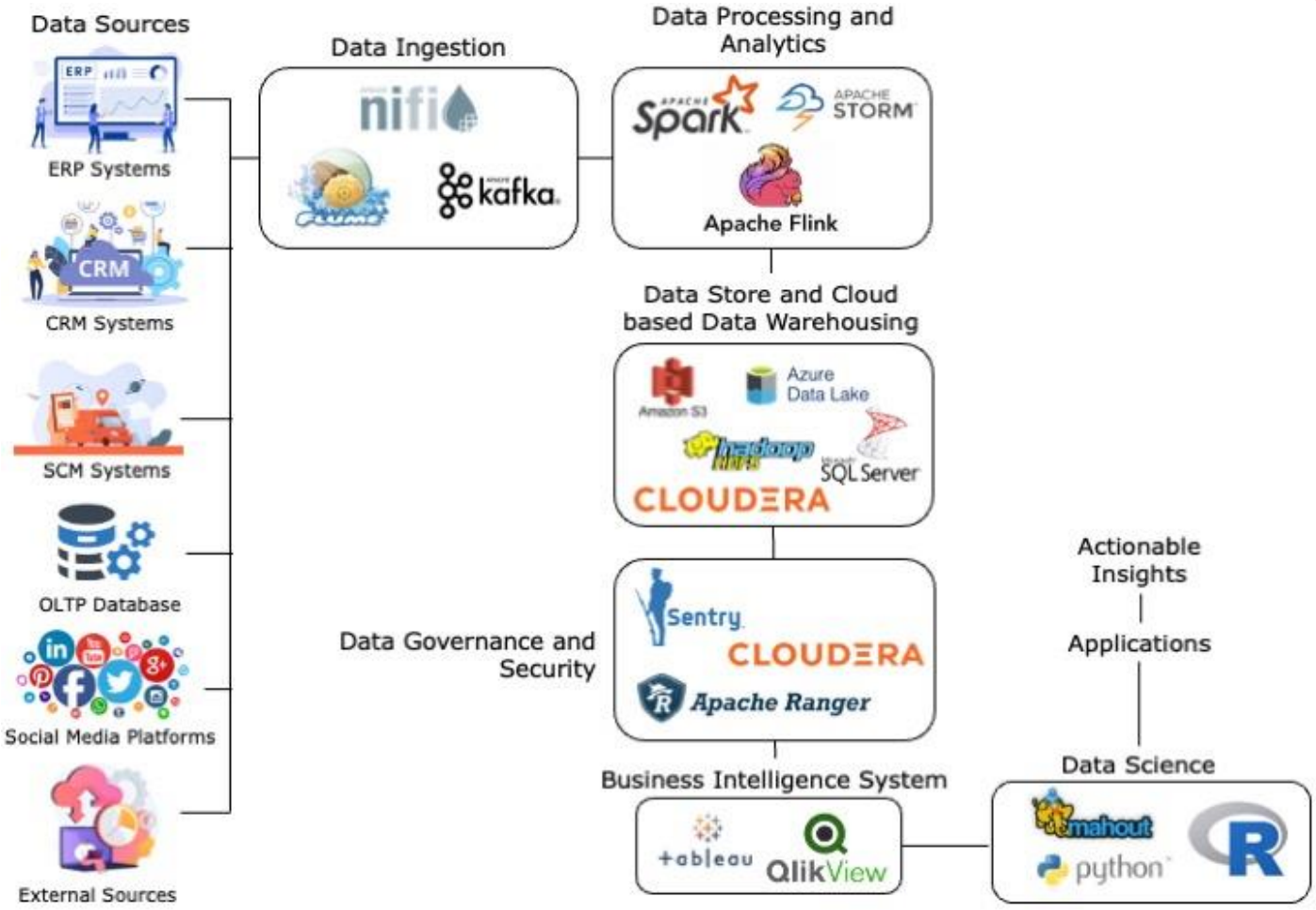

Figure 5: Proposed Infrastructure for Big Data

The proposed big data infrastructure is intended to support data scalability, data governance, and data security while also allowing for the efficient and effective processing of large volumes of data generated from various sources. Before delving into the specifics of the big data infrastructure proposed for the organisation, it is critical to understand the architecture's key components. Amazon S3 and Cloudera are two critical components that play critical roles in the infrastructure. These technologies are widely used in the industry and provide several advantages, including scalability, data security, and governance. In this context, it is critical to thoroughly discuss these technologies and how they can be integrated to create a robust big data infrastructure.

Amazon S3 (Simple Storage Service) is a cloud-based object storage service that allows for the storage and retrieval of data from any location at any time. It offers scalable and cost-effective storage solutions for a variety of applications such as data lakes, backups, and archiving. Coca-Cola can securely store and manage their data using Amazon S3, while also benefiting from its scalability and cost-effectiveness (Arun, 2023).

Cloudera, on the other hand, is a leading data management and analytics platform that enables businesses to gain insights from their data. It offers a wide range of data processing, data engineering, and data analytics tools and services. Coca-Cola can gain valuable insights from their data and improve their decision-making processes by utilising Cloudera (Cloudera, n.d).

Coca-Cola can use Amazon S3 and Cloudera in tandem to build a comprehensive data management solution. They can use Amazon S3 to store and manage their

data, as well as Cloudera to process and analyse it. Coca-Cola can streamline their data management processes and improve their overall data governance by integrating the two platforms.

In terms of scalability, Coca-Cola can scale their data storage and processing capabilities as needed by combining Amazon S3 and Cloudera. Coca-Cola can easily expand their data infrastructure to meet the demands of their growing business thanks to Amazon S3's scalable storage solutions and Cloudera's ability to handle large volumes of data.

Coca-Cola places a high priority on data security. They can benefit from advanced security features such as encryption, access control, and data backup and recovery by combining Amazon S3 and Cloudera. Amazon S3 offers industry-leading security and compliance features, whereas Cloudera provides enterprise-grade security and governance.

Coca-Cola can benefit from the integration of Amazon S3 and Cloudera in data governance. Coca-Cola can ensure that their data is properly organised and managed by utilising Amazon S3's versioning and object tagging features. Meanwhile, Cloudera's data governance and compliance tools can assist Coca-Cola in maintaining regulatory compliance and ensuring proper data use. Coca-Cola can create a comprehensive data management solution that provides scalability, data security, and data governance by combining Amazon S3 and Cloudera. This will allow Coca-Cola to better manage its data, gain valuable insights from it, and ultimately improve its business processes and decision-making capabilities. Sentry and Ranger are two software tools that can be used in a big data architecture to improve data security and governance. Cloudera created Sentry, an open-source access management tool. Administrators can use it to define and enforce fine-grained access control policies across multiple Hadoop components. Administrators can use Sentry to define roles and permissions that limit access to data, metadata, and other resources in the Hadoop ecosystem.

Sentry and Ranger, software tools that provide access control and authorization mechanisms for Hadoop ecosystems, are also highlighted in the proposed infrastructure. Sentry can be integrated with LDAP and Active Directory for user authentication and supports a variety of data sources such as HDFS, HBase, Impala, and others. This assists organisations in enforcing strict access controls and adhering to regulations such as GDPR and HIPAA (Pankaj, 2022).

Hortonworks' Ranger is another open-source tool (now part of Cloudera). It provides a framework for centralised security administration for Hadoop and other big data platforms. Ranger includes features such as role-based access control, data masking, and data encryption, as well as auditing and reporting. It enables administrators to define data access security policies as well as monitor and audit user activities. Ranger also supports LDAP and Active Directory integration for user authentication (Maxwell, 2020). Coca-Cola can achieve comprehensive security and governance across their big data platforms with Ranger, allowing them to meet a variety of regulatory requirements.

To ensure data security and governance, Coca-Cola can integrate Sentry and Ranger into their big data architecture. Sentry can define access control policies for Hadoop components like HDFS and Impala, whereas Ranger can provide

centralised security administration for the entire big data platform, including Hadoop, Hive, HBase, and others. Coca-Cola can enforce fine-grained access control policies, audit user activities, and monitor data access across their big data platform by combining Sentry and Ranger. This will aid in the protection of sensitive data, compliance with regulatory requirements, and the reduction of the risk of data breaches.

Sentry and Ranger can help Coca-Cola manage their big data platform more efficiently in terms of scalability. Administrators can ensure that users only have access to the data they need by implementing fine-grained access control policies, reducing the risk of data duplication and unnecessary data storage. This can help Coca-Cola optimise their data storage and processing resources, allowing them to scale their big data platform as needed.

Overall, Sentry and Ranger are two software tools that have been proposed to help Coca-Cola ensure data security and governance in their big data architecture. Coca-Cola can define fine-grained access control policies, audit user activities, and comply with regulatory requirements by implementing these tools. Furthermore, they can optimise data storage and processing resources, allowing them to scale their big data platform efficiently.

The proposed big data infrastructure offers numerous advantages to organisations, ranging from increased data scalability to improved data security and governance. However, putting this infrastructure in place raises a number of legal, security, and ethical concerns that must be addressed in order to protect sensitive data and adhere to legal and ethical guidelines. Additional legal, security, and ethical concerns should be addressed when implementing the proposed big data infrastructure:

Data protection laws compliance: Using a third-party service provider such as Amazon S3 raises concerns about data protection laws compliance. Coca-Cola should make certain that it has the necessary consent and permissions to store and process personal data on Amazon S3. They should also have a data retention policy in place to ensure that data is only kept for the time necessary.

Access control and authentication: Sentry and Ranger offer access control and authentication, but Coca-Cola should ensure that they are properly configured to prevent unauthorised access to sensitive data. Access should be restricted to those who require it, and access logs should be reviewed on a regular basis to detect any unusual activity.

Data security: Coca-Cola should make certain that all data stored on Amazon S3 is encrypted both in transit and at rest. Cloudera offers data encryption functionality, but Coca-Cola should also use strong encryption algorithms and keys to protect their data.

Ethical considerations: Coca-Cola should ensure that their data is used ethically and that it is not used to discriminate or harm any individuals or groups. They should also ensure that the data they collect and how it is used are both transparent.

Data retention and deletion: Coca-Cola should have a data retention and deletion policy in place to ensure that data is only kept for the time necessary. This will help to reduce the risk of unauthorised data access.

- Data breaches: Coca-Cola should have a robust data breach response plan in place to ensure that any potential data breaches are detected and addressed as soon as possible. Regular security audits and vulnerability assessments should be performed to identify and address any potential security risks.

Overall, Coca-Cola must consider all of these legal, security, and ethical concerns when implementing the proposed big data infrastructure. Coca-Cola can ensure that their data is secure, compliant with data protection laws, and used ethically by doing so.

Discussion

In this section, we will discuss the proposed solution to address the limitations identified in the previous section, as well as its key characteristics for achieving Coca-Cola's strategic goals. Based on the report's long-term goals, it can be deduced that Coca-Cola is focusing on digital transformation to improve its marketing strategies and gain insights into consumer behaviour and preferences. However, this transformation necessitates the use of massive amounts of data, necessitating the establishment of a solid big data infrastructure.

To handle large volumes of data and achieve scalability, the proposed solution in this report for Coca-Cola involves implementing Amazon S3 and Cloudera. Furthermore, data security and governance are critical concerns that must be addressed by implementing strong data security mechanisms such as encryption techniques and access controls, as well as a well-defined data governance framework with policies and procedures for data collection, storage, processing, and sharing.

Legal, security, and ethical issues have been identified as limitations in this proposed solution, which must be addressed through regular security audits and testing, as well as ensuring compliance with relevant data protection regulations. Furthermore, implementing this solution may present challenges, such as the need for skilled personnel to manage and maintain the big data infrastructure. The proposed solution is consistent with Coca-Cola's strategic goals of digital transformation, investing in existing brands, and increasing efficiency in terms of results and finances. Coca-Cola can gain valuable insights into consumer behaviour and preferences by leveraging big data, and then launch personalised marketing campaigns to increase sales and expand its online presence.

Furthermore, by leveraging cloud-based solutions and implementing a distributed computing model, the proposed solution aims to minimise infrastructure investment while maximising the capture, management, and analysis of more appropriate data. Coca-Cola can reduce the need for on-premises infrastructure and associated maintenance costs by utilising cloud-based solutions such as Amazon S3 and Cloudera. Cloud-based solutions also provide the scalability and flexibility required to handle large amounts of data while minimising infrastructure investment.

In addition, the use of a distributed computing model enables the efficient processing of large amounts of data by leveraging the computing power of multiple machines. This method maximises data capture, management, and analysis while minimising the need for costly hardware investments. Furthermore, the use of data governance tools such as Sentry and Ranger helps to ensure that the data being captured and analysed is appropriate and in accordance with applicable regulations, lowering the risk of investing in and analysing irrelevant or non-compliant data.

The proposed solution offers a cost-effective and efficient approach to data management and analysis, minimising infrastructure investment while maximising data capture, management, and analysis. This aligns with Coca-Cola's strategic goal of increasing efficiency in terms of outcomes and finances by investing in its existing brands, people, and resources while also venturing into new marketing terrains.

Finally, Coca-Cola's investment in digital transformation and data leveraging has proven to be a successful strategy in meeting its strategic goals of long-term growth. Coca-Cola has gained insights into consumer behaviour and preferences, launched personalised marketing campaigns, and expanded its online presence by investing in digital marketing, consumer engagement, data analysis, and e-commerce.

Coca-Cola can use cloud-based solutions like AWS and Cloudera to reduce infrastructure costs while increasing data capture, management, and analysis. These solutions offer scalable and cost-effective methods of storing, processing, and analysing large amounts of data. Furthermore, incorporating data governance tools such as Sentry and Ranger can ensure data security and regulatory compliance.

Overall, Coca-Cola's investment in digital transformation and data analytics has been a critical success factor in its growth strategy, and by continuing to leverage data in novel ways, Coca-Cola can drive sustainable growth and maintain its position as an industry leader.

Appendix 1

This Coca-Cola SWOT analysis shows how the organisation in charge of one of history's most recognisable brands exploited its advantages over rivals to rise to the position of second-largest beverage producer worldwide. The company sells around 500 non-alcoholic beverage brands, mostly sparkling but also include still drinks such waters, enhanced waters, juices, juice drinks, ready-to-drink teas, coffees, energy drinks, and sports drinks. The below table 3 list all the major advantages, disadvantages, opportunities, and dangers that have the greatest impact on the business (Jurevicius, 2022).

Strengths	Weakness
1. The largest market share in the beverage industry. 2. A more diverse product portfolio with a 21-billion-dollar brand 3. The competitor with the largest advertising budget. 4. The world's most well-known beverage brand. 5. Effective collaborations with bottling companies that result in some of the most extensive distribution networks in the industry. 6. Partnership with McDonald's.	1. The majority of the company's revenue comes from carbonated soft drinks. 2. Competitors' aggressive competition; Pepsi. 3. Health issues.
Opportunities	**Threats**
1. The market for ready-to-drink (RTD) coffee products in the United States has expanded. 2. A thriving tequila industry with a wide range of affordable smaller brands. 3. The coconut water market is expected to reach $8.3 billion by 2023. 4. The savoury snack sector will grow exponentially over the next three years. 5. If the value of the dollar falls, the company's income and earnings may rise.	1. Obesity concerns may cause some of the company's products to sell poorly. 2. Extending the "soda tax" to other cities or states in the United States. 3. Due to water scarcity and poor quality, the Coca-Cola Company's production costs and capacity may suffer. 4. The Coca-Cola Company's business may suffer as a result of increased competition and their capabilities.

Table 3: SWOT Analysis on Coca-Cola

Appendix 2

Coca-Cola collects a variety of data from customers for its analysis, including:

- Demographic Data: This data contains details like age, gender, income, degree of education, and occupation. The features of the company's target market are understood using this data, which is then used to adjust marketing strategies.

- **Behavioural data:** Information about past purchases, product preferences, and brand loyalty is included in this type of data. In order to comprehend customer purchasing patterns and preferences and to spot possible sales possibilities, this data is employed.

- **Social media data:** Data from social media sites like Facebook, Twitter, and Instagram is included in this category. This information is utilised to track brand-related conversations, identify influencers, and comprehend customer sentiment.

- **Geographical data:** These comprise details like location, climate, and population density. In order to improve distribution and logistics, it is necessary to understand the market circumstances and consumer preferences in various geographic areas.

- **Sales Data:** Information about product sales, revenue, and profit margins is included in sales data. The performance of the company's products is understood using this data, which is also used to inform decisions about new product development and innovation.

- **Surveys and feedback data:** Information obtained through surveys, online feedback forms, and customer service encounters is included in the category of surveys and feedback data. In order to comprehend client happiness and pinpoint areas for development, this data is utilised.

Overall, Coca-Cola gathers a variety of data from customers to learn more about consumer preferences and behaviour as well as to guide business decisions.

Appendix 3

As discussed, there are in total 14 Vs and a C defined to effectively manage and use Big Data. Table 4 depicts all the characteristics (Kapil, 2016).

Characteristics	Meaning	Description
Volume	Size of Data	Quantity of collected and stored data. Data Size is in Terabyte or Petabyte.
Velocity	Speed of Data	The transfer rate of data between source and destination.
Value	Importance of Data	Represents the business values derived from data.
Variety	Type of Data	Different formats of data, video, image, text, numbers.
Veracity	Data Quality	Accuracy of data captured. Data is not useful if the quality of it is bad.
Validity	Data Authenticity	Correctness/ Accuracy of data used to extract results in form of information.
Volatility	Duration of Usefulness	How long the stored data can be useful.
Visualization	Data Process/ Data act	Process of representing the data.
Virality	Spread Speed	Rate at which the data is broadcast by a user and received by different user.
Viscosity	Lag of Event	Time difference between the event occurred and event being described.
Variability	Data Differentiation	Efficiency at which system differentiates between noisy and important data.
Venue	Different Platform	Platforms from which data is captured; personnel systems, private & public cloud etc.
Vocabulary	Data Terminology	Terminologies like data model, data structure etc.
Vagueness	Indistinctness of existence in a Data	Refers to the reality in information that suggests little or no thought about where each might convey.
Complexity	Correlation of Data	Data is captured from different sources, and it is necessary to find out the changes in data with respect to the previously arrived data.

Table 4: Big Data Characteristics

Appendix 4

The definitions and interpretations of the 4 V's as they have been considered in the context of this report are briefly explained below (Zitter, 2022):

- **Volume**: Volume refers to the massive amount of data that overwhelms businesses. The days of storing data on local servers and managing it internally are long gone. Businesses handled terabytes of data 15 years ago. Every day, petabytes or exabytes of data, or 1,000-1,000,000 TB, are generated from sources such as transaction processing systems, emails, social media, customer databases, website lead captures, monitoring devices, and mobile apps.

- **Velocity**: It refers to the rate at which data is generated and transferred. This is a critical factor for businesses that need their data to flow quickly so that it is available when needed to make the best business decisions. A large data-using company will have a steady stream of data being produced and sent to its destination. Devices, networks, smartphones, and social media can all emit data. This data must be processed and analysed quickly, often in near real time.

- **Variety:** Variety refers to the various types of digitised data that businesses encounter, as well as the processing and insight-mining techniques for these various types of data. A company may collect data from a variety of sources, the value of which varies. Data can come from both inside and outside of a company. In terms of variety, the standardisation and distribution of all data being gathered pose a problem.

- Veracity: Veracity refers to the data's dependability and excellence. The information gathered may be incomplete, incorrect, or incapable of providing any useful, insightful information. In general, veracity refers to the level of trust in the data that has been gathered. A business can only make a profit and affect change if it has complete and accurate information. Organizations can only benefit from clean data. That is, if it is correct, flawless, dependable, consistent, impartial, and exhaustive.

Appendix 5

Below shown Marr's Strategy assists in defining and understanding the organization's strategy and how it can be met using Big Data.

Purpose Panel

Purpose: Improve the organisational profit using Big Data effectively

Ambition: Leverage the power of big data to drive our business forward and deliver on our commitment to sustainability, innovation, and exceptional customer experience.

Customer Panel

Target: Consumers aged 15-40 who are health-conscious and prefer low-calorie or no-sugar options.

Value Proposition: Ability to gain a deeper understanding of its customers and their preferences, behaviours, and purchasing patterns. This can lead to more targeted and effective marketing and sales strategies, resulting in increased revenue and customer loyalty.

Operational Panel

Partners: Suppliers, Distributors, Retailers, Communities, Technology Partners

- **Product Development:** Excel at developing new and innovative products that meet the evolving needs and tastes of its customers.
- **Marketing and Brand Building:** Known for its strong brand identity and has a long history of successful marketing campaigns. This includes creating demand for its products and building brand recognition and loyalty.
- **Supply Chain Management:** Efficient and effective supply chain management process that enables it to meet the high demand for its products while maintaining quality control.
- **Distribution:** Well-established distribution network that reaches a wide range of customers and markets, including both retail and foodservice customers.
- **Customer Service:** Value its customers and strives to provide them with high-quality customer service, including prompt and reliable delivery and responsive customer support.
- **Sustainability and Corporate Social Responsibility:** Committed to operating in a responsible and sustainable manner, and has made a number of investments to reduce its environmental impact and support the communities in which it operates.

Financial Panel

Use big data to improve its finance by leveraging data to gain a deeper understanding of customer behaviour and preferences.
- Sales and demand analysis: Analysing sales and demand data to identify trends and patterns that can inform product pricing, distribution, and marketing strategies.
- Customer behaviour analysis: Understanding customer behaviour can to tailor its marketing and product offerings to better meet customer needs, leading to increased sales and profits.
- Supply chain optimization: Big data can be used to improve supply chain efficiency, reducing waste and costs, and improving profitability.
- Fraud detection and prevention: Detect and prevent financial fraud, reducing the risk of losses and improving financial stability.
- Improved decision making: Make more informed decisions, reducing the risk of bad investments and increasing the chances of successful outcomes.
- Better pricing strategies: Analyse market trends and competition, allowing it to set more competitive prices and improve its financial performance.

Competition and Risk

- **Market competition:**
Intense competition from established and new beverage brands.
Growing popularity of health-conscious drinks.
Competition for consumer attention and market share.
- **Customer Risks:**
Changing consumer preferences and buying behaviour.
Difficulty in retaining loyal customers.
Negative consumer perceptions of sugary drinks.
- **Operational Risks:**
Supply chain disruptions and raw material shortages.
Increased costs for production, transportation, and distribution
Difficulty in expanding into new markets.
- **Financial Risks:**
Volatility in commodity prices.
Changes in economic conditions.
Weaknesses in pricing strategies.
- **IT systems and data risks:**
Cybersecurity threats and data breaches.
Dependence on technology systems and potential system failures.
Privacy concerns regarding consumer data.
- **People Risks:**
Employee turnover and recruitment challenges.
Difficulty in retaining top talent.
Potential conflicts with labour unions and employee relations issues.

Resource Panel

IT Infrastructure and Data
- Customer relationship management (CRM) system
- Supply chain management (SCM) system
- Data warehousing and analytics platform
- Point-of-sale (POS) data collection and analysis
- Social media monitoring and analysis
- Mobile app data and usage analytics
- Customer behaviour and purchase data analysis
- Inventory management and optimization
- Market research and consumer insights
- Sales and revenue tracking and analysis.

Infrastructure
- Manufacturing plants and bottling facilities.
- Warehouse and distribution centres.
- Equipment for production, packaging, and labelling.
- Delivery vehicles and transportation systems.
- Water treatment facilities.
- Raw material storage facilities.
- Energy systems, including electrical and fuel-systems.
- Office buildings and other administrative facilities.

IT infrastructure, including hardware, software, and network systems.
- Land for expansion and future growth.

People and Talent
- Skilled and experienced employees in various departments, including sales, marketing, supply chain management, finance, and IT.
- Strong leadership and management team to drive company strategy and growth.
- Employee engagement and training programs to ensure ongoing professional development and job satisfaction.
- A diverse and inclusive workplace culture that values and supports all employees.
- A strong network of suppliers and partners to support production and distribution efforts.
- Strong relationships with key stakeholders, including customers, communities, and government agencies.

Culture, Values and Leadership
- **Customer focus**: A commitment to understanding and meeting the needs of customers and stakeholders.
- **Teamwork**: Encouraging collaboration and mutual support among employees.
- **Integrity**: Demonstrating honesty and transparency in all business dealings.
- **Innovation**: Encouraging creativity and a willingness to take risks to drive growth and improvement.
- **Responsibility**: Holding oneself and others accountable for their actions and decisions.
- **Sustainability**: A commitment to reducing the company's impact on the environment and promoting sustainable business practices.
- **Diversity and Inclusion**: Promoting a workplace culture that values and respects all employees, regardless of their background or identity.
- **Continuous learning**: Encouraging and supporting ongoing personal and professional development for employees.
- **Empowerment**: Encouraging and enabling employees to take ownership of their work and make decisions that drive success.

Figure 4: SMART Strategy board for Coca-Cola

Appendix 6

The key requirements for a solution that will overcome the data collection and storage issues are explained below:

Scalability: The proposed solution for Coca-Cola must have a scalable infrastructure that can handle large volumes of data efficiently. To achieve scalability, the solution must be designed with distributed computing in mind, allowing the addition of more resources as needed. The solution must also be able to handle high-velocity data ingestion and processing to support real-time decision-making. The use of cloud-based solutions, such as AWS or Azure, can provide the necessary scalability, flexibility, and cost-effectiveness required by Coca-Cola.

Data Security: Data security is of utmost importance for Coca-Cola. The proposed solution must have robust data security mechanisms in place to protect sensitive data from unauthorized access, theft, or modification. The use of industry-standard encryption techniques, such as SSL or AES, can provide a secure data transmission and storage mechanism. The solution must also have access controls in place to ensure that only authorized personnel can access and modify the data. Regular security audits and testing must be conducted to identify vulnerabilities and ensure compliance with relevant data protection regulations such as GDPR or CCPA.

Data Governance: Data governance is critical for ensuring data quality, consistency, and compliance. The proposed solution must have robust data governance mechanisms in place to ensure data accuracy, completeness, and consistency. A well-defined data governance framework that outlines policies, procedures, and standards for data collection, storage, processing, and sharing must be developed. The framework should also include guidelines for data ownership, access controls, and data lifecycle management. The solution must also have an audit trail mechanism in place to track all data-related activities to ensure compliance with regulatory requirements. The use of data governance tools such as Collibra or Informatica can provide the necessary capabilities to ensure effective data governance.

References

Coca-Cola History (n.d.) The Coca-Cola Company. Available at: https://www.coca-colacompany.com/company/history (Accessed: January 26, 2023).

Prafull, P. (2018) Coca-Cola's marketing communication strategy: A critical analysis, LinkedIn. Available at: https://www.linkedin.com/pulse/coca-colas-marketing-communication-strategy-critical-analysis-pragya (Accessed: February 4, 2023).

Cuofano, G. (2023) Coca-Cola Organizational Structure, FourWeekMBA. What is The FourWeekMBA. Available at: https://fourweekmba.com/coca-cola-organizational-structure/ (Accessed: January 26, 2023).

Montgomery, K. et al. (2017) Big Data and the transformation of food and beverage marketing: Undermining efforts to reduce obesity?, Taylor & Francis. Available at: https://www.tandfonline.com/doi/full/10.1080/09581596.2017.1392483 (Accessed: January 26, 2023).

Joe Kaeser. (2018) Data is the 21st century's oil, The Economic Times. Available at: https://economictimes.indiatimes.com/magazines/panache/data-is-the-21st-centurys-oil-says-siemens-ceo-joe-kaeser/articleshow/64298125.cms (Accessed: January 11, 2023).

Alsghaier, H. et al. (2017) The importance of Big Data Analytics in Business: A Case Study. Available at: https://www.researchgate.net/publication/320225320_The_Importance_of_Big_Data_Analytics_in_Business_A_Case_Study (Accessed: January 28, 2023).

Panoho, K. (2019) The age of analytics and the importance of data quality, Forbes. Forbes Magazine. Available at: https://www.forbes.com/sites/forbesagencycouncil/2019/10/01/the-age-of-analytics-and-the-importance-of-data-quality/ (Accessed: January 28, 2023).

Cackett, D (2013) Information Management and Big Data A Reference Architecture. Available at: https://www.oracle.com/technetwork/topics/entarch/articles/info-mgmt-big-data-ref-arch-1902853.pdf (Accessed: January 28, 2023).

Marr, B. (2015) Big Data: Using smart big data, analytics and metrics to make better decisions and improve performance. Chichester, West Sussex, United Kingdom: Wiley (Accessed: January 28, 2023).

Ulunma (2020) Coca Cola leverages Data Analytics to Drive Innovation, Digital Innovation and Transformation. Available at: https://d3.harvard.edu/platform-digit/submission/coca-cola-leverages-data-analytics-to-drive-innovation/ (Accessed: February 8, 2023).

Meg (2019) 16 Data Visualization Tools & Guide, Talkwalker. Available at: https://www.talkwalker.com/blog/data-visualization-tools (Accessed: February 8, 2023).

Bumblauskas, D. et al. (2017) Big Data Analytics: Transforming Data to Action, Business Process Management Journal. Emerald Publishing Limited. Available at: https://www.emerald.com/insight/content/doi/10.1108/BPMJ-03-2016-0056/full/html (Accessed: February 8, 2023).

Schwartz, S. (2018) How coca-cola migrated from a single data warehouse to Global Application Deployment, CIO Dive. Available at: https://www.ciodive.com/news/how-coca-cola-migrated-from-a-single-data-warehouse-to-global-application-d/519070/ (Accessed: February 8, 2023).

Chaudhuri, S. and Dayal, U. (1997) An overview of data warehousing and OLAP technology, ACM SIGMOD Record. Available at: https://dl.acm.org/doi/10.1145/248603.248616 (Accessed: March 20, 2023).

Opara Martins, J., Sahandi, R. and Tian, F. (2016) Critical analysis of Vendor Lock-in and its impact on cloud computing migration: A Business Perspective - Journal of Cloud Computing, SpringerOpen. Springer Berlin Heidelberg. Available at: https://journalofcloudcomputing.springeropen.com/articles/10.1186/s13677-016-0054-z (Accessed: March 20, 2023).

Arun, R. (2023) What is AWS S3: Overview, features & Storage Classes explained: Simplilearn, Simplilearn.com. Simplilearn. Available at: https://www.simplilearn.com/tutorials/aws-tutorial/aws-s3 (Accessed: March 20, 2023).

Cloudera: Unlock the power of Data (no date) Talend. Available at: https://www.talend.com/resources/what-is-cloudera/ (Accessed: March 20, 2023).

Pankaj (2022) Re: Sentry integration with LDAP groups, Re: Sentry integration with LDAP Groups - Cloudera Community - 8514. Available at: https://community.cloudera.com/t5/Support-Questions/Sentry-integration-with-LDAP-Groups/m-p/31221 (Accessed: March 20, 2023).

Maxwell, E. (2020) An introduction to apache ranger, Medium. Privacera. Available at: https://blog.privacera.com/an-introduction-to-apache-ranger-655f13bcc49d (Accessed: March 20, 2023).

Jurevicius, O. (2022) Coca Cola SWOT analysis (6 key strengths in 2022), Strategic Management Insight. Available at: https://strategicmanagementinsight.com/swot-analyses/coca-cola-swot-analysis/ (Accessed: January 27, 2023).

Kapil, G. et al. (2016) A study of Big Data characteristics | IEEE conference publication. Available at: https://ieeexplore.ieee.org/document/7889917/ (Accessed: January 28, 2023).

Zitter, L. (2022) What are the 5 V's of big data?: Definition & explanation, TechnologyAdvice. Available at: https://technologyadvice.com/blog/information-technology/the-four-vs-of-big-data/ (Accessed: January 28, 2023).