

Heartland Escapes: Business Data Analysis

Big Data Analytics – CS686

Aidan Polivka

October 6, 2024

# Table of Contents

Table of Contents ..... i

Organizational Background ..... 1

    Organization Inception ..... 2

    Data Structures and Management ..... 3

Identification of Big Data ..... 4

    Data Utilization ..... 4

    Unstructured Data ..... 4

    Data Types and Business Analysis ..... 6

Big Data Analytical Tools ..... 7

    Industry Standard Big Data Tooling ..... 7

        Extract, Transform and Load (ETL) Tools..... 7

        Data Warehousing ..... 9

        Business Intelligence (BI) ..... 11

    Selection for Heartland Escapes ..... 12

Big Data Analytics System Architectures ..... 14

    Data Architecture Selection Recap ..... 14

Data Analytics Alignment with Existing Architecture.....	15
Data Analytics Training Requirements.....	16
New System Analytics Support.....	16
Big Data Analytics Techniques .....	18
Popular Big Data Use Cases .....	18
Artificial Intelligence .....	18
Natural Language Processing (NLP).....	19
Traditional Statistical Analysis .....	20
Techniques Suggested for Proposed System .....	20
Statistical Regression Analysis .....	20
Sentiment Analysis of Instagram Comments .....	21
NLP Author Description from Book Reviews .....	21
Training Considerations .....	21
Big Data Analytics Value Creation .....	23
Analyst Responsibilities and Maintenance .....	24
References.....	25

## Organizational Background

Throughout my time at Colorado Technical University, I've developed a fictional bookstore called Heartland Escapes. Previous projects include a migration plan from an on-premises system to an Azure Cloud environment, a system networking plan for cloud computing, a system security management and maintenance plan, and a database schema development plan. I think it's only natural to continue working with this fictional store here to analyze the business uses of their data and identify new utilization opportunities of both structured and unstructured data relevant to Heartland Escapes.

As stated previously, Heartland Escapes is a bookstore in Lincoln Nebraska with two locations. They have seen recent growth and consumer interest due to their expanding presence in social media. Their business model is heavily centered around hosting events at their stores including author meet and greets, summer reading programs, holiday-oriented events (like scary story readings on Halloween or Santa reads Christmas stories), and many others. Because of their marketing of these events, they've received a lot of local publicity and foot traffic in store. With the growth they've seen over the past year, Heartland Escapes would like to expand to two neighboring cities in Nebraska with a new location in Omaha Nebraska, and a new location in Beatrice Nebraska. (Polivka, 2024)

## Organization Inception

Heartland Escapes started out as a form of community necessity. The public libraries were well stocked with books but lacked in providing young people with the spark to enjoy reading. The owners of the company are avid book readers themselves and have always enjoyed sharing their literary journeys with other like-minded individuals. Heartland Escapes mission is to be a shared “Escape” for others to enjoy captivating stories and their love for books with each other. The first store opened in 2010, and it was a slow start to get people in the door with competing companies like Barnes and Noble around. As people began seeing the benefits of community-oriented storytelling, traction accrued to the point that Heartland Escapes had the financial stability to open a second store in 2017. The first year of the pandemic was difficult for the company, but after the guidelines loosened and the public started feeling more comfortable leaving their homes, business caught its stride once again. Since then, Heartland Escapes has seen nothing but growth and enthusiastic customers. (Polivka, 2024)

## Data Structures and Management

When Heartland Escapes started their journey, they developed their own point-of-sale system and inventory management API. Married to these systems are the Accounting Database and the Inventory Database. Additionally, they have a public facing website where users can search store inventory with and see store hours and event schedules.

(Polivka, 2024)

The data stores that Heartland Escapes currently has available are Microsoft SQL Databases, excel spreadsheets, generated reports from their Inventory API and Point-of-Sale system, social media platforms, google reviews, and potentially their third-party payment processor.

# Identification of Big Data

## Data Utilization

The structured data within the Heartland Escapes system is primarily Microsoft SQL Server managed databases with very limited breadth. These databases are mostly used for inventory and purchase/sale management, although a large amount of Heartland Escapes accounting is done manually using excel spreadsheets. They also run reports through their point-of-sale system and the inventory API to generate needed information about monthly revenue, inventory restocking needs and other business operation required reports. Heartland Escapes also uses a third-party payment processor that may have API endpoints to retrieve data about purchases, but that hasn't been fully explored by stakeholders.

This structured data could be utilized to create more in-depth reports about book genre sales over time. Heartland Escapes could also use some of the structured data from their social media profiles for what book genres and authors are sold after a viral post based on demographic engagement analysis.

## Unstructured Data

Heartland Escapes has a lot of unutilized unstructured data. Their social media standing mainly comes from spurts of local virality from their talented staff rather than true

advertising. There is a lot of untapped potential from the available demographic data of followers, high account traffic times, analysis of viral posts, and comment analysis.

They could use the demographic data for targeted advertising and time of day interaction rates for optimal posting times. Additionally, the viral posts can be analyzed for what they have in common that resulted in a viral response, leading to better engagement from followers and therefore more foot traffic through the doors of Heartland Escapes. Being a company with social media influence, they also have a large amount of data from google reviews that could be utilized to determine what they're doing well and what needs improvement from the perspective of customers.



## Data Types and Business Analysis

Both the structured data and unstructured data are crucial for the business analyst's work. This data is a tool in the analyst's arsenal for investigating what aspects of the business can be improved to lead to more profit. The underutilized data from Heartland Escapes' social media accounts presents a fantastic opportunity to gain new insights and find ways to attract more customers to their stores.

Structured data is important for the analyst in that it offers concrete, easy to understand insight into business processes and metrics. This also provides quantitative information to bring back to the stakeholders to explain issues, solutions and improvements to their company. Unstructured data is equally important in that it provides more qualitative information about the company. Whether this is coming from google reviews, social media comments or some other source, this data is crucial to understand customer sentiment and make improvements based on consumer feedback. Business analysts use both structured and unstructured data to form a holistic view of the business.

## Big Data Analytical Tools

Heartland Escapes needs a data warehouse capable of storing multiple types of data (structured, unstructured, and semi-structured). They also need an ETL tool to inject data from their social media sites into their warehouse, and they also need a business intelligence tool to analyze the data stored in the warehouse. Considering their organization's size, they need a tool that is easily maintained and updated and offers simple querying capabilities for analysis. Additionally, they'll need a suite of resources that are budget friendly, as their organization is not large enough to be spending exorbitant amounts of money on business analysis.

## Industry Standard Big Data Tooling

### Extract, Transform and Load (ETL) Tools

Most organizations that care about business intelligence require the use of data from many sources. In this case for Heartland Escapes, they care about data from their own databases and data from their Instagram and TikTok accounts. In order to access this data and store it in a centralized location, an Extract, Transform and Load Tool (or ETL tool for short) is required. Most of the time it's ideal to extract data from other sources and store it yourself to improve the performance of your business intelligence (BI) tools. In some cases, it might be necessary for your BI tool to perform some querying itself.

There are many ETL tools available today in the big data industry. Your choice of tooling should be influenced by utility requirements, employee experience, platform harmonization (Microsoft integrates well with other Microsoft products, for example), and cost. For example, it wouldn't make much sense to choose an AWS ETL tool if your data sources and warehouses are in Google Cloud SQL and Big Query. Also, if your desired tool only offers custom scripting in Go and your team is a .NET shop, that would be an abrupt change for most engineers. When considering platform harmony, it's important to not choose platform harmony at the cost of platform dependence.

There are a couple popular ETL options that would make sense for Heartland Escapes, considering these criteria. SQL Server Integration Studio (SSIS) would be a solid option considering they are a Microsoft shop. All their existing applications are written in C# .NET MVC and .NET Core Rest APIs. Also, their database are Microsoft SQL Server. SSIS offers custom scripting in C#, which would allow Heartland Escapes to perform requests to Instagram and TikTok APIs for data. Finally, SSIS can integrate with most data solutions, including Google Big Query and Cloud SQL. Since the rest of their system is hosted in Google Cloud Platform (GCP), integration with GCP is a strict requirement. SSIS's licensing is rolled into the enterprise license for SQL Server as well, however the infrastructure maintenance and cost may pose problems for Heartland Escapes. SSIS is not managed by Google Cloud platform, who hosts their other applications. This means that SSIS will need to run on a separate server, or within a Compute Engine Virtual Machine which is an expensive and high maintenance cloud component on GCP (Google, 2024b).

Another popular option would be ETLBox. This is a new ETL tool that's built on .NET Core, which would fit in incredibly well with Heartland Escape's existing software stack. ETLBox is built into an .NET Rest API, using network calls to trigger ETL actions. This includes Pub/Sub integration out of the box, and integration with the rest of Google Cloud Platform. Since it's built on a .NET Rest API, the Rest API can be containerized and deployed to Cloud Run, making it a fully managed tool with no infrastructure management needed. The biggest downside to ETLBox is that there is no UI pipeline development. It can only be maintained and updated by individuals with C#/Visual Basic skills (Lennartz, 2024).

The final ETL option for Heartland Escapes would be Google Cloud Dataflow. Dataflow meets many of Heartland Escapes requirements: It is a fully managed Google Cloud Platform tool. Therefore, it doesn't require any infrastructure maintenance, and it provides platform harmony by remaining within the GCP toolset. As a result of this harmony, Dataflow is easily integrated into other GCP components and tools. Dataflow also offers templates for commonly scripted flow types, making creating ETL flows quicker and simpler. One of the major bonuses to Dataflow is the integrated monitoring and observability that comes with using this tool on Google Cloud Platform. The downside is that Dataflow does not offer C# scripting, the only scripting available is in Python or Java. An additional downside is platform dependency (Google, 2024c).

## Data Warehousing

After the ETL Tool "extracts" and "transforms" the data, it needs to be able to "load" it into some resource or collection of resources. This is the host for the organization's big

data and will be the data source for the organization's business intelligence tool. In choosing a Data Warehousing solution, it's important to have some of the same considerations as for the ETL Tool. It needs to be able to fulfill your company's requirements, platform harmony is a bonus, it needs to be highly performant and cost effective. For Heartland Escapes, there are really two natural options that match all these specifications.

The first solution would be a combination of Google Cloud Storage and Google Big Query. Big Query offers storage of both unstructured and semi-structured data types. Big Query's uses of columnar storage making it optimized for fast query times and efficient data processing. Google Cloud Storage is a great tool for unstructured data storage, and it offers different storage costs per document usage frequency. It also allows you to create automated retention policies and storage type resolutions. For example, if Heartland Escapes does not typically use documents three years or older, those can be moved to archive storage therefore reducing the storage cost. Both Cloud Storage and Big Query operate on a pay-as-you-go pricing model.

The second solution would be a combination of Snowflake and Google Cloud Storage. Snowflake has somewhat recently partnered with Google and can be fully integrated into GCP (Ichhpurani, 2019). Snowflake offers very similar benefits to Big Query, although it's slightly more performant and slightly more costly (also a pay-as-you-go pricing model) than Big Query. Snowflake is easily integrated into Google Cloud Storage as well. The major benefit that Snowflake offers over Big Query is that it's able to be used across

various platforms. This removes some platform dependency while also providing platform harmony.

## Business Intelligence (BI)

The business intelligence tool is the presentation side of a big data technology stack. This is the tool that is used to query the compiled data and display it to business analysts. Heartland Escapes has a couple natural options for a business intelligence tools, considering the options that were selected for data warehousing.

The first natural choice would be Google's Looker Studio. Looker Studio free to use and integrates natively with Google Cloud Storage and Google Big Query. It also offers multi-cloud support, meaning that if Heartland Escapes chooses to migrate cloud platforms to AWS, they'd be able to do so without destroying the existing BI analytics dashboards. Looker offers third-party connectors, which means that it will be able to integrate with Snowflake as well. Supposedly, the user interface is intuitive, which makes development of dashboards easy for technical and non-technical individuals (Google, 2024a). As a result, these dashboards should be easily maintained for the Heartland Escapes stakeholders. There is no guarantee that third party connectors are free to use, so that may be a pitfall for Looker if Snowflake is chosen as the data warehousing tool (Snowflake, Inc, 2024).

The second choice for Heartland Escapes is Power BI. Power BI is a very common business intelligence tool that offers many similar benefits to Looker Studio. It also offers direct integration with Big Query, Cloud Storage, and Snowflake. Power BI offers a free

edition, however there are more collaboration and processing capabilities using the Power BI paid versions (Kenneth, 2022).

## Selection for Heartland Escapes

Because the rest of the Heartland Escapes system is containerized and/or portable, it's important that the chosen big data technology stack doesn't pigeonhole them into a cloud platform dependency. Any configuration of the options outlined within this section are completely viable solutions for Heartland Escapes. The best organization of these tools that would offer Heartland Escapes a cost effective, easily maintainable solution with a familiar tech stack would be:

- ETL Tool: ETLBox
  - Platform independence and containerization will make this tool highly portable for Heartland Escapes
  - Familiar development language will make this tool easily maintained by the Heartland Escapes team
  - Containerized deployment with Cloud Run removes any need for infrastructure management, making the system easily maintainable
- Warehousing Tool: Snowflake & Google Cloud Storage
  - Snowflake is fully managed, so there is no need to manage infrastructure
  - Snowflake being multi-cloud offers more simple portability

- Snowflake in combination with cloud storage provides coverage of all potential data formats
- Although Snowflake is slightly more costly, removing the platform dependency is worth it.
- BI Tool: Google Looker Studio
  - Looker is free to use without any paywall behind collaboration
  - It's easy to use and the UI is intuitive enough for non-technical individuals to participate in dashboard development
  - Offers native support for Cloud Storage, and there are third-party connectors to easily integrate Snowflake



# Big Data Analytics System Architectures

## Data Architecture Selection Recap

The selected big data architecture consists of three primary components: An ETL tool, a warehousing platform, and a BI tool. The selected Extract, Transform, and Load tool, ETLBox, was chosen because it is highly compatible with the existing architecture, it's highly portable, platform independent, and suitable for the current development skills of Heartland Escapes engineers. ETLBox can also be containerized, this paired with Google Cloud Platform's Cloud Run component removes any need for infrastructure management.

The chosen warehousing tool is a combination of Snowflake and Google Cloud Storage. Between the two, there is storage coverage of all data types. Snowflake is also multi-cloud, reducing platform dependencies to GCP. Snowflake is a commonly used data warehousing solution, making it well documented and easier to develop upon. The final piece to Heartland Escapes data analytics architecture is the chosen business intelligence tool. Google Looker Studio was chosen because it's free, multi cloud, and has an intuitive dashboard building environment for non-engineers.

All these sub-components of the larger data analytics system were chosen because they integrate well with the existing system, and they integrate well with one another. These sub-components are highly compatible with the current cloud platform, and together they form a cost effective, maintainable data analytics solution for the Heartland Escapes stakeholders.

## Data Analytics Alignment with Existing Architecture

The data analytics architecture fits well with the existing architecture. Parts of the e-commerce system and the point-of-sale system are hosted using GCP's Cloud Run component. This is the same component that will host the ETL tool, which doesn't require any infrastructure management. Both the point-of-sale system and e-commerce system are written in C#, which is the same language used for ETLBox. ETLBox has event-based messaging support, allowing other platform tools like Google Cloud Scheduler and Google Pub/Sub to interact with the ETL tool.

Snowflake has native support in Google Cloud Platform, which ensures a basic level of alignment with the existing architecture. Snowflake also offers third-party connectors, extending portability and connectivity with software systems. ETLBox should have no problems connecting to Snowflake, and other parts of the system like Google Cloud Logging and Cloud Storage should be able to stream data to Snowflake without too much additional configuration. Google Cloud Storage is a part of Google Cloud Platform and is easily connected to in .NET. Google offers .NET libraries to write files to and read files from Google Cloud Storage.

Looker Studio also has native support in Google Cloud Platform. Although it doesn't natively support Snowflake, it also has third-party connectors that are easily configured.

## Data Analytics Training Requirements

Some training may be required to utilize the ETLBox library. Although the Heartland Escapes engineers have experience writing in .NET C#, learning how to connect the existing SQL solutions, develop transformations and loading them into Snowflake and Cloud Storage may require some new learning.

There may require some minor training with Snowflake, especially considering loading Cloud Storage documents into Snowflake and transforming them into something useable by Looker. The engineers will need to learn how to use Snowflake and Cloud Storage together.

Looker Studio should require very little training, seeing as it's a fairly intuitive platform. There will need to be some general training for the engineers on connecting Snowflake to Looker and developing Snowflake queries for Looker.

## New System Analytics Support

One of the primary purposes of this new data analytics system is to support analysis of Heartland Escapes operations in the context of their social media platforms. This system can accomplish this goal using these steps as an example:

1. ETLBox extracts and loads data from the Cloud SQL instances to the Snowflake warehouse
2. ETLBox extracts and loads data from the Instagram API to Google Cloud Storage

3. Snowflake hosts a query that compiles data from Cloud Storage and Snowflake tabular data that can be used to build a dashboard
4. Business Analysts utilize Looker Studio to perform said query in Snowflake, and build a dashboard based on transformations of the returned data

Business analysts can request other queries to be developed to support Heartland Escapes system analysis. If analysts need data from other sources, Heartland Escapes engineers can create new endpoints within ETLBox to extract and load data into the data warehousing system.

## Big Data Analytics Techniques

Big data can be utilized via many applications and techniques. There's a variety of data types and use cases for big data aggregation and analysis, so it's only natural that collections of data can be used in different ways.

### Popular Big Data Use Cases

#### Artificial Intelligence

There are many use cases of big data that result in an artificial intelligent tool. Artificial Intelligence (also known as Machine Learning) utilizes a mass of data separated into known inputs and outputs. This data is “trained” on a model resembling a simulated network of neurons inspired by the human brain. Once this neural network has been trained on the mass of data, it should be able to predict an output based on new inputs. The output's accuracy is dependent upon the amount and quality of data it was trained upon (Stryker & Kavlakoglu, 2024).

An example of this would be the old Captchas where users would need to “prove they're not a robot”. Many people who interact with the internet often may recall needing to choose pictures of cats from a set of photos to access a website. These Captchas served an additional purpose, which was to train Artificial Intelligence to determine whether a photo contains an image of a cat. When the user chooses the photos, the correct answer is known by the system. A neural network is trained on the data generated by users interacting

with the Captcha, and then the neural network will be able to pick out images that contain a photo of a cat.

This is a similar principle that can be utilized in other systems that collect large amounts of data. This includes techniques like predictive analysis (use of historical data as input into a neural network to predict future behavior and events), data mining (use of machine learning and historical data to discover patterns, correlations, and anomalies in large datasets), and deep learning (a layered version of machine learning for more complex pattern recognition)(Mucci, 2024).

## Natural Language Processing (NLP)

NLP is a technique used to analyze and use human language in text form. NLP is built using deep machine learning and offers many advantages for processing data within a company's big data solution. NLP includes both sentiment analysis and text mining. Sentiment analysis can be used to pull emotional expression from text: typically a tertiary expression of positive, negative, and neutral. Text mining can be used to extract insights from free form text data (Holdsworth, 2024).

These aspects of natural language processing can be instrumental in analyzing data from a system the contains large amounts of unstructured data. A strong example of NLP that is used today is Amazon's compilation of customer reviews. All customer reviews on a product are processed to generate a single consolidated review of general customer sentiment around the product. This way, other customers that want a quick look at the

quality of a product as experienced by individuals who have used it can do so with something more nuanced than a 5-star system.

## Traditional Statistical Analysis

Before there was machine learning, there was good, old fashioned statistical analysis. This practice includes the manual interpretation of data and discovery of trends through queries and dashboards. Three different types of statistical analysis are:

1. Descriptive Statistics – Summarizes data using mathematical techniques like averages, medians, and variances.
2. Inferential Statistics – Predictions about a data set based upon a sample of such data.
3. Regression Analysis – Models relationships between dependent and independent variables (e.g. predicting sales based on social media interaction) (Siavoshi, 2024).

## Techniques Suggested for Proposed System

### Statistical Regression Analysis

With the proposed big data architecture, Heartland Escapes could use data from google reviews, social media platforms, and their own databases to create correlational relationships. A primary example of this would be predicting customer foot traffic based on user interaction with social media sites. Using the data accrued from both the Heartland

Escapes stores and their social media site, it can be determined whether this correlation exists, and if so, taken advantage of.

### Sentiment Analysis of Instagram Comments

For events that are posted on Instagram, the likes and comments can be used to gauge customer sentiment and excitement. Using Natural Language Processing and Sentiment analysis, comments can be analyzed to determine common things that customers are saying about upcoming events and determine a ratio for positive responses to the posted event. Additionally, the view-to-like ratio can also help in this determination for general customer excitement.

### NLP Author Description from Book Reviews

In the e-commerce site, natural language processing can be used to parse customer comments about an author from book reviews. This data can be compiled across reviews of all the author's works, and a general statement about the author's writing style can be concocted. This could be incredibly helpful for Heartland Escapes customers, and also help set Heartland Escapes apart from competing bookstores.

### Training Considerations

These techniques and technologies are incredibly helpful to businesses trying to get the most out of their big data. However, some of these techniques come with massive cost and knowledge requirements. It's important to consider the monetary and time investment



when considering building a business intelligent dashboard or output from a big data solution. The regression analysis of predicting customer foot traffic from Instagram interactions, for example, should be a relatively simple query to build and analyze. This is utilizing traditional techniques for data analysis and comparison.

Even analysis solutions that use machine learning can be a simple task to complete. Consider the example of using Instagram comments to determine general sentiment for an upcoming event. There are pre-built natural language processors that can be used for this, the most obvious for Heartland Escapes would be Google Cloud NLP. Google Cloud NLP is prebuilt and available to businesses that have their data hosted in Google Cloud Platform, and integrates well with ETLBox. The only additional training needed by the data engineer would be to understand how to use Google Cloud NLP, and maybe some training on cleaning the data from the Instagram comments (i.e. stripping out irrelevant comments, removing emojis, etc). This would be an easily achievable task for the Heartland Escapes team without too much additional training.

On the other hand, developing a tool that analyzes all comments across all of an author's works, cherry-picks information about the author from those comments, and compiles them into a single story about the author is a much more complicated task. The amount of employee and machine learn training required has the potential to be enormous, even with a prebuilt NLP.

## Big Data Analytics Value Creation

The value created for Heartland escapes by adding a big data solution has been defined by the problem statement at the beginning of this document. Heartland Escapes desired a way to view the relationship of their day-to-day business affairs from the context of their social media activity. With this big data solution, Heartland Escapes business data and social media data is aggregated into one place. Now they can query information about social media posts and their business processes at the same time, potentially drawing correlations between the two. With enough data, Heartland Escapes may be able to predict increased flow of foot traffic to their store based on social media interactions on event posts, and gauge customer sentiment about their store from comments on these posts.

Before this big data solution was in place, Heartland Escapes didn't have any insight into their data from Instagram. Now that this data is available, they will be able to perform all kinds of queries that could provide conclusions about the efficacy of their social media efforts. For example, they may want to know generally how their sales perform within 48 hours of a social media post compared to other dates. Or how traffic to their e-commerce site is impacted by social media traction. Maybe they can predictively tune their e-commerce service system scaling based on social media post dates from this data.

## Analyst Responsibilities and Maintenance

Considering that this is a whole addition to the existing system, the engineers will have more to maintain with the inception of the new business intelligence solution. However, the job of the business analyst will be simplified. Now that there is a single place where all the businesses valuable data is stored and can be queried upon, the business analyst has much less research to do. They can also easily request that the data warehouse ingests data from other sources, or that queries on other data sources be built directly into Looker dashboards. This should give the analyst much more data to work with, and hopefully new insights about business processes to bring to stakeholders.

The maintenance required for this system should be minimal. That was one of the primary factors in building this model since Heartland Escapes has very few payrolled software engineers. The ETL tool is built on .NET C#, which is familiar to the Heartland Escapes engineering staff. It can be easily containerized and deployed on a Cloud Run instance, which has minimal infrastructure maintenance concerns (strictly CI/CD pipeline development). The choice of Snowflake and Cloud Storage as the warehousing solution also results in minimal infrastructure maintenance. It's hosted in Google Cloud Platform, which will make interservice communication simple to configure. Since Snowflake is multi-cloud and ETLBox is containerized, this part of the system can be easily lifted and shifted to another cloud platform if need be. Finally, looker studio as the dashboarding tool should require minimal maintenance, and the development of looker dashboards is boasted to be simple enough for non-technical folks to build.

## References

- Polivka, A. D. (2024). Heartland Escapes: Database System Overview. Accessed September 8, 2024.
- AWS. (n.d.). What is Structured Data? - Structured Data Explained - AWS. Amazon Web Services, Inc. <https://aws.amazon.com/what-is/structured-data/>. Accessed September 8, 2024.
- AltexSoft Inc. (2023, November 1). Unstructured Data: Examples, Tools, Techniques, and Best Practices. Medium. <https://altexsoft.medium.com/unstructured-data-examples-tools-techniques-and-best-practices-c0fefa57f741#:~:text=Unstructured%20data%20storage%20%20Scalability.%20Unstructured%20data%20has>. Accessed September 8, 2024.
- Terra, J. (2020, May 5). What is a Business Analysis and What does Business Analyst Do. Simplilearn.com. <https://www.simplilearn.com/what-is-a-business-analyst-article>. Accessed September 8, 2024.
- Google. (2024a). Welcome to Looker Studio! - Looker Studio Help. Support.google.com. <https://support.google.com/looker-studio/answer/6283323?hl=en>. Accessed September 15, 2024.
- Google. (2024b, September 10). SQL Server Integration Services (SSIS). Google Cloud. <https://cloud.google.com/sql/docs/sqlserver/ssis>. Accessed September 15, 2024.

Google. (2024c, September 11). Dataflow overview. Google Cloud.

<https://cloud.google.com/dataflow/docs/overview>. Accessed September 15, 2024.

Herzberg, B. (2021, December 19). What Is Snowflake Data Warehouse? A Tutorial | Built In.

Built In. <https://builtin.com/articles/snowflake-data-warehouse>. Accessed September 15, 2024.

Ichhpurani, K. (2019, June 4). Announcing Snowflake on Google Cloud Platform. Google

Cloud Blog; Google Cloud. <https://cloud.google.com/blog/products/data-analytics/announcing-snowflake-on-google-cloud-platform>. Accessed September 15, 2024.

Kenneth, G. (2022, June 1). Top 6 Power BI Benefits For Your Business. Panoply.io; Panoply.

<https://blog.panoply.io/benefits-of-power-bi>. Accessed September 15, 2024.

Lennartz, A. (2024). Quickstart. ETLBox. [https://www.etlbox.net/docs/getting-](https://www.etlbox.net/docs/getting-started/quick-start/)

[started/quick-start/](https://www.etlbox.net/docs/getting-started/quick-start/). Accessed September 15, 2024.

Rogojan, B. (2021, March 1). Snowflake vs BigQuery Pricing and Performance Comparison.

Panoply.io; Panoply. <https://blog.panoply.io/snowflake-vs-bigquery-comparing-pricing-performance-and-usability>. Accessed September 15, 2024.

Snowflake, Inc. (2024). Use the Snowflake Connector for Google Looker Studio | Snowflake

Documentation. Snowflake.com. <https://other-docs.snowflake.com/en/connectors/google-looker-studio-connector>. Accessed September 15, 2024.

Holdsworth, J. (2024, June 6). What is Natural Language Processing? IBM.

<https://www.ibm.com/topics/natural-language-processing>. Accessed September 29, 2024.

Mucci, T. (2024, August 12). What is Predictive AI? IBM.

<https://www.ibm.com/think/topics/predictive-ai>. Accessed September 29, 2024.

Siavoshi, M. (2024, July 22). An Overview of Descriptive vs. Inferential Statistics. Statology.

<https://www.statology.org/an-overview-of-descriptive-vs-inferential-statistics>.

Accessed September 29, 2024.

Stryker, C., & Kavlakoglu, E. (2024, August 16). What is Artificial Intelligence (AI)? IBM.

<https://www.ibm.com/topics/artificial-intelligence>. Accessed September 29, 2024.