

# PROJECT TWO: GAP ANALYSIS & PROPOSAL

Presented by  
Joaquin Esguerra Jr.

# GAP ANALYSIS





# Current and Future States

## Current States

- Kansas City Motors database is not compatible with the rest of the company's database.
- The extracted data displays missing data, errors, and inaccurate values from Kansas City Motors.
- Data is stored in Kansas City Motors defunct AS400 database in scientific format. ◦ The rest of the company uses the SQL database which means Kansas City Motors defunct database is not compatible.

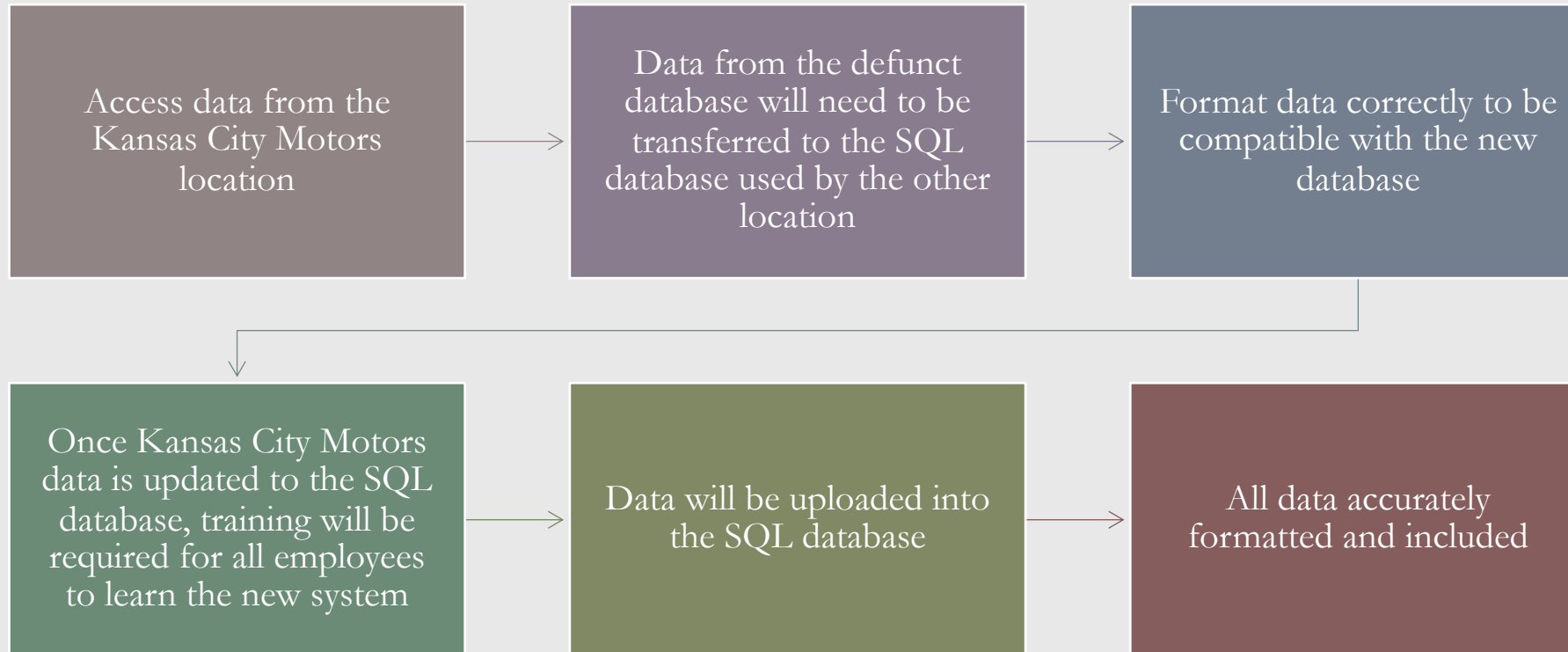
## Future States

- Kansas City Motors data needs to be transformed into an SQL format.
- Data values will need to be inserted for the given blank values, along with error values that need to be replaced to match the format of the rest of the company's data.
- After correctly formatted, the data collected from Kansas City Motors will need to be moved to the companies data warehouse.



# Identification of the Gap

- Due to Kansas City Motors data, that is being stored in a defunct database, it does not align with the other locations databases.
- Missing data values, whether error values or the values in scientific format, from the Kansas CityMotors location.
- Bridging the gap within the data will require the data that is taken from the Kansas City Motors location to be formatted to be compatible with the data from the other locations. This will allow the data to be merged into the data warehouse with the entire company.



# Move From Current State to Future State

# Visualize the Gap Analysis

Kansas City Motors database is not up to the standard with the rest of the company

- Data needs to be taken from the current database and transferred to the SQL database
- While accessing the data errors and blank values were indicated, along with scientific formatted values

Data will need to be reformatted to match the SQL database format

- Utilize given references to ensure all data from the old database is included within the new database
- After formatting the data and checking for accuracy, the data needs to be transferred to the data warehouse

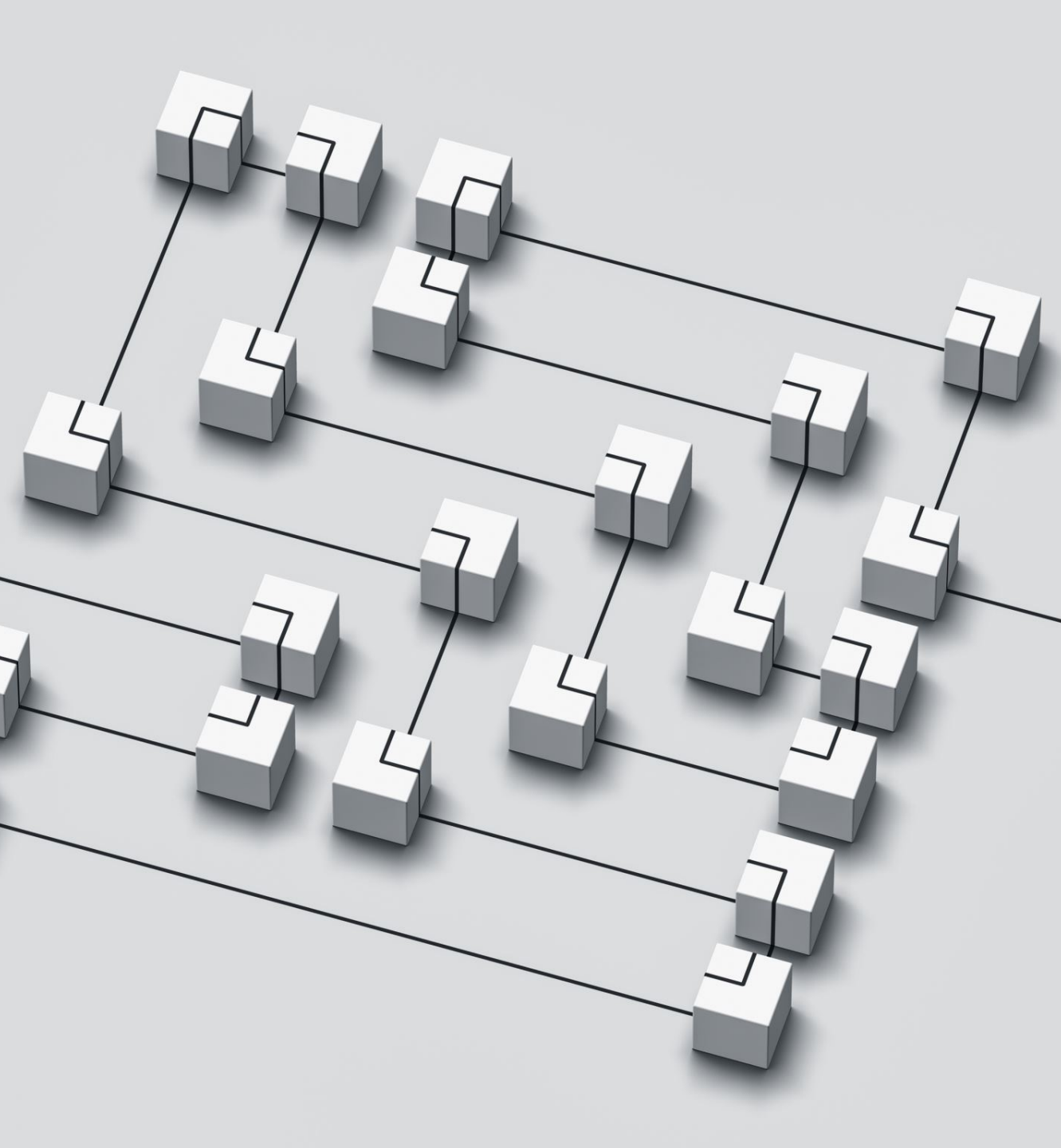
Kansas City Motors current database is a defunct database

- Once formatting the data and checked for accuracy, it needs to be transferred to the data warehouse
- Training will need to be given to all employees for accuracy using the new database



# Summarize the Gap Analysis

- All the company locations need to have access and be using the same data warehouse.
  - This will ensure all data can be easily accessed and provide accurate results.
- All data values need to be checked over and empty values corrected.
  - Inaccurate data will potentially lead to inaccurate results.
- Utilizing the ETL process would be beneficial to ensure that all data is formatted correctly from the Kansas City Motors location, and all other locations for the overall goal of moving the data into one data warehouse.
- The use of the SQL database by the company, requires Kansas City Motors to be equipped with the same database and use the SQL database moving forward.



# ETL (Extract, Transform, and Load) Process

- Extract
  - Copying data from one or more possible sources that could be in various formats. This data would be put in a data depository to maintain data quality.
- Transform
  - Taking the data, then translating it into the format that matches the target system. This includes changing data types, splitting fields/combining, and applying complex formulas. (Kimachia, 2024).
- Load
  - Data that has been transformed will then be loaded into the target system, which could include a database or a data warehouse. Data analyst would then analyze.



# Sensitive or Confidential Information Handling

- With the information given, there seems to be limited security measures put in place to protect the data. However, that is just based on the given information. Protections that have not been disclosed could be in place.
- Provided information contains only motor vehicle sales which I've determined has little importance in terms of security protective needs.
- Information, such as private customer or employee information, there would need to be greater limitations with who has access and how easy it is to be shared within the company or outside of.





# PROPOSAL TO DATA STEWARD

# Issues Identified in Project One

- Kansas City Motors data brought up to the standards and compatible with the SQL database used by the rest of the company.
- The SQL database that is used by the rest of the company requires a specific format, which the defunct database that Kansas City Motors currently uses doesn't meet.
- The data that was taken from Kansas City Motors includes errors, blank values, and inaccurate values.



# Production and Testing Environments

- In the world of cybersecurity, a sandbox environment, which is known as an isolated virtual machine in which unsafe software code can potentially execute without affecting local applications or network resources. (Juviler 2022)
- The sandbox environment would be beneficial for the company to use in the case that a new program is used or anything that could result in issues to the system. A sandbox would be the buffer to ensure nothing happens to the data.
- To allow for better production, a backup system needs to be put in place to store a copy of the data in case of extenuating situations. This would go to work during the ETL process.





## Additional Data Resources

- For the company to thrive, additional data resources would be beneficial for the overall practice of the company.
- Resources, including articles on the uses of the ETL process, could be beneficial to the overall company process
- Access to more data sources will lead to accuracy with the given data and will help the company be successful.

# References

- *The Dama Guide to the Data Management Body of Knowledge*. Guide to the System Engineering Body of Knowledge. (n.d.). [https://sebokwiki.org/wiki/The\\_DAMA\\_Guide\\_to\\_the\\_Data\\_Management\\_Body\\_of\\_Knowledge](https://sebokwiki.org/wiki/The_DAMA_Guide_to_the_Data_Management_Body_of_Knowledge)
- Juviler, J. (2022, June 22). *What is a sandbox environment? [+ how to create one]*. HubSpot Blog. <https://blog.hubspot.com/website/sandbox-environment>
- Kimachia, K. (2024, April 22). *What is ETL? (extract, transform, load): The ultimate guide*. TechRepublic. <https://www.techrepublic.com/article/what-is-etl/>

