Course 3 resources and citations

Module 1: Introduction to data transformation in the cloud

Resources

Helpful resources and tips

- [Common problems with labs/lab troubleshooting](#)

- [Google Docs Help Center](#)

- [Google Sheets Help Center](#)

- [How to use Google Slides](#)

The data journey is a flexible path

- [7 data collection methods in business analytics](#)

Impacts of the data journey on decision making

- [Business analytics in action: Real-world examples and use cases](#)

Techniques Used to Transform Data, Part 2

- [Importance of proper formatting in data collection](#)

Storage options for transformed data

- [BigQuery overview](#)

Citations

Stages of the data journey

- Statistics Canada. (2021, October 22). *Data journey*.

The data journey is a flexible path

- Cymetrix. (2023, January 11). *Business analytics in action: Real-world examples and use cases*. Linkedin.

- Statistics Canada. (2021, October 22). *Data journey*.

- Weinmeister, K. (2020, April 13). From raw data to machine learning model, no coding required. *Google Cloud Blog*.

Impacts of the data journey on decision making

- Cymetrix. (2023, January 11). *Business analytics in action: Real-world examples and use cases*. Linkedin.

- Statistics Canada. (2021, October 22). *Data journey*.

- Weinmeister, K. (2020, April 13). From raw data to machine learning model, no coding required. *Google Cloud Blog*.

Steps for effective data collection

- Chowdhury, S. (2022, June 17). *Concepts of data staging: Interview prep guide*. Linkedin.

- Cote, C. (2021, December 2). 7 data collection methods in business analytics. *Business Insights Blog*.

The process stage ensures clean and consistent data

- Aziza, B. (2021, December 2). Unlocking opportunities with data transformation. *Google Cloud Blog*.

- Choudhury, A. (2021, January 22). *Top 8 data transformation methods*. Analytics India Magazine.

- Linkedin. (n.d.). *What are some common data transformation techniques and when to use them?*

[Supplemental] Process and summarize data

- Purdue University. (n.d.). *Descriptive statistics*. Purdue Online Writing Lab.

- The University of Utah. (n.d.). *Central tendency & variability*.

Techniques Used to Transform Data, Part 2

- Hamilton, T. (2020, March 31). What are Data Unions? How do they work? Which ones can I use? *Streamr*.

- W3 schools. (n.d.). *SQL joins*.

Manual and automated transformation

- Aziza, B. (2021, December 2). Unlocking opportunities with data transformation. *Google Cloud Blog*.

- Google Cloud. (n.d.). *What is ETL?* Retrieved July 20, 2023.

Storage options for transformed data

- Enagandula, R. (2023, May 17). *Database vs data warehouse vs data lake*. Linkedin.

Module 2: Handle raw data with data pipelines

Resources

Overview of data pipeline stages and benefits

- Data Pipelines: An Overview

- What Data Pipeline Architecture should I use?

Strategies and tools for efficient data ingestion

- Big data ingestion

Data transformation steps to ensure reliable data

- [What are the most common data cleaning techniques?](#)

Transformation approaches for securing data

- [FBI Tech Tuesday: Protecting Against Personally Identifiable Information (PII) Theft](#)

- [How to safeguard personally identifiable information](#)

Data dictionaries and monitoring help maintain data integrity

- [Definitive Guide to Data Monitoring](#)

- [What Is a Data Dictionary?](#)

Analysis and visualization after the data pipeline

- [How to Meet WCAG (Quick Reference)](#)

- [Introduction to SQL](#)

- [Top data visualization tools for business professionals](#)

Citations

Data pipelines in the cloud

- Dowsett, C. (2022, December 22). *What is a data pipeline?* Built In.

- Google Cloud. (n.d.). *What is ETL?* Retrieved July 20, 2023.

Overview of data pipeline stages and benefits

- Knight, M. (2023, March 2). Data pipelines: An overview. *DATAVERSITY*.

The difference between ELT and ETL

- Google Cloud. (n.d.). *What is ETL?* Retrieved July 20, 2023.

- Jameel, M. Y. (2022, August 26). *ETL vs ELT?* Linkedin.

Data ingestion methods that suit your needs

- Gomede, E. (2023, August 2). Batch vs. streaming data ingestion: Choosing the right approach for efficient data processing. *Medium*.

- Rajan, V. (2023, July 25). *What is data ingestion and why it is important*. Linkedin.

- Thallam, R. (2020, September 16). BigQuery explained: How to ingest data into BigQuery so you can analyze it. *Google Cloud Blog*.

Strategies and tools for efficient data ingestion

- Gomede, E. (2023, August 2). Batch vs. streaming data ingestion: Choosing the right approach for efficient data processing. *Medium*.

- Rajan, V. (2023, July 25). *What is data ingestion and why it is important*. Linkedin.

Data mapping helps ensure consistent data

- Google Cloud. (2024, January 4). *Data mapping overview*.

Data transformation steps to ensure reliable data

- Aziza, B. (2021, December 2). Unlocking opportunities with data transformation. *Google Cloud Blog*.

Introduction to profiling and cleaning data

- Elgabry, O. (2019, March 2). The ultimate guide to data cleaning. *Towards Data Science*.

- Google for Developers. (2022, July 18). *Representation: Cleaning data*. Machine Learning. Retrieved July 25, 2023.

- Indeed Editorial Team. (2022, October 12). *What is data profiling? (Definition, techniques and types)*. Indeed.

Common ways to manipulate data

- CESSDA ERIC. (n.d.). *File formats and data conversion*. Data Management Expert Guide.

- Kline, L. (2021, October 7). Understanding data enrichment: A comprehensive guide | Hightouch. *The Data Activation Blog*.

- Linkedin. (n.d.). *What are the benefits and challenges of data standardization and normalization?*

- Linkedin. (2024, January 13). *What are the best tools and methods for data conversion and loading?*

Data transformation techniques that achieve different business needs

- Kulshrestha, A. (Ace), & Ellis, S. (2019, June 26). Take charge of your data: How tokenization makes data usable without sacrificing privacy. *Google Cloud Blog*.

- Linkedin. (n.d.). *What are the benefits and challenges of data standardization and normalization?*

- Linkedin. (2023, December 8). *What are some effective strategies for data deduplication in the ETL process?*

Transformation approaches for securing data

- Brennan, B., & Smith, K. (2022, May 24). *FBI Tech Tuesday: Protecting against personally identifiable information (PII) theft*. Federal Bureau of Investigation.

Different approaches to loading data

- Chetry, N. (2023, August 12). Incremental data load approach and implementation strategy. *Medium*.

- Google Cloud. (2024, January 12). *Introduction to loading data*.

Overview of data validation strategies and rules

- Metwalli, S. A. (2023, March 7). *What is data validation?* Built In.

- Nene, N., & Fu, J. (2021, July 21). Introducing the data validation tool for EDW migrations. *Google Cloud Blog*.

Data dictionaries and monitoring can help you maintain data integrity

- Indeed Editorial Team. (2022, September 30). *Definitive guide to data monitoring*. Indeed.

- UC Merced Library. (n.d.). *What is a data dictionary?*

Analysis and visualization after the data pipeline

- Stobierski, T. (2021, January 12). Top data visualization tools for business professionals. *Business Insights Blog*.

- W3 schools. (n.d.). *Introduction to SQL*.

Module 3: Cloud data optimization strategies

Resources

Interview tip: Explain impact

- Interview Warmup

Citations

Challenges of data transformation

- Aziza, B. (2021, December 2). Unlocking opportunities with data transformation. *Google Cloud Blog*.

- Google Cloud. (n.d.). *Google's guide to building a data-driven culture*. Retrieved May 3, 2023.

- Google Cloud. (n.d.). *What is big data?* Retrieved May 4, 2023.

- Google Cloud. (n.d.). *What is ETL?* Retrieved May 3, 2023.

Analyze a data transformation plan

- LeBlanc, D. (2019, June 27). Creating actionable customer segmentation models. *Google Cloud Blog*.

Data aggregation can make it easier to extract meaningful insights

- Google Career Certificates. (2021, June 18). *Aggregating data for analysis* [Video]. YouTube.

- Visiers, I. S. J. (2019, June 11). How to efficiently process both real-time and aggregate data with Cloud Dataflow. *Google Cloud Blog*.

Consequences of duplicate data and how to eliminate it

- Google Cloud. (2024, January 4). *Use the legacy streaming API*.

- Khan, Z. (2021, August 23). Handling duplicate data in streaming pipeline using Pub/Sub & Dataflow. *Google Cloud Blog*.

- Statistics and Data. (n.d.). *How duplicate data can damage the credibility of your statistics!*.

Overview of how joins combine data from different tables

- Looker Studio Help. (n.d.). *How blends work*. Retrieved May 10, 2023.

- Tai, E. (2022, February 7). Tips for joining data—Prevent unwanted data loss and data duplication. *The Data School*.

- Thallam, R. (2020, September 30). BigQuery explained: Working with joins, nested & repeated data. *Google Cloud Blog*.

SQL strategies to join data effectively

- Banhudo, G. (2020, November 27). SQL performance tips #1. *Towards Data Science*.

- Google Cloud. (2024, January 4). *Optimize query computation*.

- W3 schools. (n.d.). *SQL joins*.

Data derivation to combine data and obtain new insights

- Ahmad, I., Stepin, N., & Suliman, S. (2021, October 18). Where data meets IP – Protecting business data in a commercial context. *Data Protection Report*.