



# Data Exploration and Preparation with BigQuery : A Practical Guide to Cleaning, Transforming, and Analyzing Data for Business Insights

Kahn, Mike

Monografía

Leverage BigQuery to understand and prepare your data to ensure that it's accurate, reliable, and ready for analysis and modeling Key Features Use mock datasets to explore data with the BigQuery web UI, bq CLI, and BigQuery API in the Cloud console Master optimization techniques for storage and query performance in BigQuery Engage with case studies on data exploration and preparation for advertising, transportation, and customer support data Purchase of the print or Kindle book includes a free PDF eBook Book Description Data professionals encounter a multitude of challenges such as handling large volumes of data, dealing with data silos, and the lack of appropriate tools. Datasets often arrive in different conditions and formats, demanding considerable time from analysts, engineers, and scientists to process and uncover insights. The complexity of the data life cycle often hinders teams and organizations from extracting the desired value from their data assets. Data Exploration and Preparation with BigQuery offers a holistic solution to these challenges. The book begins with the basics of BigQuery while covering the fundamentals of data exploration and preparation. It then progresses to demonstrate how to use BigQuery for these tasks and explores the array of big data tools at your disposal within the Google Cloud ecosystem. The book doesn't merely offer theoretical insights; it's a hands-on companion that walks you through properly structuring your tables for query efficiency and ensures adherence to data preparation best practices. You'll also learn when to use Dataflow, BigQuery, and Dataprep for ETL and ELT workflows. The book will skillfully guide you through various case studies, demonstrating how BigQuery can be used to solve real-world data problems. By the end of this book, you'll have mastered the use of SQL to explore and prepare datasets in BigQuery, unlocking deeper insights from data. What you will learn Assess the quality of a dataset and learn best practices for data cleansing Prepare data for analysis, visualization, and machine learning Explore approaches to data visualization in BigQuery Apply acquired knowledge to real-life scenarios and design patterns Set up and organize BigQuery resources Use SQL and other tools to navigate datasets Implement best practices to query BigQuery datasets Gain proficiency in using data preparation tools, techniques, and strategies Who this book is for This book is for data analysts seeking to enhance their data exploration and preparation skills using BigQuery. It guides anyone using BigQuery as a data warehouse to extract business insights from large datasets. A basic understanding of SQL, reporting, data modeling, and transformations will assist with understanding the topics covered in this book

**Título:** Data Exploration and Preparation with BigQuery A Practical Guide to Cleaning, Transforming, and Analyzing Data for Business Insights

**Edición:** 1st ed

**Editorial:** Birmingham Packt Publishing, Limited 2023 2023

**Descripción física:** 1 online resource (264 pages)

**Contenido:** Cover -- Copyright -- Contributors -- Table of Contents -- Preface -- Part 1: Introduction to BigQuery -- Chapter 1: Introducing BigQuery and Its Components -- Technical requirements -- What is BigQuery? -- Understanding how BigQuery works -- Dremel, the execution engine -- Colossus distributed storage -- The Borg compute platform -- Jupiter network infrastructure -- BigQuery administration and access -- Tools for administration -- Understanding identity and access management -- BigQuery best practices and cost management -- Best practices -- Understanding and controlling costs -- Extending your data -- BigQuery ML -- External datasets -- External connections -- Summary -- References -- Chapter 2: BigQuery Organization and Design -- Technical requirements -- Understanding BigQuery's resource hierarchy -- Organizations, folders, and projects -- BigQuery-specific resources -- BigQuery storage -- Exploring architecture patterns -- The centralized enterprise data warehouse -- The decentralized data warehouse -- The cross-org data exchange -- Schema design -- Table design -- Summary -- Part 2: Data Exploration with BigQuery -- Chapter 3: Exploring Data in BigQuery -- Technical requirements -- What is data exploration? -- Fundamentals -- Data life cycle -- Common challenges and solutions -- Introduction to exploring data in BigQuery -- Exploring data in the BigQuery SQL workspace -- Exploring schema and table structure -- Exploring data using SQL -- Exploring data using the bq command-line interface -- Exploring data with visualization tools -- Enhancing data exploration in BigQuery -- Advanced approaches -- Best practices -- Summary -- Chapter 4: Loading and Transforming Data -- Technical requirements -- Exploring data loading techniques -- Batch loading data -- Streaming ingestion of data -- Scheduled loading of data Situations where you do not need to load data -- Data transformation with BigQuery -- Evaluating ETL and ELT approaches for data integration -- Hands-on exercise - data loading and transformation in BigQuery -- Understanding the scenario -- Loading data from a local file -- Preparing and transforming data -- Summary -- Chapter 5: Querying BigQuery Data -- Technical requirements -- Understanding query structure -- Action command - the SELECT clause -- Location command - the FROM clause -- Filtering command - the WHERE clause -- Selection handling commands - the GROUP BY, HAVING, ORDER BY, and LIMIT clauses -- Understanding data types -- Using expressions and aggregations -- Expressions -- Aggregations -- Joining tables -- Inner joins -- Outer joins -- Using functions -- Advanced querying techniques -- Subqueries -- Window functions -- Common table expressions -- Array functions -- Saving, sharing, and scheduling queries -- Optimizing queries -- Troubleshooting queries -- Summary -- Further reading -- Chapter 6: Exploring Data with Notebooks -- Technical requirements -- Understanding the value of using notebooks -- Jupyter notebooks -- Using Workbench notebook instances in Vertex AI -- Creating a managed notebook -- Executions and schedules -- Hands-on exercise - analyzing Google Trends data with Workbench -- Using Colab notebooks -- Comparing Workbench instances and Colab -- Summary -- Further reading -- Chapter 7: Further Exploring and Visualizing Data -- Technical requirements -- Understanding data distributions -- Why is it important to understand data distributions? -- Uncovering relationships in data -- Exploring BigQuery data with Google Sheets -- Connecting to Sheets from BigQuery using Explore with Sheets -- Connecting to BigQuery using Connected Sheets -- Column statistics -- Collaboration with BigQuery data in Sheets Visualizing BigQuery data using Looker Studio -- Creating the right visualizations -- Hands-on exercise - creating visualizations with Looker Studio -- Commonly created charts -- Calculations in visualization tools -- Data quality discovery while creating visualizations -- Filtering data in visualizations -- Integrating other visualization tools with BigQuery -- Summary -- Further reading -- Part 3: Data Preparation with BigQuery -- Chapter 8: An Overview of Data Preparation Tools -- Technical requirements -- Getting started with data preparation -- Clearly defining your data preparation goals -- Evaluating your current data quality -- Data cleaning and transformation -- Validating prepared data -- Data preparation approaches -- Data preparation tools -- Visual data preparation tools -- Query and code-based tools -- Automated data preparation -- Summary -- Further reading -- Chapter 9: Cleansing and Transforming Data -- Technical requirements -- Using ELT for cleansing and transforming data -- Assessing dataset integrity -- The shape of the dataset -- Skew of the dataset -- Data profiling -- Data validation -- Data visualization -- Using SQL for data cleansing and transformation -- SQL data cleansing strategies and examples -- SQL data transformation strategies and examples -- Writing query results -- Using Cloud Dataprep for visual cleansing and transformation --

Summary -- Further reading -- Chapter 10: Best Practices for Data Preparation, Optimization, and Cost Control -- Technical requirements -- Data preparation best practices -- Understanding your data and business requirements -- Denormalizing your data -- Optimizing schema design -- Considering nested and repeated fields -- Using correct data types -- Data cleansing and validation -- Partitioning and clustering -- Optimizing data loading -- Best practices for optimizing storage Long-term and compressed storage -- Cross-cloud data analytics with federated access model and BigQuery Omni -- Backup and recovery -- Best practices for optimizing compute -- Analysis cost options -- Query optimization -- Query optimization cheat sheet -- Monitoring and controlling costs -- Query plan and query performance insights -- Monitoring, estimating, and optimizing costs -- Controlling costs -- Summary -- Further reading -- Part 4: Hands-On and Conclusion -- Chapter 11: Hands-On Exercise - Analyzing Advertising Data -- Technical requirements -- Exercise and use case overview -- Loading CSV data files from local upload -- Data preparation -- Standardizing date formats -- Data exploration, analysis, and visualization -- Analyzing ads and sales data -- Return on ad spend -- Visualizations -- Summary -- References -- Chapter 12: Hands-On Exercise - Analyzing Transportation Data -- Technical requirements -- Exercise and use case overview -- Loading data from GCS to BigQuery -- Uploading data files to Google Cloud Storage -- Loading data into BigQuery -- Data preparation -- Data exploration and analysis -- Visualizing data with BigQuery geography functions -- Summary -- Further reading -- Chapter 13: Hands-On Exercise - Analyzing Customer Support Data -- Technical requirements -- Exercise and use case overview -- Data loading from CSV upload -- Data preparation -- Data exploration and analysis -- Count of ticket\_type across both datasets -- The most common support issues using ticket\_subject data -- Average resolution time per ticket\_type -- Customer demographics using customer\_age and customer\_gender -- Analyzing emotions with sentiment analysis -- Creating a connection -- Granting access to the external connection service account -- Creating a model -- Querying the model -- Summary -- References and further reading Chapter 15: Summary and Future Directions -- Summary of key points -- Chapter 1, Introducing BigQuery and Its Components -- Chapter 2, BigQuery Organization and Design -- Chapter 3, Exploring Data in BigQuery -- Chapter 4, Loading and Transforming Data -- Chapter 5, Querying BigQuery Data -- Chapter 6, Exploring Data with Notebooks -- Chapter 7, Further Exploring and Visualizing Data -- Chapter 8, An Overview of Data Preparation Tools -- Chapter 9, Cleansing and Transforming Data -- Chapter 10, Best Practices for Data Preparation, Optimization, and Cost Control -- Chapter 11, Hands-On Exercise - Analyzing Advertising Data -- Chapter 12, Hands-On Exercise - Analyzing Transportation Data -- Chapter 13, Hands-On Exercise - Analyzing Customer Support Data -- Future directions -- More integration with AI and ML -- Generative AI -- Natural language queries -- DataOps -- Hybrid and multi-cloud data analysis -- Zero-ETL and real-time analytics -- Data governance and privacy -- Federated learning -- Data clean rooms -- Data monetization -- Additional resources -- Final words -- Index -- Other Books You May Enjoy

**ISBN:** 1-80512-342-4

**Materia Título preferido:** Google BigQuery Google Analytics

**Materia:** Data mining

**Enlace a formato físico adicional:** Print version Kahn, Mike. Data Exploration and Preparation with BigQuery Birmingham : Packt Publishing, Limited,c2023

---

## Baratz Innovación Documental

- Gran Vía, 59 28013 Madrid
- (+34) 91 456 03 60
- [informa@baratz.es](mailto:informa@baratz.es)