

Student name: Youssef Ashraf

Submission date:

Table of Contents

Section 1: Data Warehouse Fundamentals.....	2
Section 2: Data Lakehouse Fundamentals	6
Section 3: Practical project: Modeling, Building a Data Warehouse and BI Reports for Gravity Bookstore.....	8
Use-case Background.....	8
Requirements.....	9

To be filled by ITI:

Reviewed by:

Review date:

Grade:

Section 1: Data Warehouse Fundamentals

Multiple Choice Questions

1. Which of the following tools is used to help an organization towards building and enabling both Business Intelligence (BI) and advanced analytics use-cases?
 - A) Data Lakehouse
 - B) Data Lake
 - C) Data Warehouse
 - D) All of the above**
 - E) A & C
2. A data warehouse is which of the following?
 - A) Can be updated by end users
 - B) Contains numerous naming conventions and formats
 - C) Organized around business subject areas**
 - D) Contains only current data
3. _____ are slice-and-dice techniques that allow you to view multidimensional information from different perspectives.
 - A) OLAP**
 - B) OLTP
 - C) A database
 - D) An operational database
4. What is a subset of a data warehouse in which only a focused portion of the data warehouse information is kept?
 - A) Data mining tool
 - B) Data mart**
 - C) Data model
 - D) None of the above
5. What do integrity constraint rules help you ensure?
 - A) Governance of the data
 - B) Quantity of the data**
 - C) Quality of the data
 - D) Performance tuning
6. A goal of data mining includes which of the following?
 - A) To explain some observed event or condition
 - B) To confirm that data exists
 - C) To explore and categorize data with new patterns
 - D) To create a new data warehouse
 - E) A & C**
 - F) All of the above

7. All the features below describing the data warehouse, except:
- A) Subject-oriented
 - B) Time variant
 - C) Encapsulated
 - D) Non-volatile
8. Which of the following describes a data warehouse?
- A) Collection of information
 - B) Information gathered from many different knowledge workers
 - C) Used as a source for analytical processing
 - D) All of the above
9. A database that supports OLTP is often called a(n) _____ database.
- A) OLTP
 - B) Operational
 - C) Production
 - D) Working
10. _____ is the manipulation of information to support decision making.
- A) Online transaction processing (OLTP)
 - B) Operational database
 - C) Online data warehouse
 - D) Online analytical processing (OLAP)
11. Which of the following is supported by an operational database?
- A) Online transaction processing (OLTP)
 - B) Online analytical processing
 - C) Online checking
 - D) Online research processing (OLAP)
12. A multidimensional representation of information is referred to as a:
- A) Database
 - B) Data warehouse
 - C) Hypercube
 - D) Metadata
13. When is it appropriate to 'de-normalize' a relational database design for a Data Warehouse database?
- A) When disk space is low.
 - B) When memory is low.
 - C) When the analysis requirements are understood.
 - D) When the database design is no longer expected to change.

14. What is an Operational Data Store (ODS)?

- A) A set of databases that serve as a 'staging' area to facilitate consolidating data from several, distributed-source systems.
- B) A set of databases that support OLAP.
- C) A set of databases that support reporting from an application system.
- D) A set of databases that provide integrated operations data to serve the organization's day-to-day activities.

15. What is a 'snowflake' schema?

- A) The dimension tables are 'normalized'.
- B) The dimension tables can refer to more than one fact table.
- C) All recurring groups of attributes are completely removed from dimension tables.
- D) A schema that can be implemented only with an MDDB Database Management

16. The process of migrating the data from a source to other destination is known as:

- A) Extract, Transform and Load (ETL)
- B) Move, Transform and Load (MTL)
- C) Extract, Transform and Save (ETS)
- D) Extract, Migrate and Load (EML)

17. Which of the following is a business benefit of a Data Warehouse?

- A) Customers are happier
- B) Reduction in governance interference
- C) Decision makers will be able to make more decisions each day
- D) Improves morale of the business analysts

18. The two major tasks of the transformation are:

- A) Data replication and redundancy
- B) Data quality and cleansing
- C) Data classification and verification
- D) Data stability and modeling

19. The dimension table in the data warehouse stores:

- A) Business measures
- B) Data about business perspectives
- C) Metadata
- D) Data dictionary

20. The fact table in the data warehouse stores:

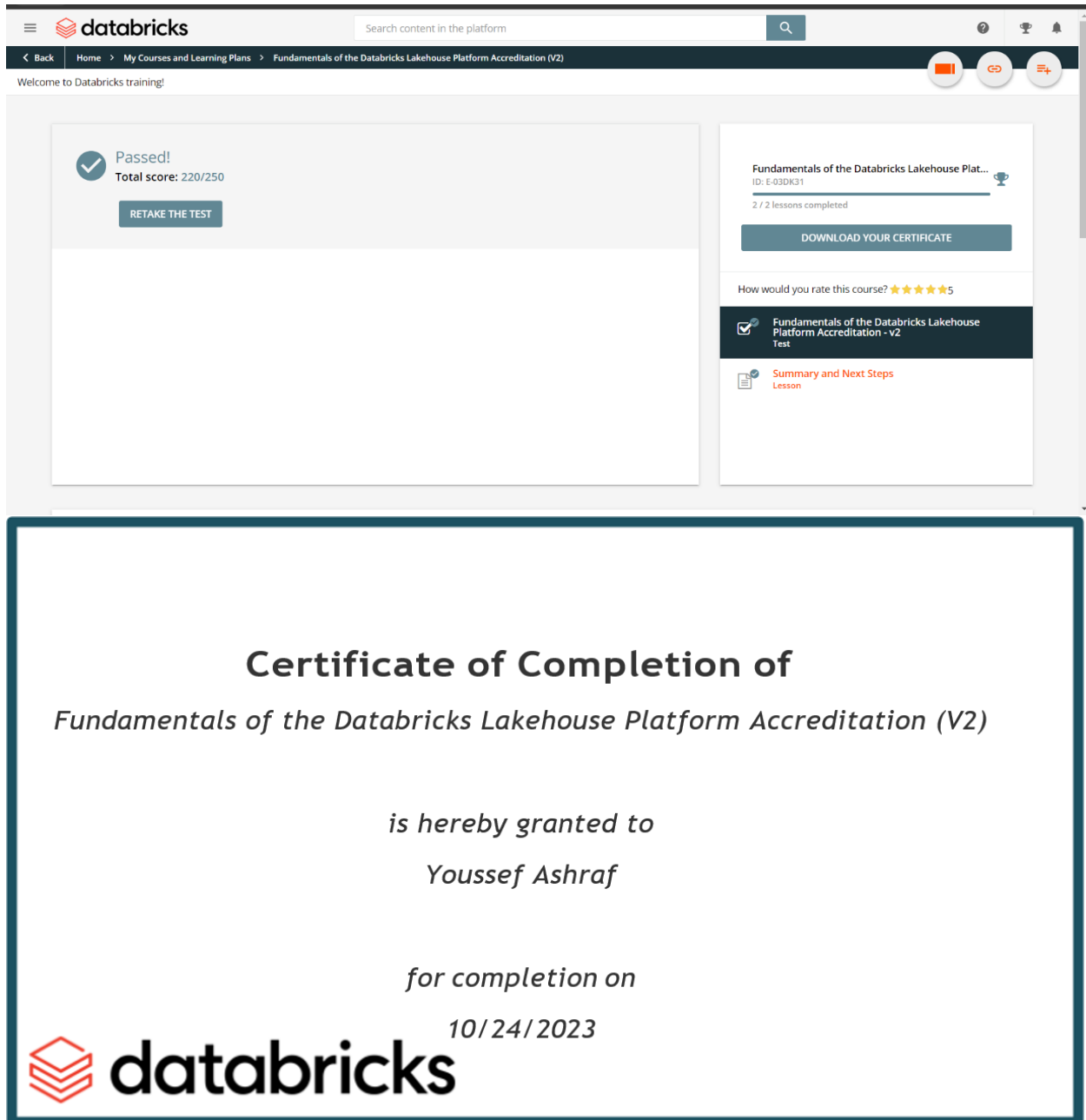
- A) Metadata
- B) Data about business perspectives
- C) Business measures
- D) Data dictionary

21. Which of the following is NOT a Data Warehouse process:
- A) Change Data Capture (CDC)
 - B) Transaction completion on operational systems
 - C) Data cleansing
 - D) Load the data warehouse and populate data marts
22. Which of the following data type is the counterpart of 'Reference Data':
- A) Metric Data
 - B) Derived Data
 - C) Event Data
 - D) Derived Data
23. Which type of data transformation is most likely to create new information from existing data:
- A) Derivation
 - B) Summarization
 - C) Translation
 - D) Conversion
24. Source-to-Target mapping is a prerequisite activity for:
- A) Choosing data sources
 - B) Design of target database
 - C) Design of data transformation
 - D) Data warehouse auditing
25. Which of the following is NOT a significant consideration of Data Architecture for a DWH:
- A) Data flow
 - B) Data structure
 - C) Data purging and archiving
 - D) Data integration

Section 2: Data Lakehouse Fundamentals

Please create a community account and conduct the Databricks Lakehouse Fundamentals Accreditation¹:

[Databricks Accredited Lakehouse Platform Fundamentals | Databricks](#)



Passed!
Total score: 220/250
[RETAKE THE TEST](#)


Fundamentals of the Databricks Lakehouse Plat...
ID: E-03DK31
2 / 2 lessons completed
[DOWNLOAD YOUR CERTIFICATE](#)

How would you rate this course? ★★★★★5

Fundamentals of the Databricks Lakehouse Platform Accreditation - v2
Test
[Summary and Next Steps](#)
Lesson

Certificate of Completion of
Fundamentals of the Databricks Lakehouse Platform Accreditation (V2)

is hereby granted to
Youssef Ashraf

for completion on
10/24/2023
 **databricks**

¹ Hint: certificate issuing may take up to 24 hours after assessment completion until you have it via email.

Provide below a screenshot of the certificate or its URL from: <https://credentials.databricks.com/>

Section 3: Practical project: Modeling, Building a Data Warehouse and BI Reports for Gravity Bookstore

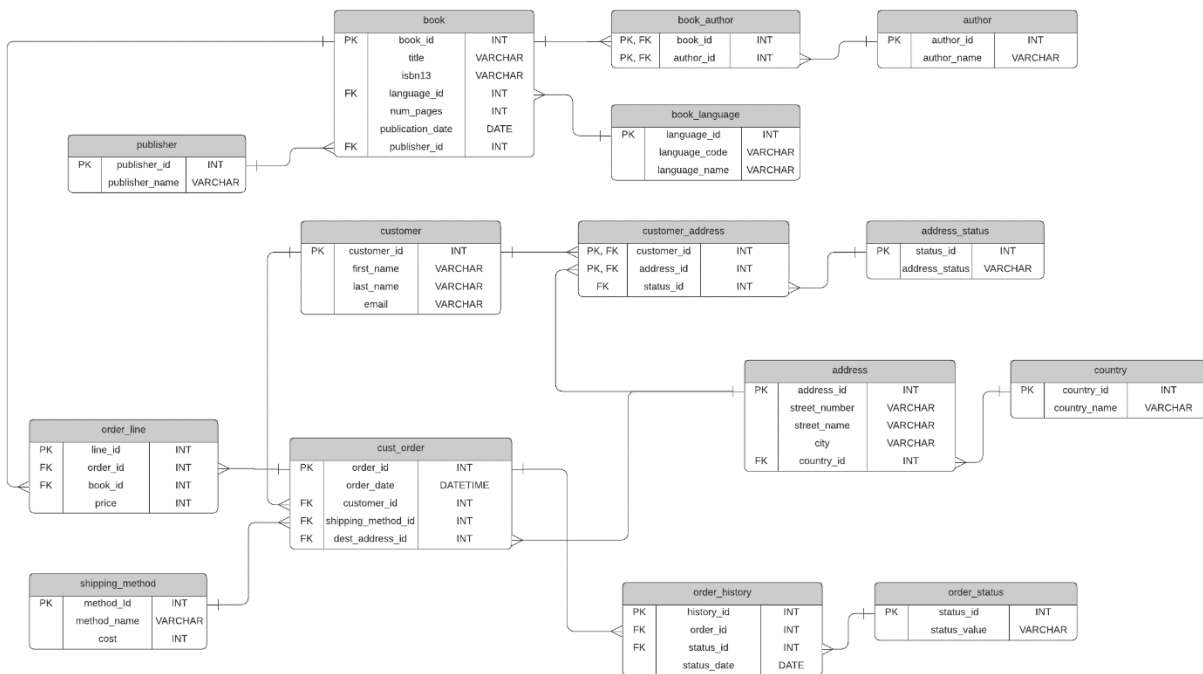
Use-case Background

Gravity Bookstore is a database for a fictional bookstore called that captures information about books, customers, and sales.

Please download the SQL files from the link below and run them in a new database named 'gravity_books' in your local SQL Server DB engine:

[databasestar/sample_databases/sample_db_gravity/gravity_sqlserver](https://github.com/databasestar/sample_databases/sample_db_gravity/gravity_sqlserver) · GitHub

ERD of 'gravity_books' transactional database:



Tables description:

- **book**: a list of all books available in the store.
- **book_author**: stores the authors for each book, which is a many-to-many relationship.
- **author**: a list of all authors.
- **book_language**: a list of possible languages of books.
- **publisher**: a list of publishers for books.
- **customer**: a list of the customers of the Gravity Bookstore.
- **customer_address**: a list of addresses for customers, as a customer can have more than one address, and an address has more than one customer.
- **address_status**: a list of statuses for an address, because addresses can be current or old.
- **address**: a list of addresses in the system.

- **country:** a list of countries that addresses are in.
- **cust_order:** a list of orders placed by customers.
- **order_line:** a list of books that are a part of each order.
- **shipping_method:** the possible shipping methods for an order.
- **order_history:** the history of an order, such as ordered, cancelled, delivered.
- **order_status:** the possible statuses of an order.

Requirements

1. Model and develop 'gravity_books_dwh' Data Warehouse **figure1**
 - a. Please provide the DDL statements of table creation and a screenshot from the DWH ERD.(

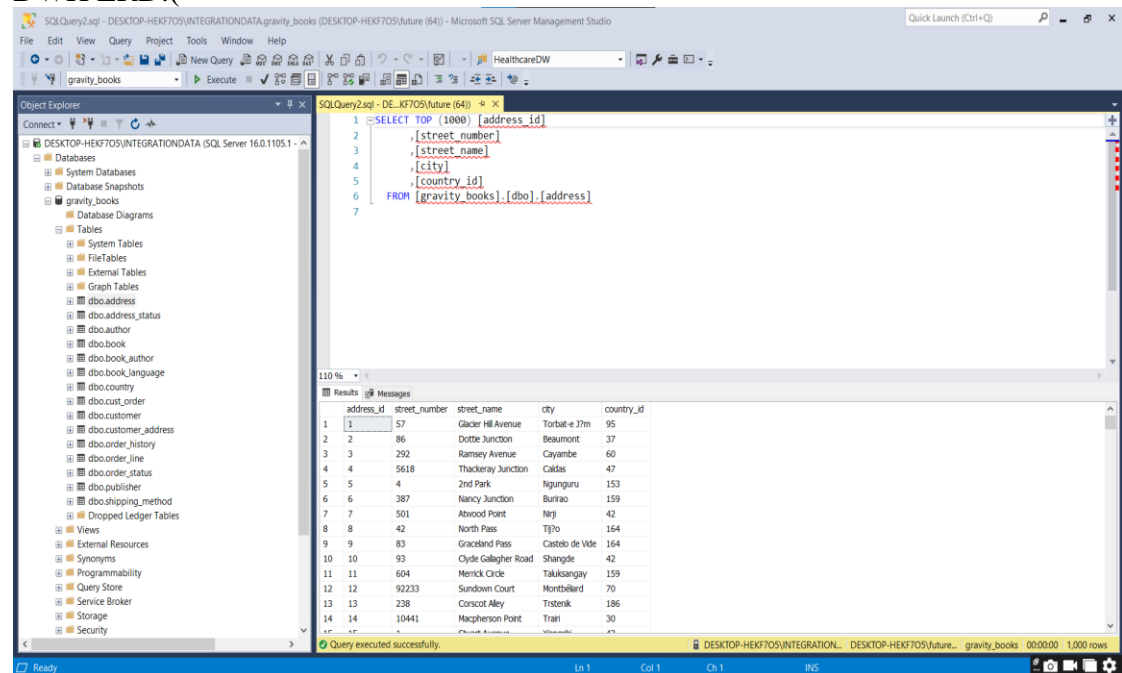


Figure 1

2. Define which approach (star schema, snowflake) of data warehouse used in your solution, and why.

I have used a star schema approach for the data warehouse. A star schema is a simple and efficient way of organizing data for analysis and reporting. It consists of a single fact table that contains the measures of interest and multiple dimension tables that describe the attributes of the fact table. That's what I do one fact table **order_fact** and all dimension tables this is a star schema **figure2**



Figure 2

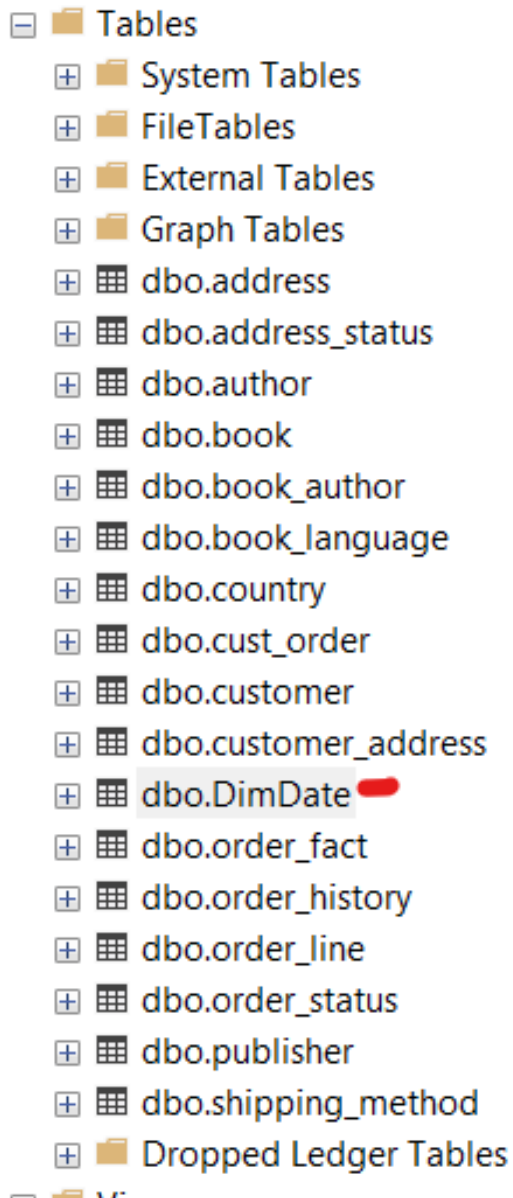
It reduces the number of joins required to query the data, which improves performance and simplifies queries.

It avoids redundancy and inconsistency in the dimension tables, which improves data quality and integrity.

It is easy to understand and maintain by business users and analysts.

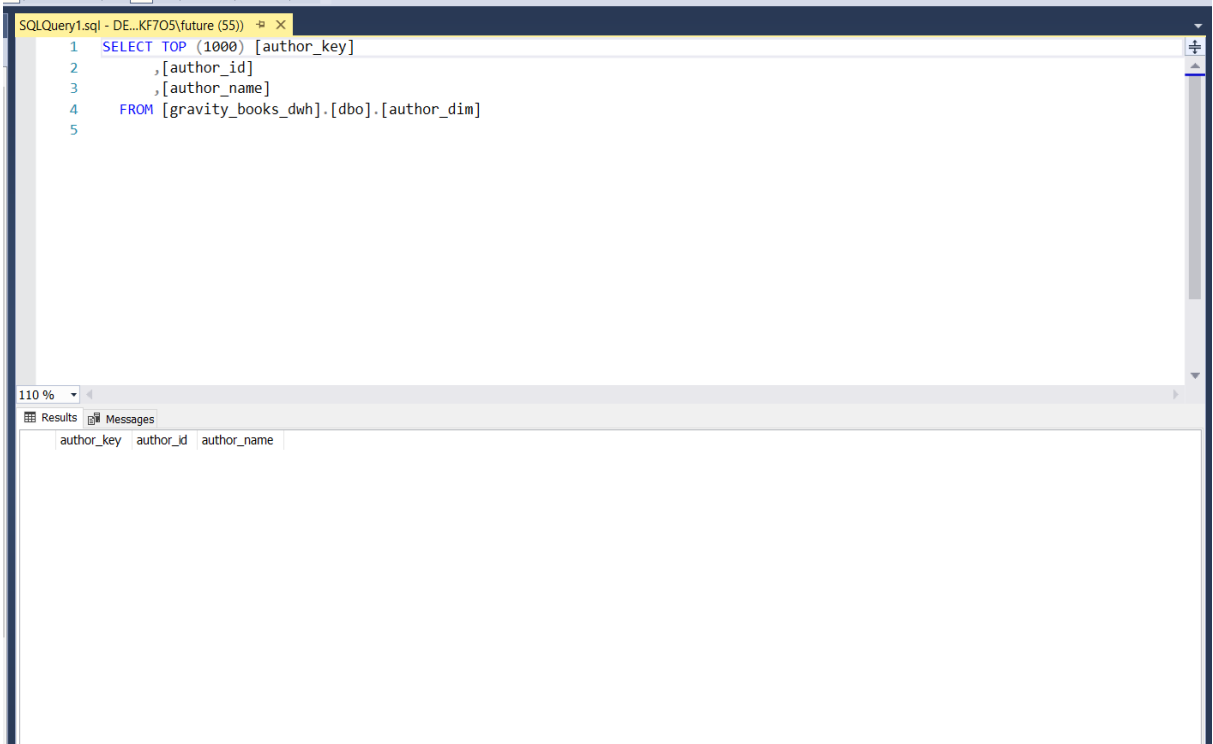
3. Define a method to check and maintain the integrity between the fact and the dimensions (SQL).

4. Please, add a date dimension to the system to track the historical changes.

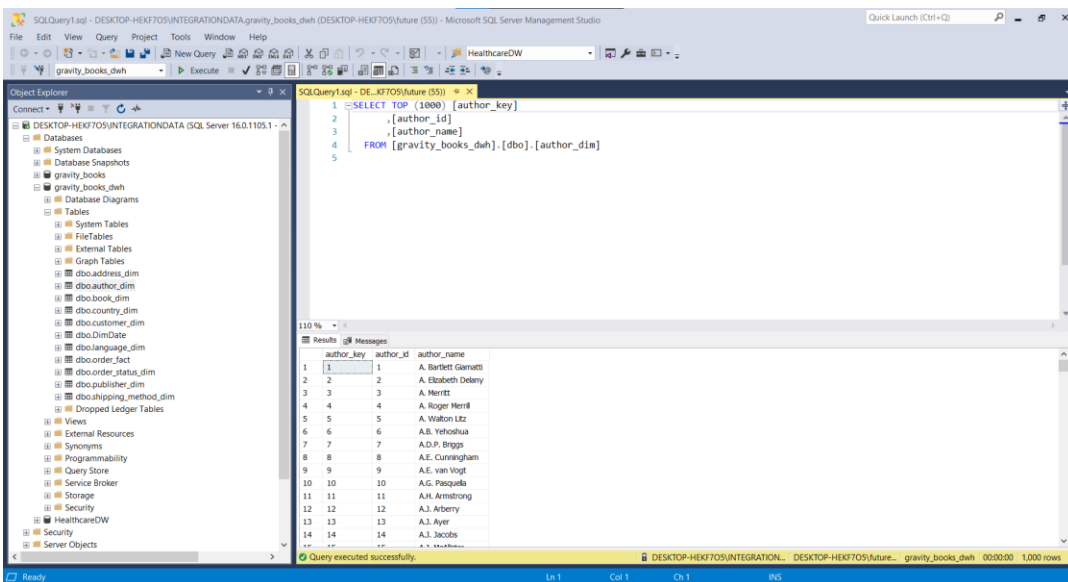


- Design an SSIS project to populate the data from 'gravity_books' transactional database into the new target data warehouse 'gravity_books_dwh'.

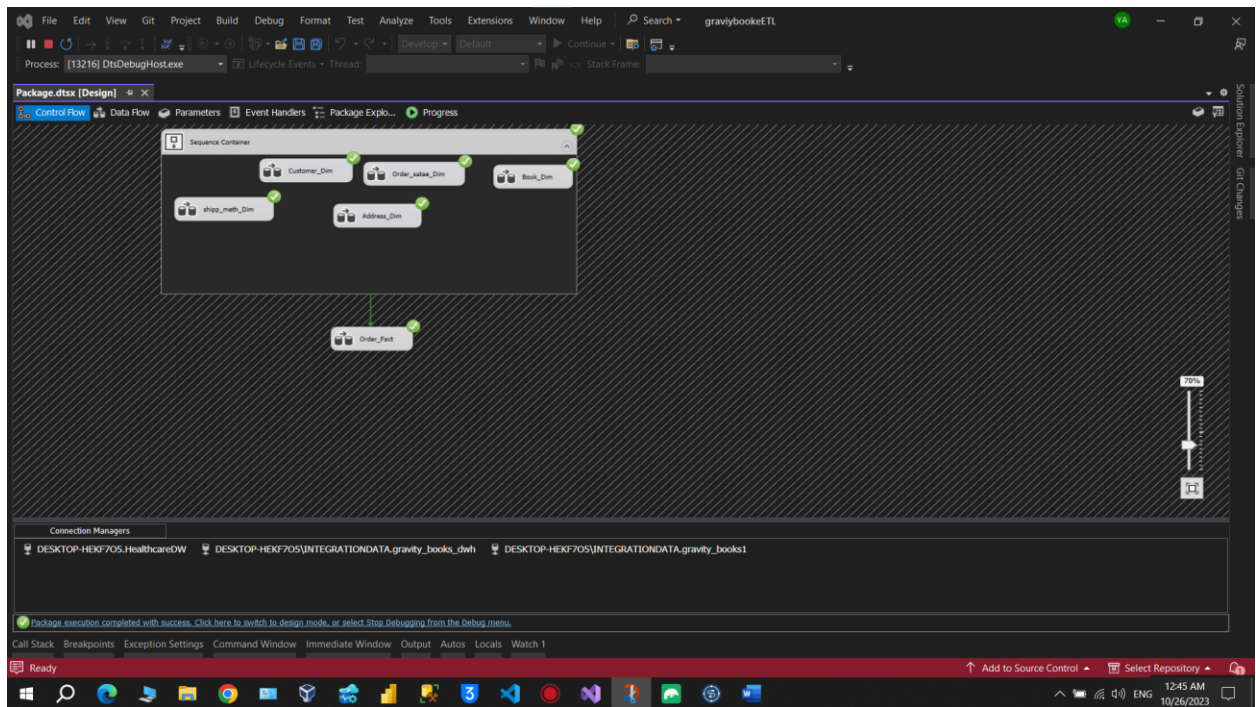
Figure 3 show the DWH empty not transfer Data from source gravity_books to destination gravity_books_DWH



Data Transfer successfully from source to destination In Figure4

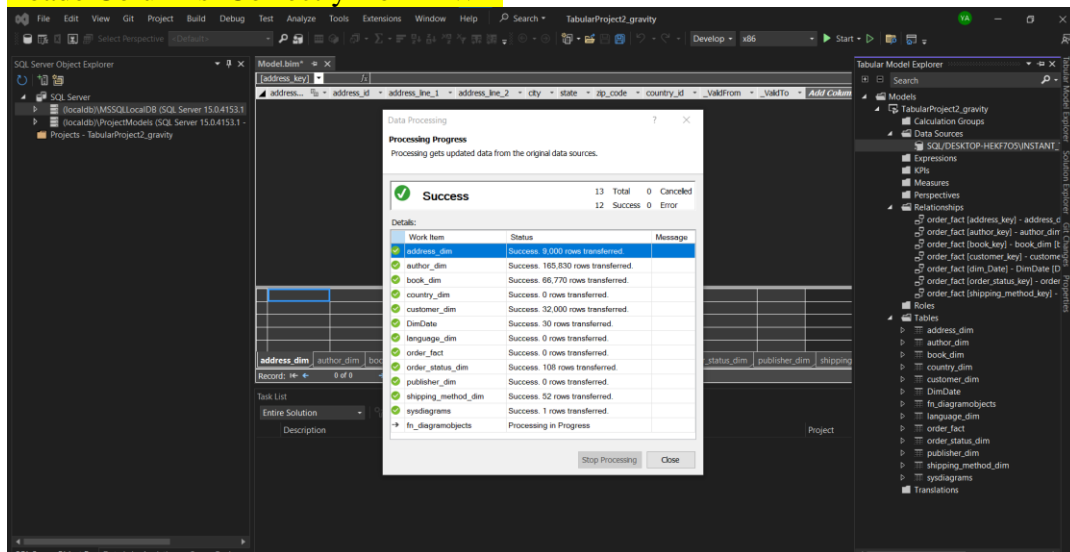


All Dimensions And Fact Tables The Solution In SSIS

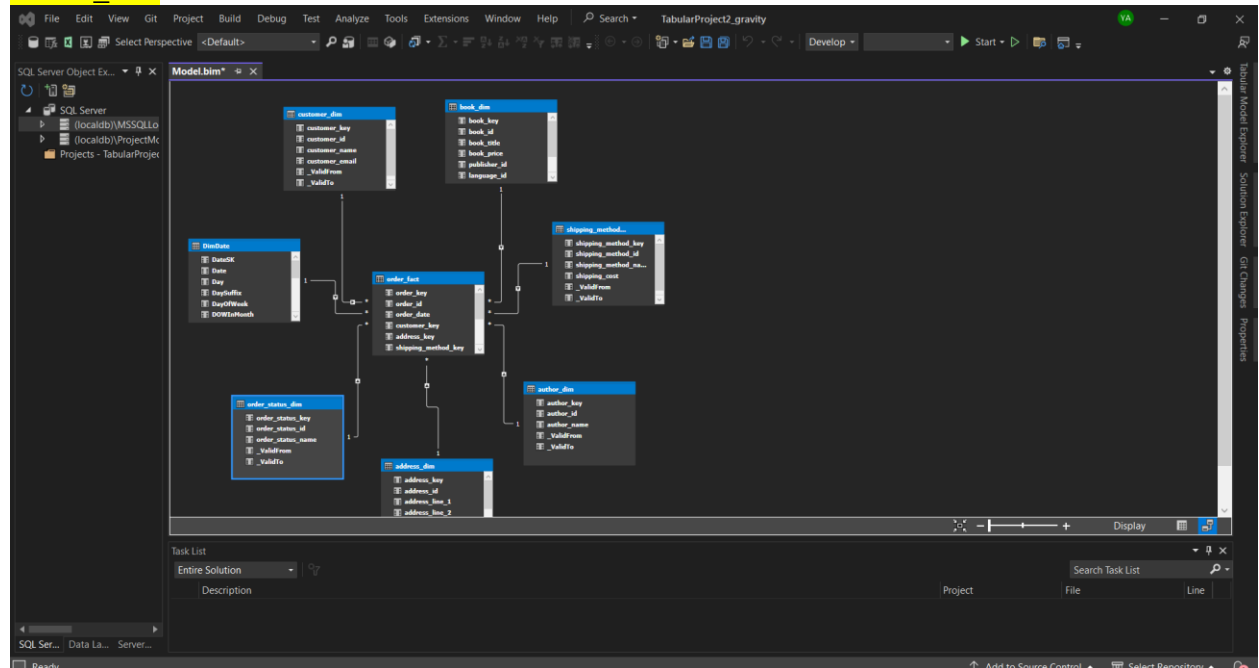


- Design an SSAS project (Tabular mode) and provide the main deliverables of the cube browsing.

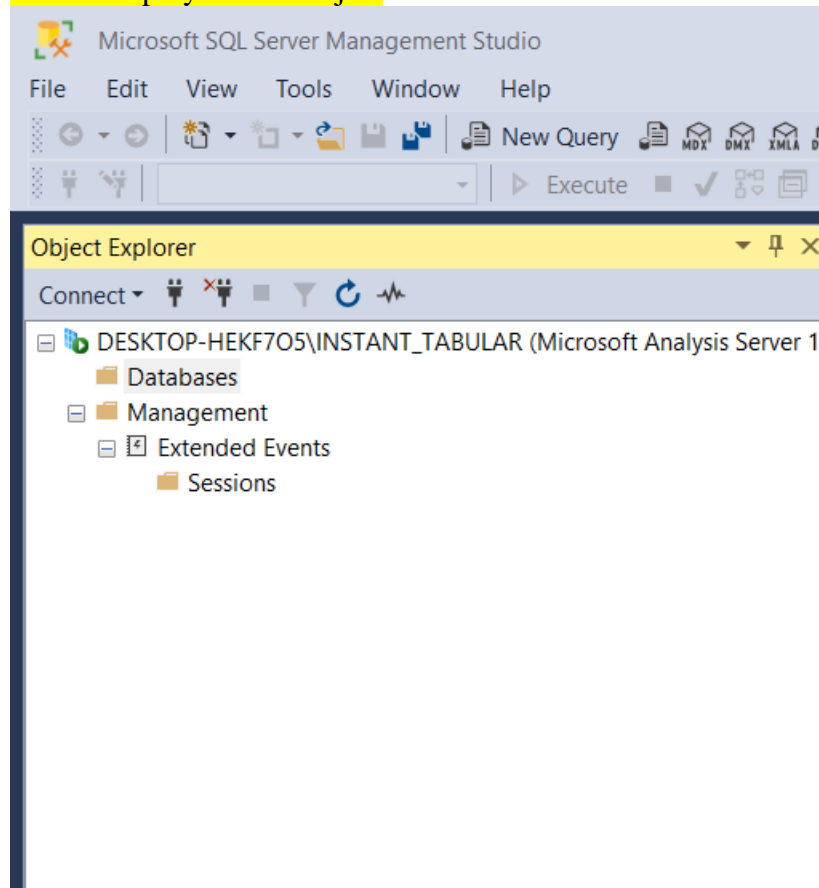
Load Columns Correctly from DWH



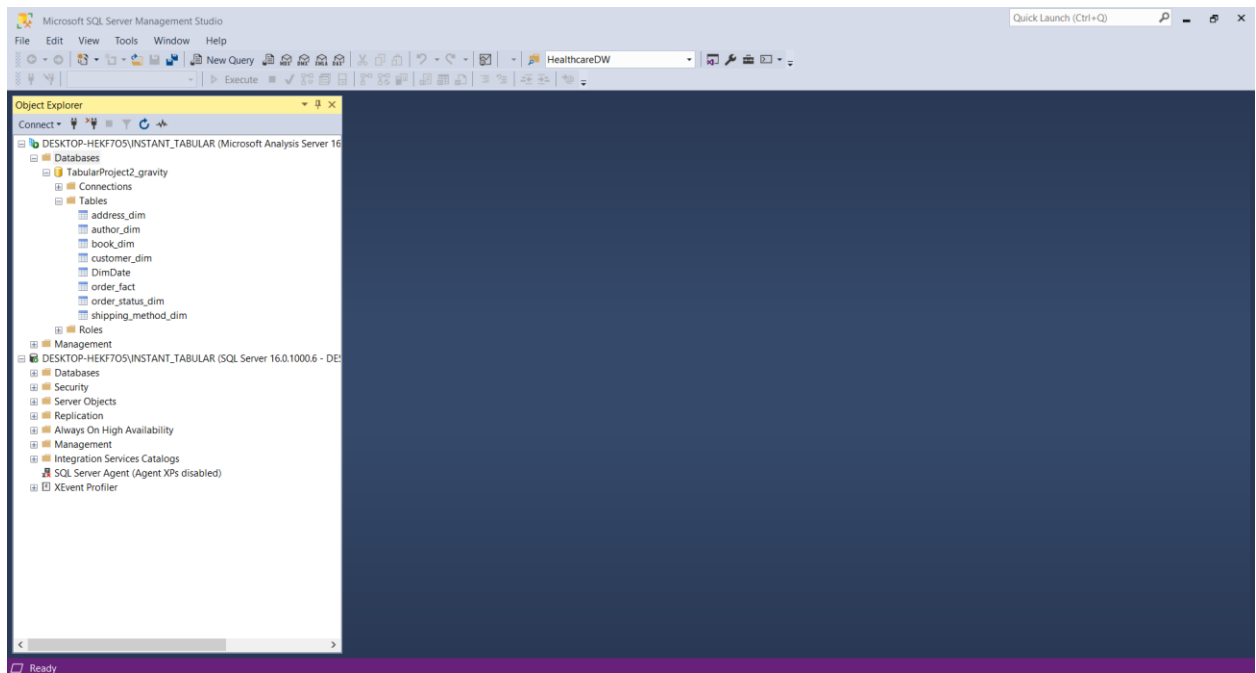
Model_dim



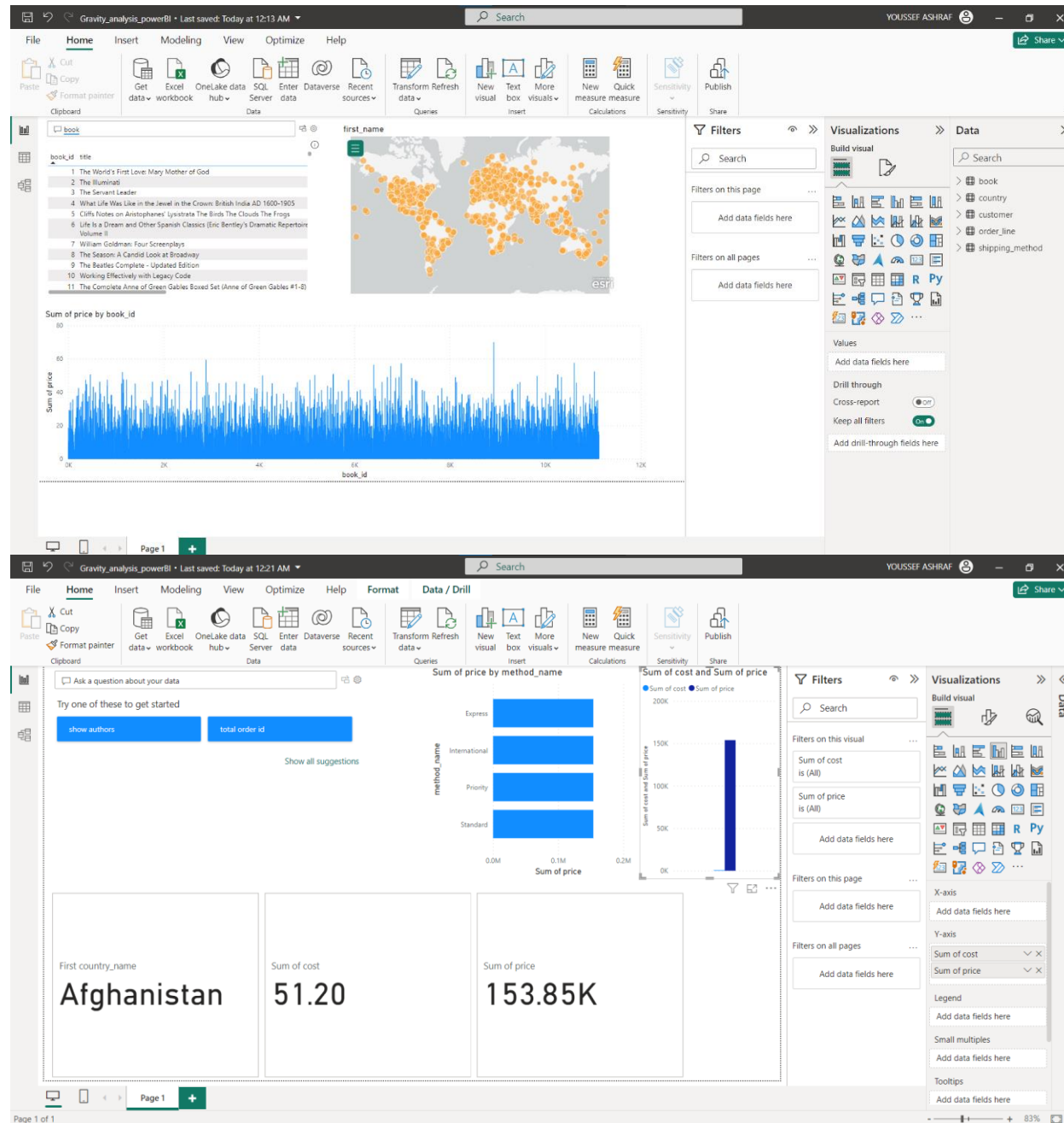
Before Deploy SSAS Project



After Deploy SSAS Tabular MODE



7. Using either PowerPivot in Excel or Power BI Desktop, create BI self-service reporting for exploring gravity books OLAP cube.
 - a. Please use the self-service canvas to explore the data and get business insights from it.
 - b. Create sample analytical reports and one dashboard.



- Please provide screenshots from the output of each point above:
- Compress the entire solution files:
 - DWH DDL statements (format .sql)
 - ETL SSIS project
 - OLAP SSAS project
 - PowerPivot Excel or Power BI (format .xlsx or .pbix)