**Yêu cầu lựa chọn câu hỏi**

1. Tạo 2 bảng, mỗi bảng 15 câu hỏi tương ứng với chủ đề chuỗi 2 buổi webinar   
   Buổi 1: Model the data (25–30%)  
   Buổi 2: Prepare the data (25–30%)
2. Các yêu cầu chuyên môn khác

Buổi 1:

* 5 câu lý thuyết: Câu hay dễ sai
* 5 câu DAX practical: Có tình huống
* 5 câu Modeling practical: Có tình huống, dễ sai, gồm Cardinality, Active/ Inactive Relationship, ...

Buổi 2:

* 4 câu Data Source
* 4 câu Clean & Transform
* 3 câu Append/ Merge Queries
* 4 câu xử lý lỗi và tối ưu Power Query

Ưu tiên: Ghép 2–3 câu để thành một tình huống thực tế → tăng sự liền mạch

**Bảng câu hỏi**

Bảng 1 (15 câu hỏi buổi 1) | Deadline: 16/05  
Bảng 2 (15 câu hỏi buổi 2) | Deadline: 19/05

★: Câu dễ ★★: Câu vừa ★★★: Câu khó

**BUỔI 1**

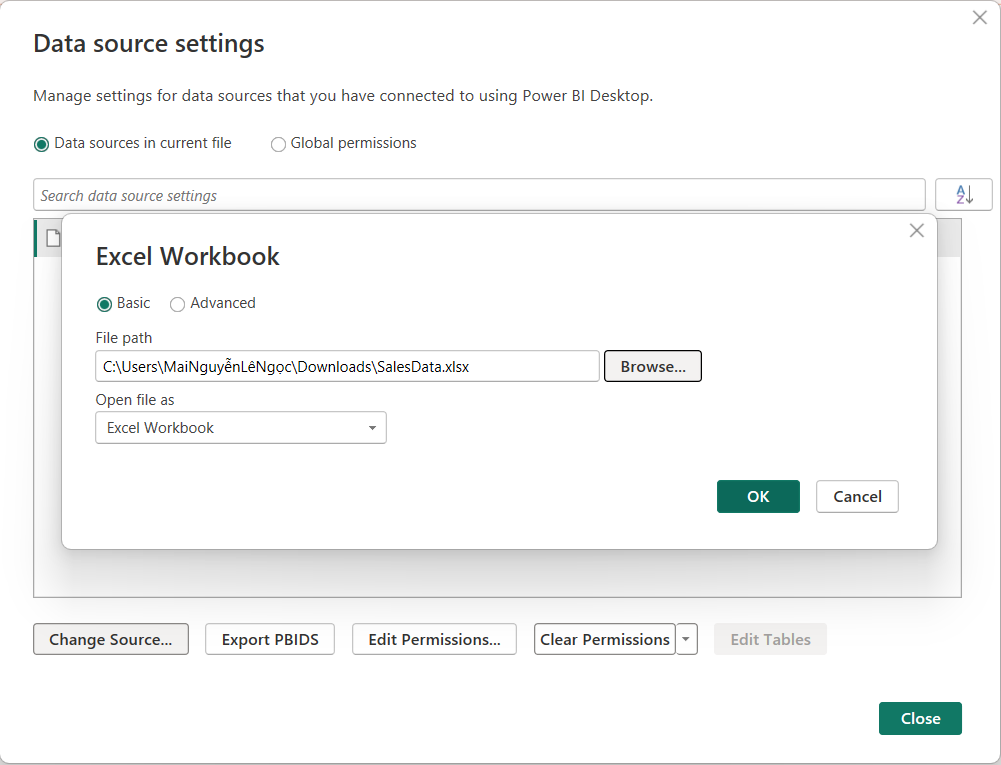
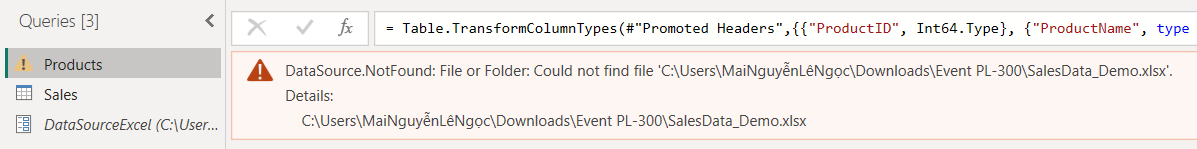
|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | **Questions** | **Answers** | | | |
| **LÝ THUYẾT** | | | | | |
| 1  ★ | Which data analyst task has a critical performance impact on reporting and data analysis? | **Model** | Analyze | Visualize | Share reports in Power BI Service |
| 2  ★★ | What does data granularity mean? | **Data granularity is the level of detail that is represented in the data** | Data granularity is the filter direction that is associated with the relationship between two columns | Data granularity is a many-to-many relationship | Data granularity is a type of schema design |
| 3  ★ | What is the difference between a fact table and a dimension table? | **Fact tables contain observational data while dimension tables contain information about specific entities within the data** | Fact tables contain information about specific entities while dimension tables contain information about observational data | There is no difference | Dimension tables tell you about specific roles while fact tables tell you information about facts that are associated with those roles in Power BI |
| 4  ★ | What is cardinality? | **Cardinality is the direction that the data flows in a relationship between two tables** | Cardinality is the granularity of the data | Cardinality is how long it takes for the data to load | Cardinality is a type of visual element |
| 5  ★ | Is it possible to create a relationship between two columns if they are different DATATYPE columns? | **No, both columns in a relationship must be sharing the same DATATYPE** | Yes, if the cardinality of the relationship is set to Many-to-Many | Yes, the above is fully supported in latest version of Power BI Desktop | Yes, if you enable Auto-detect relationships feature in Power BI Desktop |
| **DAX (Case Study)**  Link Files Power BI: [Bộ 30 câu hỏi](https://datapotanalytic.sharepoint.com/:f:/r/sites/Marketing-InboundMarketing/Shared%20Documents/Inbound%20Marketing/03%20Event/01%20WS%20Online/2025/WID250528_PL300_B%C3%A1n%20voucher/Webinar%20ch%E1%BB%AFa%20%C4%91%E1%BB%81%20PL%20300%20%202105%200406/Professional%20Material/B%E1%BB%99%2030%20c%C3%A2u%20h%E1%BB%8Fi?csf=1&web=1&e=2AkK59) | | | | | |
| You are working on a Sales Performance report for ABC Company.  Management has the following requests: | | | | | |
| 6  ★ | First, they want to compare current month’s sales with the previous month to track performance trends.  Which two functions will help you compare dates to the previous month? | **CALCULATE and PREVIOUSMONTH** | TOTALYTD and PREVIOUS MONTH | CALCULATE and TOTALTYD | TOTALYTD and DATEDIFF |
| 7  ★ | Second, the Sales table contains both Order Date (active relationship) and Ship Date (inactive relationship). The management team needs specific measures based on Ship Date, requiring you to handle the inactive relationship properly.   How do you use an inactive relationship in a single measure? | **Use the USERELATIONSHIP function** | Change the active relationship in the Modeling tab | Delete one relationship | Use the TREATAS function to simulate the relationship in the measure |
| 8 ★ | Third, the management team wants to track the running total of sales throughout the year to understand overall sales momentum.  They ask you to build a measure that calculates cumulative SalesAmount up to the current date.  Which DAX function should you use to calculate the running total? | DATESMTD | PREVIOUSMONTH | **TOTALYTD** | SUMX with ALLSELECTED |
| 9  ★ | The HR manager wants to assess salary fairness by identifying employees whose salaries are above the 50th percentile across the organization.  Which DAX function should you use to calculate? | **MEDIANX over the Employee table** | AVERAGEX over the Salary table | CALCULATE with a filter on salaries above the average | SUMX over the Employee table |
| 10  ★ | Finally, the management team wants to display company-wide total revenue in a card visual but also allow users to filter by Region elsewhere on the page.  You notice that your Total Revenue measure changes when slicers are applied. You are asked to prevent slicers from affecting this specific KPI, regardless of user selections.    What should you do in your DAX measure? | **Use the ALL or REMOVEFILTERS function to override the filter context** | Change the slicer interaction setting to “None” | Add DISTINCT to the measure to remove duplicates | Use a different visual to isolate the KPI from the filters |
| **MODELING** | | | | | |
| 11  ★★ | You are building a Power BI report for a retail company.  The Sales table contains over 10 million rows and includes two date columns: OrderDate and ReturnDate.  There is a single shared Date table in your model.  You have created the following relationships:   * An active relationship between Sales[OrderDate] and Date[Date] * An inactive relationship between Sales[ReturnDate] and Date[Date]   The manager requests a line chart that shows total returned sales by month, using the ReturnDate column.  Which measure should you create to meet this requirement? | ReturnedSales = SUM(Sales[SalesAmount]) | **ReturnedSales =**  **CALCULATE(**  **SUM(Sales[SalesAmount]),**  **USERELATIONSHIP(Sales[ReturnDate], ‘Date’[Date])**  **)** | ReturnedSales =  CALCULATE(  SUM(Sales[SalesAmount]),  Date[Date] = Sales[ReturnDate]  ) | ReturnedSales =  CALCULATE(  SUM(Sales[SalesAmount]),  FILTER(Sales, Sales[ReturnDate] <> BLANK())  ) |
| 12  ★★ | You are building a data model for a retail company.    You have the following tables:   * Sales: includes fields OrderID, ProductID, OrderDate, CustomerName * Product: includes fields ProductID, ProductName, Category * Returns: includes fields OrderID, ReturnDate     You want to connect Returns to Sales, so you can analyze return data by product, category, and order date.  What is the best way to configure the relationship between Returns and the other tables? | Create a relationship between Returns[ProductID] and Product[ProductID] directly | **Create a relationship between Returns[OrderID] and Sales[OrderID], and rely on existing relationships between Sales and Product** | Create a relationship between Returns[ReturnDate] and Sales[OrderDate] | Duplicate the Product table to connect directly to Returns using ProductID |
| 13  ★ | Your data model contains the following tables:  • Date table: 25K rows that are updated rarely  • Product table: 2K rows that are updated monthly  • Sales table: 1M rows that are updated regularly, and changes need to be reflected immediately  You need to identify the best storage mode for the data model tables. | **Product - Import, Sales - Direct query; Date - Import** | Product - Import, Sales - Import; Date - Direct query | Product - Direct query, Sales - Direct query; Date - Import | Product - Import, Sales - Import; Date - Import |
| 14  ★★★ | In your data model, you have three tables:   * Customer: contains customer details * OrderDetails: contains all customer orders * Region: contains the region assignment for each customer   There is a one-to-many relationship from Customer to OrderDetails, and from Region to Customer.    You need to create a report that shows total sales by region, but also allows users to filter by region and see only the relevant orders in visualizations.    However, some orders may not have a matching customer, and you do not want these unmatched orders to affect regional totals.  How should you configure the relationships to support this analysis and ensure accurate filtering? | **One-to-many from Customer to OrderDetails, and one-to-many from Region to Customer, both with single-directional filters** | One-to-many from Customer to OrderDetails with bi-directional filtering, and one-to-many from Region to Customer with single-directional filtering | One-to-many from Customer to OrderDetails, and one-to-many from Region to Customer, both with bi-directional filters | One-to-many from OrderDetails to Customer with bi-directional filtering, and one-to-many from Customer to Region with single-directional filtering |
| 15  ★ | In your Power BI report, you have columns that are necessary for calculations but should not be visible to end users.    How can you configure the data model to hide these columns from report viewers while still using them in your measures? | Delete the columns from the data model | **Mark columns as “Hide in report view” in Power BI Desktop** | Remove the columns from the query editor | Create a new table excluding those columns |

**BUỔI 2**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **DATA SOURCE** | | | | | |
| 16  ★  ✔ | When would you need to access the Data Source Settings? | If you need to connect to a new data source | If you need to edit an existing query | **If the file name or location changes** | All of the above |
| You are building a Power BI sales report that pulls data from multiple environments: Development, Test, and Production, all hosted in Azure SQL. | | | | | |
| 17  ★★★ | You have a dataset that uses the development database as a data source. You need to configure the dataset so that you can easily change the data source between the development, test, and production database servers from powerbi.com. Which should you do? | **Create a parameter and update the queries to use the parameter** | Create a JSON file that contains the database server names. Import the JSON file to the dataset | Create a query for each database server and hide the development tables | Set the data source privacy level to Organizational and use the ReplaceValue Power Query M function |
| 18  ★★✔ | You create a parameter named DataSourceExcel that holds the file name and location of a Microsoft Excel data source. You need to update the query to reference the parameter instead of multiple hard-coded copies of the location within each query definition.  Solution: You modify the source step of the queries to use DataSourceExcel as the file path.  Does this meet the goal? | **Yes** | No |  |  |
| 19  ★★★ | You have two Microsoft SQL Server database servers named SQLProd and SQLDev. SQLDev contains the same tables as SQLProd, but only a subset of the data in SQLProd.    You create a new Power BI Desktop model that uses 120 tables from SQLDev.    You plan to publish the Power BI file to the Power BI service.    You need to connect the model to the tables in SQLProd. The solution must minimize administrative effort.    What should you do from Query Editor before you publish the model? | Create a new connection to SQLProd, and then import the tables from SQLProd | Delete the existing queries, and then add new data sources | **Configure the Data source settings** | Edit the source of each table query |
| **CLEAN & TRANSFORM** | | | | | |
| 20  ★ | Which of the following selections are not the best practices for naming conventions in Power BI?  You can choose more than one answer. | **Abbreviate column names** | Rename columns to have spaces in them | Replace values that have integers with human readable results | **Use special characters or symbols in column names** |
| 21  ★★✔ | You have a CSV file that contains user complaints. The file contains a column named Logged. Logged contains the date and time each complaint occurred. The data in Logged is in the following format: 2018-12-31 at 08:59.  You need to be able to analyze the complaints by the logged date and use a built-in date hierarchy.  What should you do? | **Split the Logged column by delimiter “at” and change the data type of the first part to Date** | Change the data type of the Logged column to Date | Apply a transformation to extract the last 11 characters of the Logged column and set the data type of the new column to Date | Add a conditional column that outputs 2018 if the Logged column starts with 2018 and set the data type of the new column to Whole Number |
| 22  ★★✔ | You import an Excel spreadsheet into Power Bl Desktop. The spreadsheet contains manufacturers, models, and specifications of the computers deployed in your company.  You need to examine the quality of the data. The results of Data Preview are shown in the exhibits.  For each of the following statements, select Yes if the statement is true. Otherwise, select No. | Statement | | Yes | No |
| The DevicelD column is a candidate for the primary key for this table. | | X |  |
| The DeviceType column can be used as a dimension. | |  | X |
| A visualization containing the average amount of storage available for each DeviceType can be created using the Storage column. | |  | X |
| 23  ★★✔ | You have a table of Purchase Orders (POs), a sample of which is shown in the exhibit. There are relationships with other tables on the PO Number, PO Date, and Vendor columns.  You need to optimize model performance.  Which two actions should you perform? Each correct answer presents part of the solution. | **Remove the PO prefix from the PO Number column and change the data type to Whole Number** | Change the format of the PO Amount column to Currency | Mark the Purchase Orders table as the date table | **Add a new column to the Purchase Order table and perform a lookup to set the Vendor Id from the Vendor column** |
| **APPEND/ MERGE QUERIES** | | | | | |
| 24  ★ | If you want to concatenate 2 query results with all the combined rows, which operation should you perform? | **Append** | Merge | Join | Combine Column |
| 25  ★  ✔ | You have Employee, Contractor, and Consultant tables in a data model, as shown in the exhibit.  You need to combine these tables to create a new table. The Name column of the newly created table can be used as a primary key for the purpose of simplicity.  Which transformation should you use? | Merge queries as new | **Append queries as new** | Append queries | Merge queries |
| 26  ★★ | You create a dataset using multiple data sources that contain similar tables.  You need to reduce the number of tables in the dataset to make the data model simpler and the creation of visuals easier.  Which method should you use to combine the tables for each scenario? To answer, drag the appropriate method to each scenario. A method may be used once, more than once, or not at all. | Scenario | Append Queries | Merge Queries | Combine Files |
| A folder containing JSON documents containing the same structure |  |  | X |
| Multiple tables with a common key between them |  | X |  |
| Multiple tables with the same columns containing data for each month | X |  |  |
| **XỬ LÝ LỖI & TỐI ƯU POWER QUERY** | | | | | |
| 27  ★★★ | You import sales data using Power Query to apply a series of transformation steps after the query has been executed.  There are errors, as shown in the exhibit for the Rename Column step and the exhibit for Cells.  You need to resolve the errors without losing any data.  How should you resolve the errors? To answer, select the appropriate option for each part. | | | | |
|  | a. Step-level error: | Change the Data source Settings | **Remove the step from Query Settings** | Use the Remove Errors feature | Duplicate the step and rename it |
|  | b. Cell-level error: | Remove the column | Use the Remove Errors feature | **Use the Replace Errors feature** | Change data type to Text |
| 28  ★ | You import an HR dataset into Power BI Desktop. You need to see how many empty and error rows are in a dataset. Which two data quality options can you use to meet your goal? Each correct answer presents a complete solution. | **Column quality** | **Column profile** | Column distribution | Custom column |
| 29  ★ | You open a query in Power Query Editor. You need to identify the percentage of empty values in each column as quickly as possible. Which Data Preview option should you select? | **Column quality** | Column distribution | Column profile | Show whitespace |
| 30  ★ | What can be achieved by removing unnecessary rows and columns? | **Deleting unnecessary rows and columns will reduce dataset size and its good practice to load only necessary data into your data model** | It is not necessary to delete unnecessary rows and columns, and it is a good practice to keep all metadata intact | Deleting unnecessary rows and columns can damage the structure of the data model | Removing rows and columns will improve visual quality of charts |

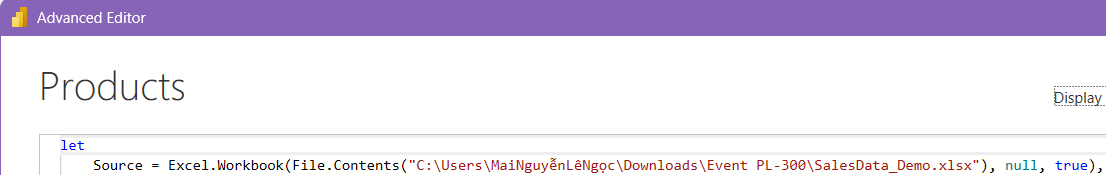
**Giải thích đáp án các câu hỏi**

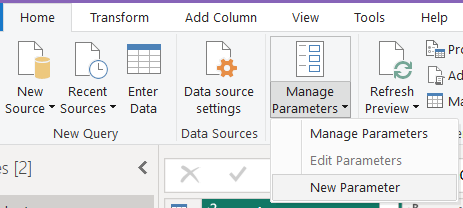
16.

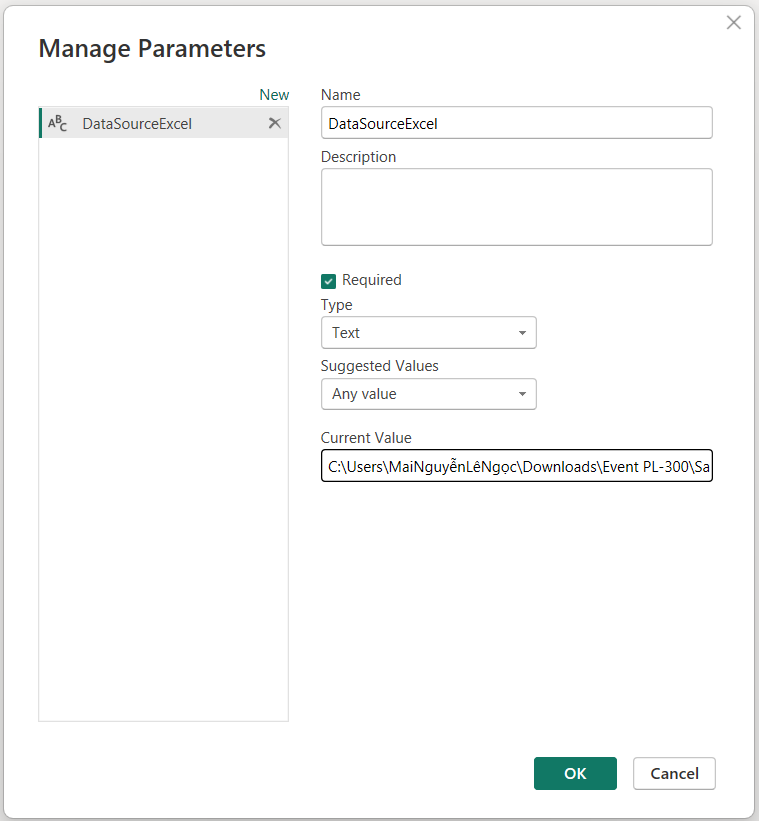


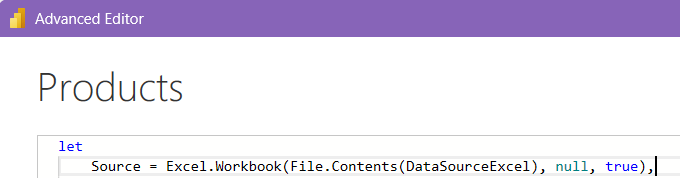
17. Trong PBI service, bạn không thể chỉnh sửa nguồn dữ liệu, thay vào đó bạn nên sử dụng một tham số (parameter) để lưu trữ các chi tiết kết nối, chẳng hạn như tên CSDL, không nên sử dụng một dòng chữ ghi nhận một kết nối cố định. Điều này sẽ giúp bạn quản lý việc kết nối dữ liệu thông qua cổng web của PBI service, hoặc sử dụng API ở các bước tiếp theo, một cách dễ dàng hơn.

18. Sử dụng parameter trong Data Source được ứng dụng trong nhiều trường hợp khác nhau, từ kết nối đến các nguồn dữ liệu khác nhau được định nghĩa trong Query Parameters, đến việc load các nhóm các cột khác nhau vào PBI.



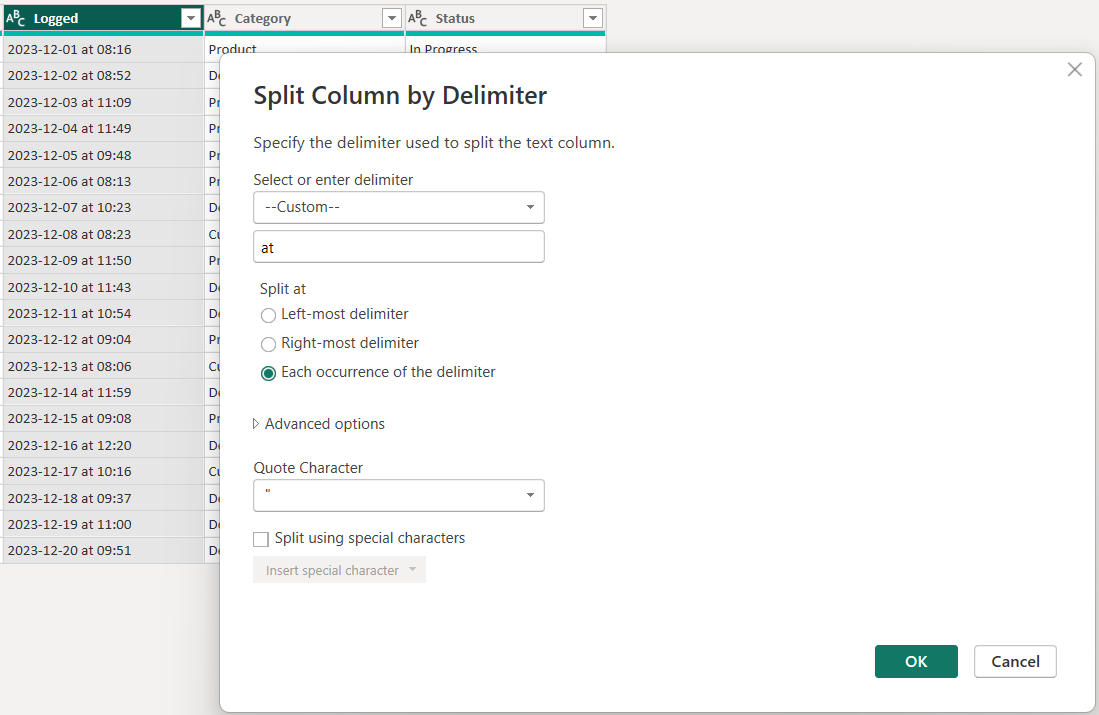


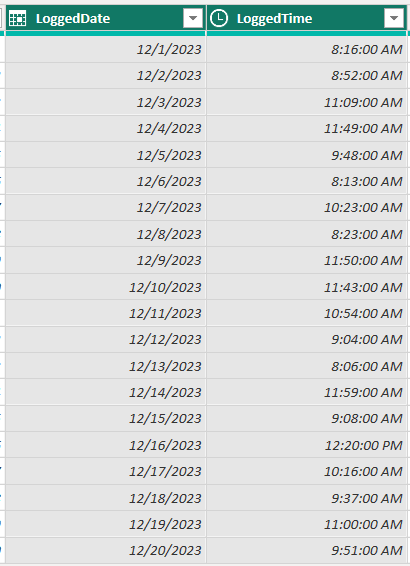


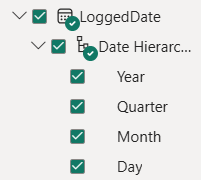


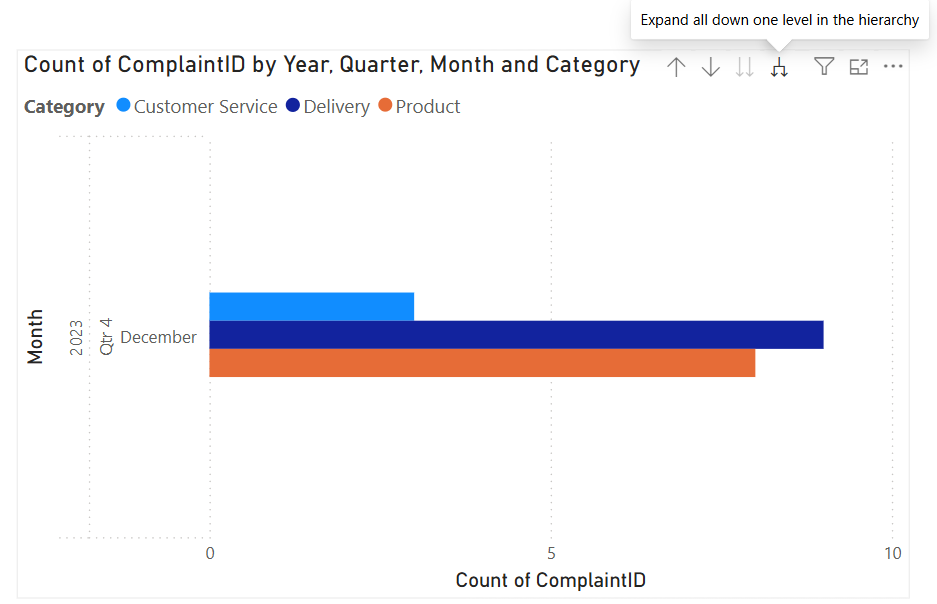
20. Các từ viết tắt/ chữ cái, biểu tượng đặc biệt dẫn đến nhầm lẫn vì chúng thường bị lạm dụng hoặc không được thống nhất chung.

21. Để sử dụng built-in-date hierarchy, bạn cần đặt kiểu dữ liệu của 2 cột mới đã tách thành Date & Time.

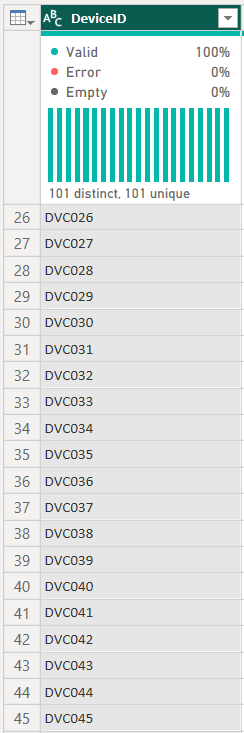




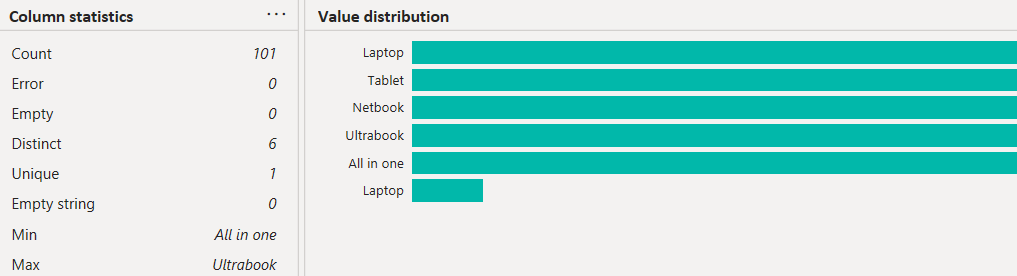


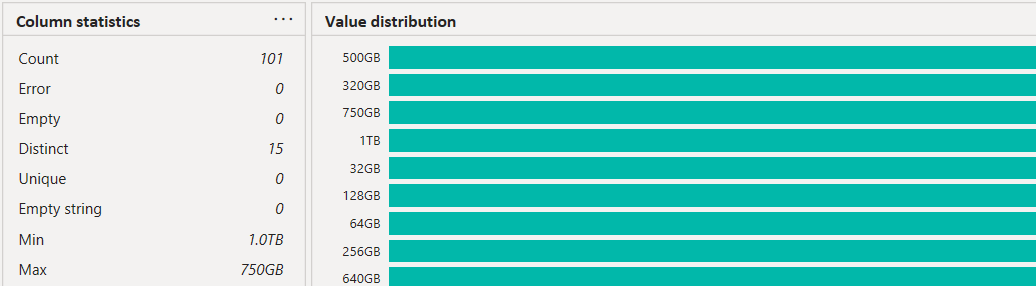


22.

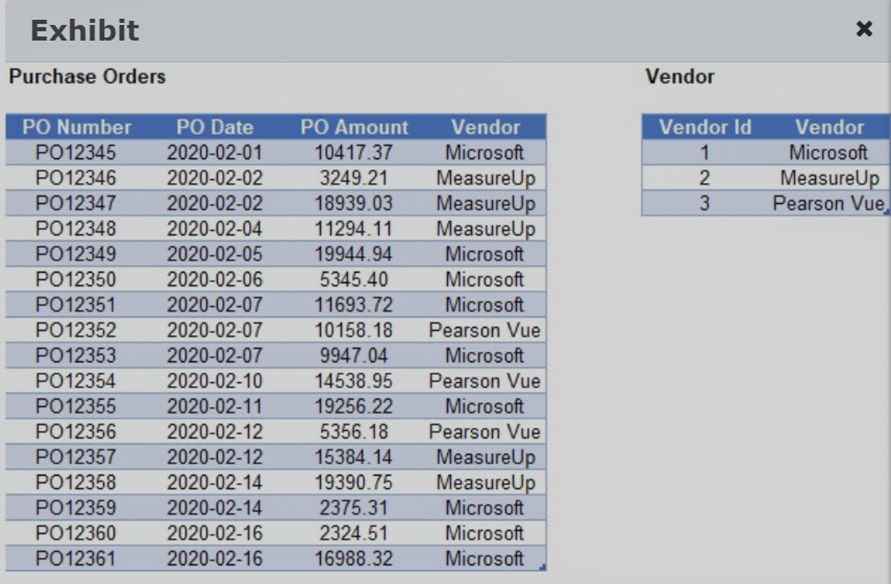


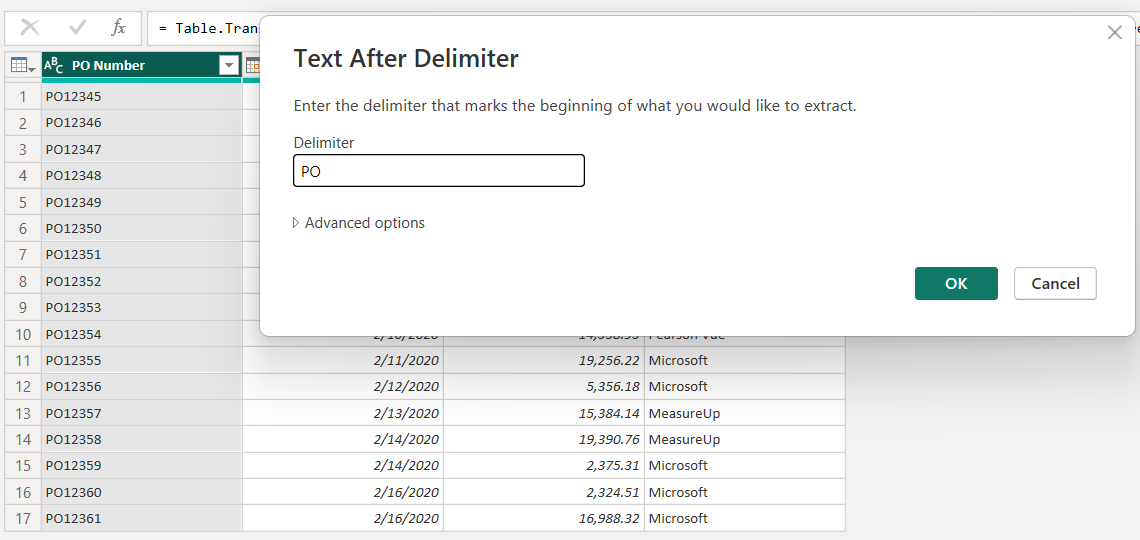
The DevicelD column is a candidate for the primary key for this table. When the distinct and unique counts are the same, the column contains unique values. The DevicelD column can be used as the primary key for this table.

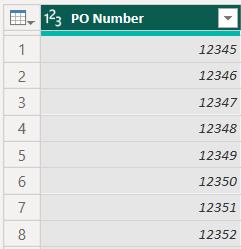
The DeviceType column cannot be used as a dimension. The DeviceType column has unique values, but two of the labels are Laptop. This is probably because there are spaces on one of the Laptop labels. You must cleanse the column to remove spaces, or use Replace Values, before you can use it as a dimension.

A visualization containing the average amount of storage available for each DeviceType cannot be created using the Storage column. The Storage column is a Text column. You cannot average a text column. The data is also in different units of measure: GB and TB. You would need to perform a series of transformations to extract the numeric value and convert TB into GB.

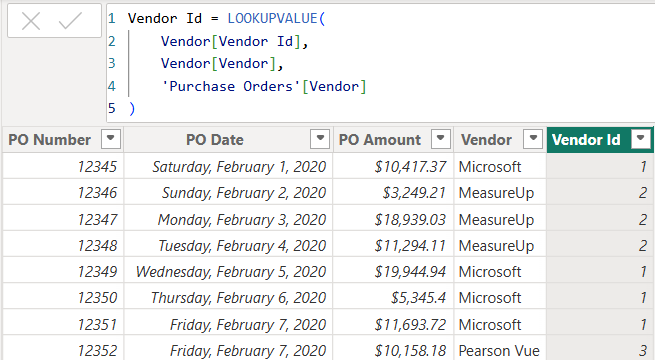
23.



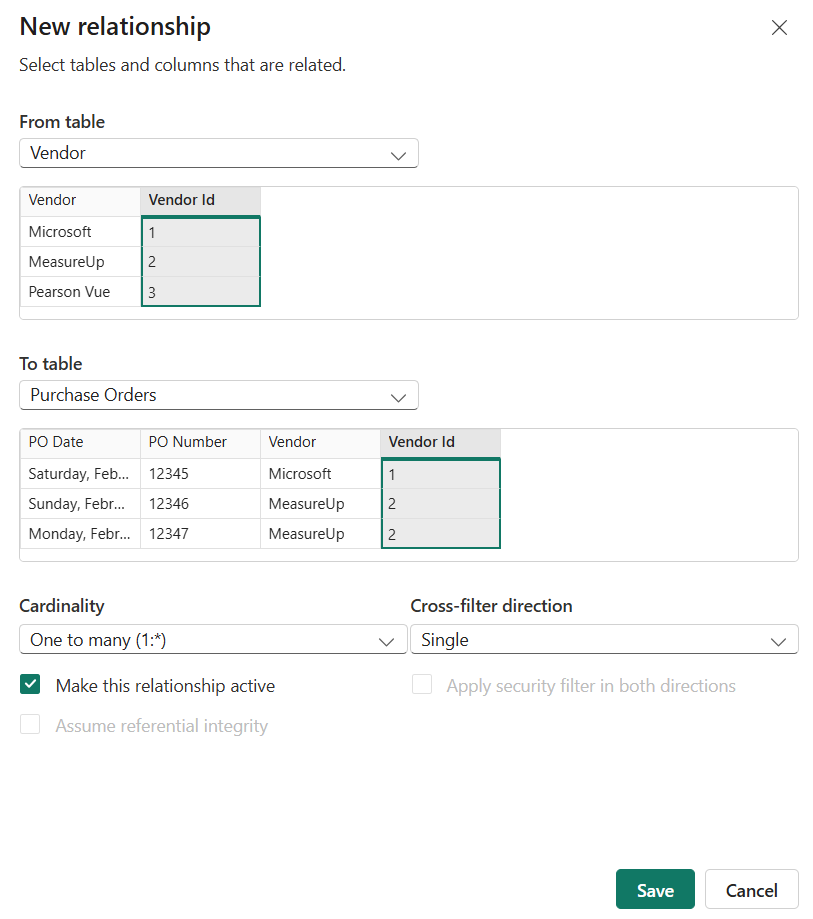




Bạn nên loại bỏ tiền tố PO ở cột PO Number và thay đổi kiểu dữ liệu thành Whole Number. Mô hình sẽ đạt được hiệu suất cao nhất với dữ liệu dạng số, sử dụng mã hóa giá trị. Cột văn bản sử dụng mã hóa băm (hash) sẽ ảnh hưởng xấu đến hiệu suất.

  
Bạn cũng nên thêm một cột mới vào bảng Purchase Order và thực hiện lookup để đặt Vendor Id theo cột Vendor. Cột này nên sử dụng hàm sau:

Vendor Id = LOOKUPVALUE(Vendor[Vendor Id],Vendor[Vendor],'Purchase Order'[Vendor]).



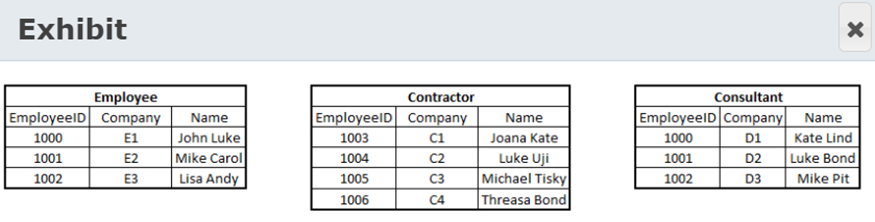
Sau đó, bạn nên tái tạo mối quan hệ bằng cách sử dụng Vendor ID chứ không phải Vendor Name. Sử dụng dữ liệu dạng số trong các mối quan hệ thay vì dữ liệu văn bản sẽ cải thiện hiệu suất mô hình.

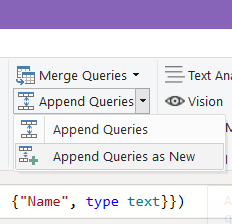
Bạn không nên thay đổi định dạng của cột PO Amount bởi vì định dạng cột không ảnh hưởng đến hiệu suất của mô hình.

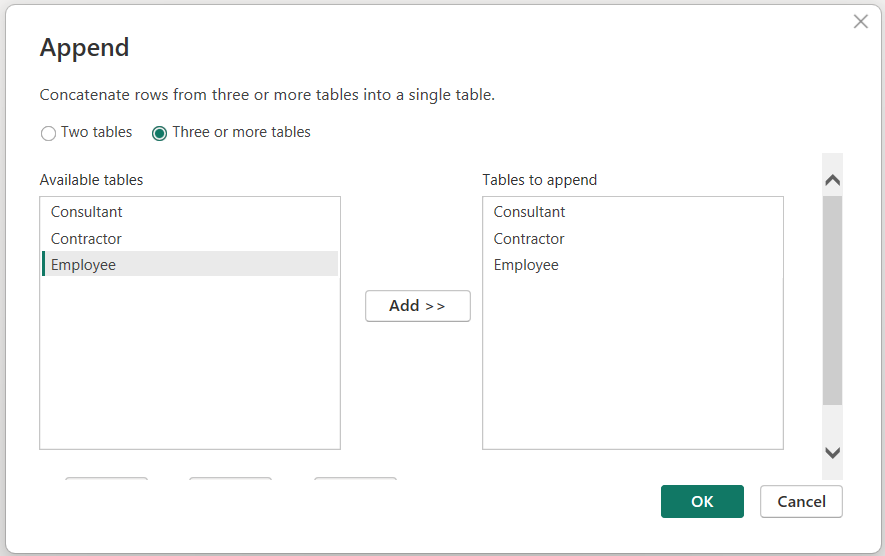
Bạn không nên để bảng Purchase Orders làm bảng ngày tháng. Bảng ngày tháng yêu cầu các ngày sát nhau và duy nhất. Bảng Purchase Orders không đáp ứng được một trong hai yêu cầu này. Bạn nên tạo bảng Ngày tháng riêng.

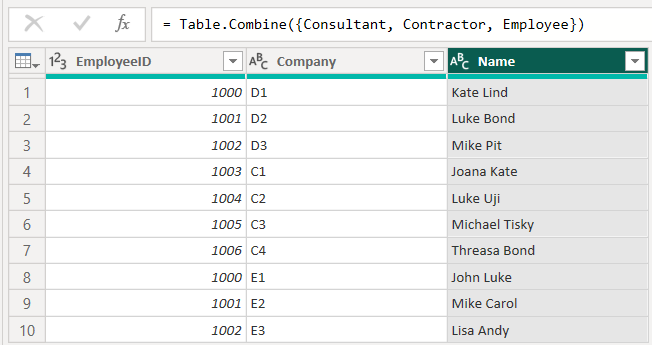
24. Append sẽ lấy hai bảng và kết hợp nó thành một truy vấn. Truy vấn kết hợp sẽ có nhiều hàng hơn trong khi vẫn giữ nguyên số cột.

25.







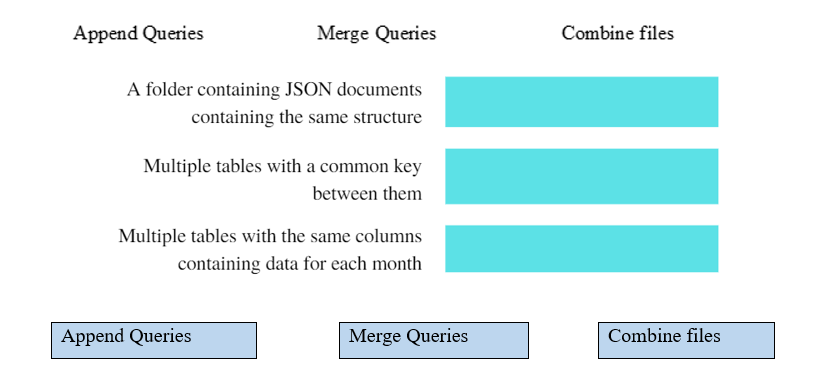
  
You should use the Append queries as new transformation to combine employee, contractor, and consultant tables into a new table. This transformation allows you to combine two or more tables into a new table if all participating tables have the same schema. This is like a union in SQL. Since all the values in the Name column are unique, you can use this column as a primary key for newly created table.

You should not use the Append queries transformation. This transformation allows you to combine two or more tables into an existing table if all participating tables have the same schema. This is like a union in SQL.

You should not use the Merge queries transformation. This transformation allows you to merge two or more tables into an existing table based on a column that is common between all participating tables. This is like a join in SQL.

You should not use the Merge queries as new transformation. This transformation allows you to merge two or more tables into a new table based on a column that is common between all participating tables. This is like a join in SQL.

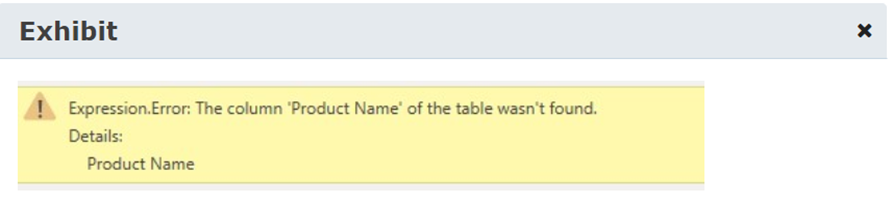
26.

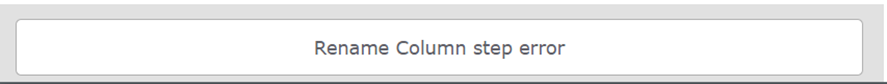
Bạn nên sử dụng Combine Files (binaries) để kết hợp các tệp có cùng loại tệp và cấu trúc trong một thư mục và tạo một bảng duy nhất. Bạn có thể nhập Excel, JSON và văn bản bay bằng phương pháp Combine Files.

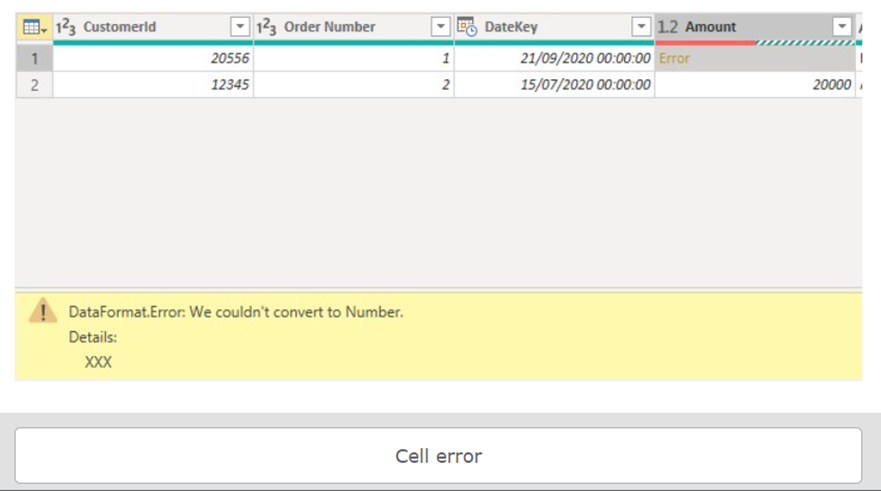
Bạn nên sử dụng phương pháp Merge Queries để kết hợp hai bảng thành một bằng cách sử dụng một cột phổ biến trong cả hai bảng. Phương thức Merge Queries thực hiện một hoạt động tương tự như một JOIN trong SQL. Bạn có thể chỉ định việc hợp nhất thực hiện liên kết bên trong hay bên ngoài.

Bạn nên sử dụng phương pháp Append Queries để kết hợp hai bảng thành một bằng cách thêm các hàng từ cả hai bảng vào một bảng. Các bảng phải có cấu trúc giống nhau. Phương thức Append Queries nối các hàng thành một bảng.

27.







You should remove the step from Query Settings. The error is probably caused by the column in error no longer being in the data. The column has either been removed or renamed. By deleting the step, you can let the other transformations be applied.

You should not use Remove Errors to fix the step-level error. This feature removes rows with errors. It does not solve an error in a step

You should not change the Data source settings to fix the step-level error. Data source settings contain the data source type, its location, and credentials for accessing the data. Changing the data source settings will not resolve the step error.

You should use the Replace Errors option to fix the cell-level error. This allows you to specify a replacement value for the cells in error. The data in the cell could not be converted to a numeric value. This is because the value contains alpha characters. Only one cell is affected, so this is not an operation error.

28. You can use the Column quality and Column profile options to check empty or error values. Column quality allows you to analyze valid, error, or empty values for all columns in a single view. Column profile allows you to analyze value distribution along with empty or error values for the selected column.

You should not use the Column distribution option to check empty or error values. Column distribution allows you to show distinct and unique values for all columns in a single view.

You should not use the Custom column option to check empty or error values. The custom column option allows you to create a new column from Power Query editor either by using an example or providing a column formula.

29. Column quality: Trong hộp Column quality, chúng ta có thể nhìn thấy tỷ lệ dữ liệu có ý nghĩa (valid), lỗi, và ô trống trong cột được chọn.

30. Xóa các dòng và các cột không cần thiết sẽ làm giảm kích thước dataset, chúng ta sẽ chỉ cần load các dữ liệu cần thiết vào data model.