**Project: Building an ETL Pipeline from Azure SQL to Azure Data Lake using Azure Data**

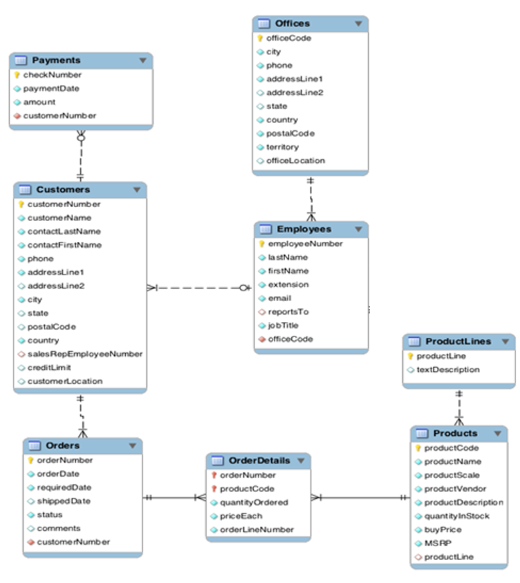
### Overview

* Objective: Automate data extraction from **Azure SQL DB** into **Azure Data Lake Storage**.
* Tools Used: **Azure SQL, Azure Data Studio, ADF, ADLS, Parquet/CSV**.
* Pipeline Architecture:

A blue and green logo with black text

AI-generated content may be incorrect.

**We have the following ERD for a IlluminaireSales database schema.**

****

### Phase 1: Create the OLTP Database

1. **Set Up an OLTP Database in Azure SQL Database**

**A screenshot of a computer

AI-generated content may be incorrect.**

**A screenshot of a computer

AI-generated content may be incorrect.**

1. **Open Azure Data Studio > New Connection > Use Microsoft Entra ID**

**A screenshot of a computer

AI-generated content may be incorrect.**

To avoid enabling public access for everyone but to connect from my computer, added the IP to the firewall.

**A computer screen with a computer screen

AI-generated content may be incorrect.**

1. **Check if the database is Normalized (3NF)**

A table is in 3NF if:

* It is in 2nd Normal Form (2NF) (no partial dependencies).
* It has no transitive dependencies (non-key columns should depend only on the primary key).

|  |  |
| --- | --- |
| **Table** | **Observations** |
| Offices | No issues |
| Employees | No issues |
| Customers | City, State, Country are redundant (could be normalized into a Geography table) |
| ProductLines | No issues |
| Products | No issues |
| Orders | No issues |
| OrderDetails | No issues |
| Payments | No issues |

Proceeding with the existing schema which is almost 3NF.

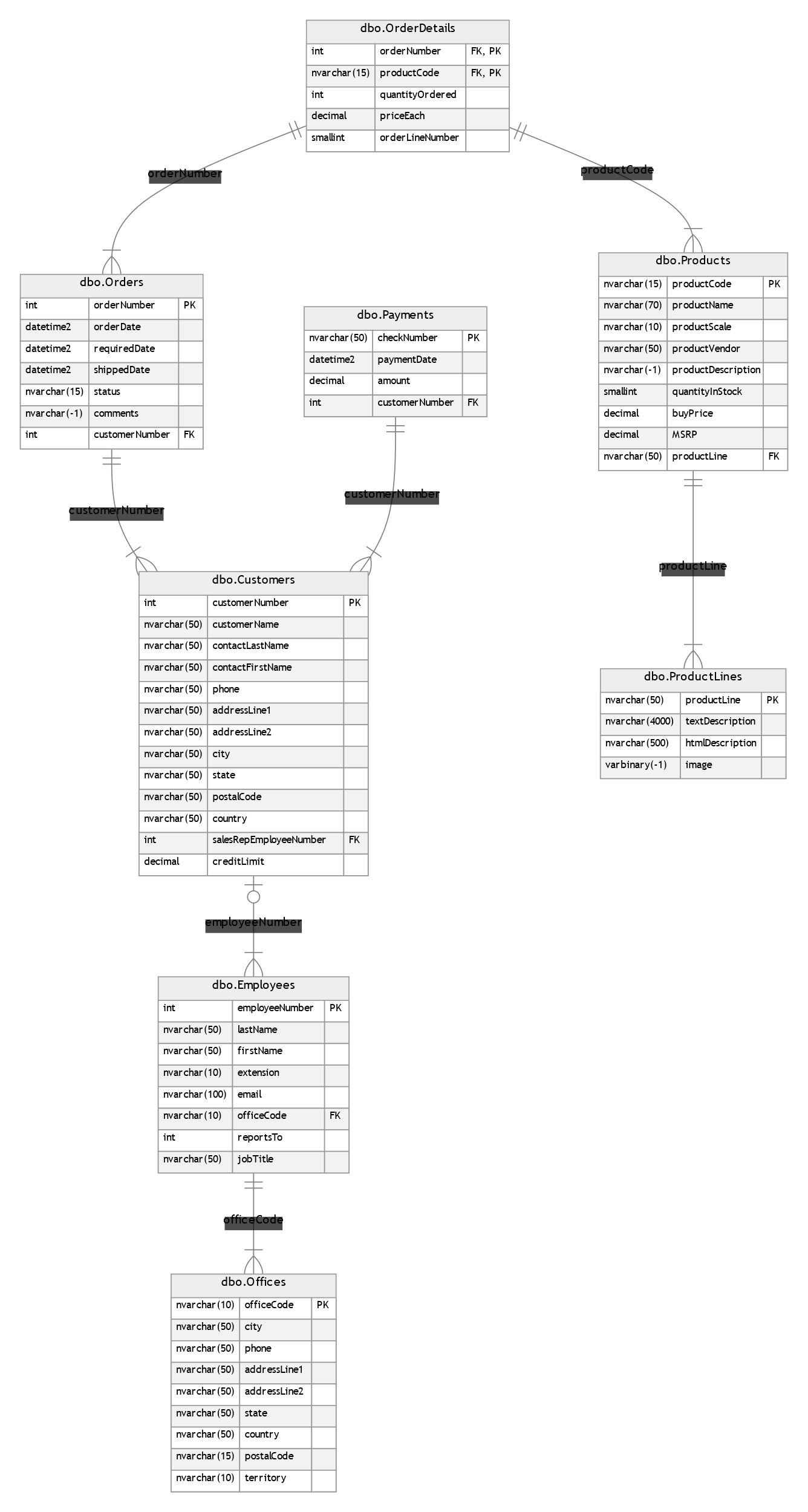
1. **Created an Azure SQL Database ‘IlluminaireSales’ using Azure Data Studio**

A screenshot of a computer

AI-generated content may be incorrect.

A screenshot of a computer

AI-generated content may be incorrect.



OLTP database setup is complete!

### Phase 2: Implementing the ETL pipeline

1. **Extract Data from OLTP into Azure Data Lake (ADLS) Using Azure Data Factory (ADF)**
   1. Create an Azure Data Lake Storage (ADLS) Account
      1. Storage Account Name: illuminaireadls
      2. Choose the same region as your Azure SQL Database.
      3. Enable Hierarchical Namespace (Required for Data Lake Storage Gen2)
      4. Create a container in ADLS -> Set Public Access Level to Private. A screenshot of a computer

         AI-generated content may be incorrect. A screenshot of a computer

         AI-generated content may be incorrect. A screenshot of a computer

         AI-generated content may be incorrect. A screenshot of a computer

         AI-generated content may be incorrect.
2. **Create an Azure Data Factory (ADF) Pipeline**
   1. Set up ADF
      1. Region: Same as Azure SQL DatabaseA screenshot of a computer

         AI-generated content may be incorrect. A screenshot of a computer

         AI-generated content may be incorrect.
   2. Create a Parameterized ADF Pipeline to Copy All 8 Tables
      1. Create a parameter for TableName - This parameter will allow us to pass any table dynamically into the pipeline. A screenshot of a computer

         AI-generated content may be incorrect.
      2. Add a ForEach Loop to Process All Tables

@json('["Customers","Orders","Products","Payments","Employees","OrderDetails","Offices","ProductLines"]')A screenshot of a computer

AI-generated content may be incorrect.

* + 1. Add the Copy Data Activity Inside the ForEach Loop and add the **source** (Azure SQL Database) where the OLTP data is sitting by creating a linked serviceA screenshot of a computer

       AI-generated content may be incorrect.This parameter will hold the table name dynamically from the ForEach loop.A screenshot of a computer

       AI-generated content may be incorrect. This tells ADF to use the value of the TableName parameter instead of a hardcoded table. A screenshot of a computer

       AI-generated content may be incorrect. This ensures that for each iteration, the table name is dynamically assigned from the list in the ForEach loop. A screenshot of a computer

       AI-generated content may be incorrect.
    2. Now configure the Sink by creating an ADLS linked serviceA screenshot of a computer

       AI-generated content may be incorrect.

A screenshot of a computer

AI-generated content may be incorrect.

* + 1. Now validate the pipeline: **A screenshot of a computer

       AI-generated content may be incorrect.**
    2. Upon Debug, the pipeline runs successfully. A screenshot of a computer

       AI-generated content may be incorrect.
    3. Publish all

### Troubleshooting performed during pipeline creation:

A screenshot of a computer

AI-generated content may be incorrect.

Cannot connect to SQL Database. Please contact SQL server team for further support. Server: 'illuminaire-server.database.windows.net', Database: 'IlluminaireSales', User: ''. Check the linked service configuration is correct, and make sure the SQL Database firewall allows the integration runtime to access.

Cannot open server 'illuminaire-server' requested by the login. Client with IP address '40.119.9.182' is not allowed to access the server. To enable access, use the Azure Management Portal or run sp\_set\_firewall\_rule on the master database to create a firewall rule for this IP address or address range. It may take up to five minutes for this change to take effect.A screenshot of a computer

AI-generated content may be incorrect.

Issue: Azure Data Factory (ADF) Managed Identity Not Listed in Microsoft Entra ID

A screenshot of a computer

AI-generated content may be incorrect.

A screenshot of a computer

AI-generated content may be incorrect.

A screenshot of a computer

AI-generated content may be incorrect.

Assign Database Permissions

Once your ADF Managed Identity is set as an Azure SQL Admin, grant it permissions at the database level. A screenshot of a computer

AI-generated content may be incorrect.

A screenshot of a computer

AI-generated content may be incorrect.