SSIS Project – Part 2

1. Correct the "Extract Internet Sales" package by adding a Lookup that uses the following query. The source for this data is Products DB on the localhost.
   Send a file with unmatched records to your email address.

```sql
SELECT
        p.ProductKey,
        p.ProductName,
        p.StandardCost,
        p.Color,
        p.ListPrice,
        p.[Size],
        p.Weight,
        p.Description,
        ps.ProductSubCategoryKey,
        ps.ProductSubcategoryName,
        pc.ProductCategoryKey,
        pc.ProductCategoryName
FROM
        Product p
        JOIN ProductSubcategory ps ON p.ProductSubcategoryKey = ps.ProductSubCategoryKey
        JOIN ProductCategory pc ON ps.ProductCategoryKey = pc.ProductCategoryKey
```

2. Create a package called "Load Products". This package will populate product dimensions (`DimProductCategory, DimProductSubcategory, DimProduct`) in your Data Warehouse DB. Use the scripts below to get source data for these dimensions.

```sql
-- use your staging DB table resulted from step 1 as a source
SELECT DISTINCT ProductCategoryBusinessKey, ProductCategoryName
FROM dbo.InternetSales

-- sub category has a parent relationship to category, you will have to include a parent lookup
SELECT DISTINCT
        ProductCategoryBusinessKey,
        ProductSubcategoryBusinessKey,
        ProductSubcategoryName
FROM dbo.InternetSales

-- note parent relationship to sub category
SELECT DISTINCT
        ProductSubcategoryBusinessKey,
        ProductBusinessKey,
        ProductName,
        StandardCost,
        Color,
        ListPrice,
        Size,
        Weight,
        Description
FROM dbo.InternetSales
```

3. Remove the 1000 records limit in "Extract Internet Sales" package and run it. Modify the "Load Products" package to include a Lookup transformation to insert new records into the respective table and update existing ones. Run the "Load Products" package.

4. Create a "Load Geography" package to load only new records into the DW. Make sure to only load new records. Remove the 1000 limit in the "Extract Customers" package, and run it. Run the load Geography package again.

```sql
-- use your staging DB table resulted from step 1 as a source
SELECT DISTINCT
        PostalCode,
        City,
        StateProvinceName,
        CountryRegionCode,
        CountryRegionName
FROM dbo.Customers
```

5. Remove all records from the Customers table in your staging DB, as from running the packages multiple times without any restrictions, we ended up having duplicates. Add a lookup transformation to only load new customers and run the package first for 1000 first records, and then without limit. Make sure there is no duplicate records.

6. Create a package "Load Customers". Use your staged Customers data to populate the DimCustomers in your Data Warehouse. Be careful, there is a relationship to the dimGeography.

7. Remove all records from the Internet sales table in your staging DB. Modify the extract internet sales query to work with a date limit. Only push orders created in month of July 2001. Run the package

8. Create "Load Internet Sales" package to transfer the data from your staging table into your DW fact table. Empty the Staging table after the transfer.

9. Modify your "Extract Internet Sales" load one month (following) of data at every run. Hint, use SQL query to get the maximum date from the fact table.