





Homework 1c

Transforming & Staging





Transform

Change 'DateA' to Date



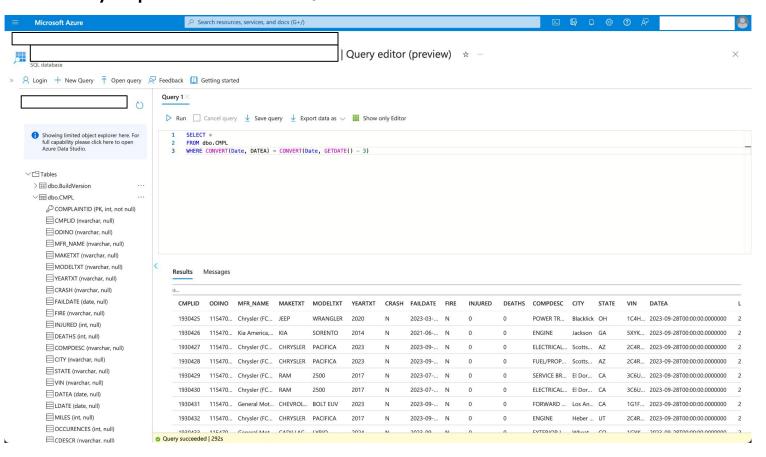
Staging

Copy Data to SQL DB Table



HW1c Final Output

Most Recently Updated Rows/Observations from the SQL Table Data

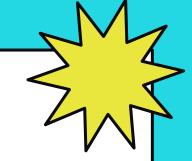




HOMEWORK 1 GRADED ASSIGNMENT

- Building on Homework 1a and 1b
- Finalize structure of data into SQL Server Database
- Assignment to be graded





Homework 1c

Transforming & Staging Deployed Services



SQL Database



Data Factory





O. Deploy HW 1c template



Open Azure CloudShell



Clone Github Repository (ALREADY CLONED)

"Git Pull" if needed!



Create Azure Template



Deploy Azure Template

*Hint

PS /home/dk98> cd './OMDSMod4Homework/Homework 1c/Transforming and Staging/'
PS /home/dk98/OMDSMod4Homework/Homework 1c/Transforming and Staging> bash ./formTemplate.sh
Template and parameters created successfully: ./template/template.json and ./template/parameters.json





Deployment Result

SQL server East US

dba09cc5
SQL database East US



SQL

Azure SQL Database

Azure SQL Server





1. Go to Azure SQL Database

SQL server East US

dba09cc5
SQL database East US

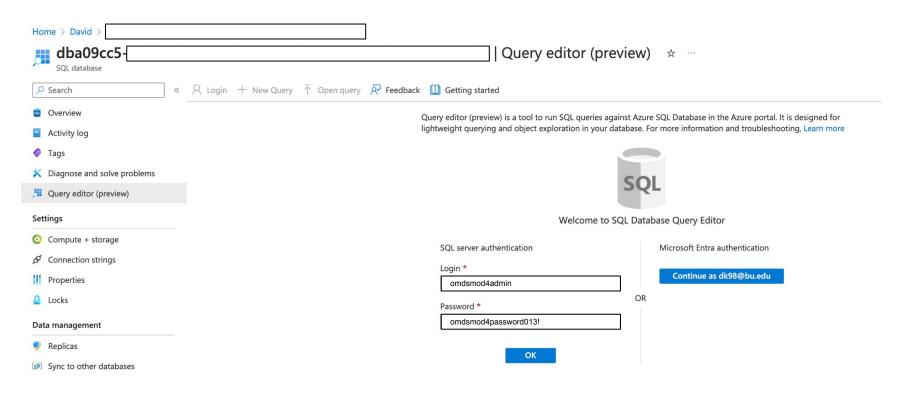


Azure SQL Database



Azure SQL Server

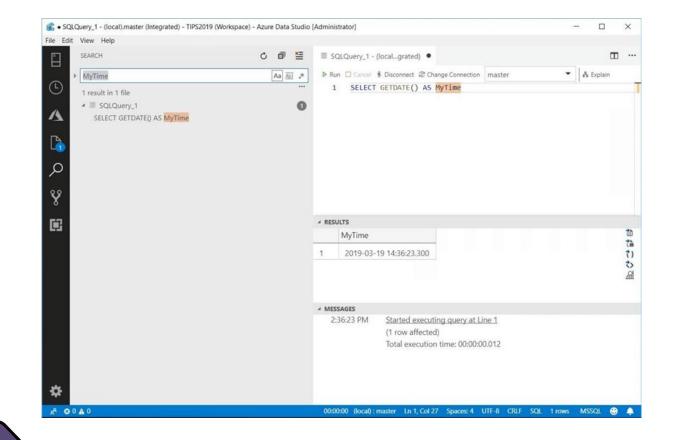
1.1 Access Query Editor in SQL DB





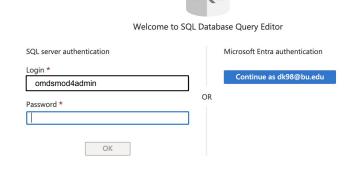






1.2 Authentication

- 1. Deployed our template:
 - a. Login: omdsmod4admin
 - b. Password: omdsmod4password013!
- 2. Created SQL Database Manually:
 - a. Own Credentials



*Firewall Error

If Prompted,

Give Access to your own IP Address

1.3 Create Table in SQL Database

CREATE TABLE complaints

CMPLID int primary key,

Look at <u>Reference</u> File to create the following query

CREATE TABLE Complaints

Appendix A. Complaints File Characteristics

| Field# | Name | Type/Size | Description |
|--------|------------|-------------|---|
| 1 | CMPLID | CHAR (9) | NHTSA'S INTERNAL UNIQUE SEQUENCE NUMBER. IS AN UPDATEABLE FIELD, THUS DATA FOR A GIVEN RECORD POTENTIALLY COULD CHANGE FRO |
| 2 | ODINO | CHAR (9) | ONE DATA OUTPUT FILE TO THE NEXT. |
| | | | ALSO, IF LDATE IS PRIOR TO DEC 15, 2002, THIS NUMBER MAY BE REPEATED FOR MULTIPLE PRODUCTS OWNED BY THE SAME COMPLAINANT. |
| | MFR_NAME | CHAR (40) | |
| 4 | MAKETXT | | VEHICLE/EQUIPMENT MAKE |
| 5 | MODELTXT | | VEHICLE/EQUIPMENT MODEL |
| 6 | YEARTXT | CHAR (4) | MODEL YEAR, 9999 IF UNKNOWN or N/A |
| 7 | CRASH | CHAR (1) | WAS VEHICLE INVOLVED IN A CRASH, 'Y' OR |
| 8 | FAILDATE | CHAR (8) | MODEL YEAR, 9999 IF UNKNOWN OR N/A WAS VEHICLE INVOLVED IN A CRASH, 'Y' OR DATE OF INCIDENT (YYYYMMDD) |
| 9 | FIRE | CHAR (1) | WAS VEHICLE INVOLVED IN A FIRE 'Y' OR 'N |
| 10 | INJURED | NUMBER (2) | NUMBER OF PERSONS INJURED |
| 11 | DEATHS | NUMBER (2) | NUMBER OF FATALITIES |
| 12 | COMPDESC | CHAR (128) | SPECIFIC COMPONENT'S DESCRIPTION |
| 13 | CITY | | CONSUMER'S CITY |
| 14 | STATE | CHAR (2) | CONSUMER'S STATE CODE |
| 15 | VIN | CHAR (11) | VEHICLE'S VIN# |
| 16 | DATEA | CHAR (8) | DATE ADDED TO FILE (YYYYMMDD) |
| 17 | LDATE | CHAR (8) | DATE COMPLAINT RECEIVED BY NHTSA (YYYYMM |
| 18 | MILES | NUMBER (7) | VEHICLE MILEAGE AT FAILURE |
| 19 | OCCURENCES | NUMBER (4) | NUMBER OF OCCURRENCES |
| 20 | CDESCR | CHAR (2048) | DESCRIPTION OF THE COMPLAINT |
| 21 | CMPL TYPE | CHAR (4) | SOURCE OF COMPLAINT CODE: |



CREATE TABLE complaints (CMPLID int primary key,

ODINO nvarchar(max), MFR_NAME nvarchar(max), MAKETXT nvarchar(max), MODELTXT nvarchar(max), YEARTXT nvarchar(max), CRASH nvarchar(max), FAILDATE date, FIRE nvarchar(max), INJURED int, DEATHS int, COMPDESC nvarchar(max), CITY nvarchar(max), STATE nvarchar(max), VIN nvarchar(max), **DATEA** date, LDATE date, MILES int, OCCURENCES int, CDESCR nvarchar(max),

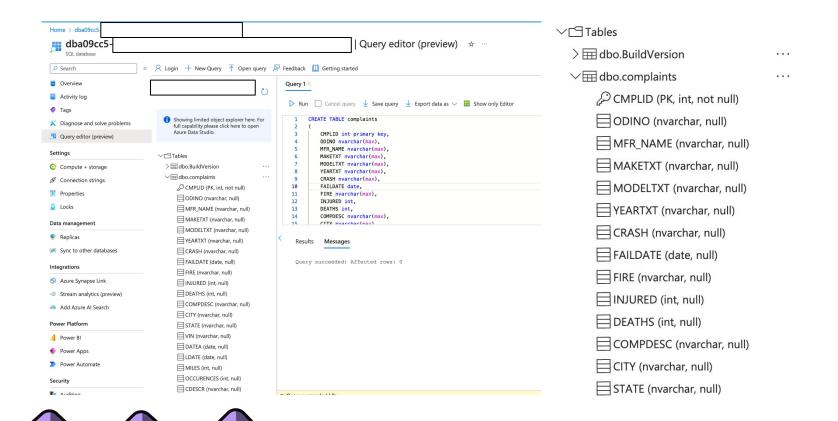
CMPL_TYPE nvarchar(max), POLICE_RPT_YN nvarchar(max), PURCH DT date, ORIG_OWNER_YN nvarchar(max), ANTI_BRAKES_YN nvarchar(max), CRUISE_CONT_YN nvarchar(max), NUM_CYLS int, DRIVE_TRAIN nvarchar(max), FUEL_SYS nvarchar(max), FUEL_TYPE nvarchar(max), TRANS_TYPE nvarchar(max), VEH_SPEED int, DOT nvarchar(max), TIRE_SIZE nvarchar(max), LOC_OF_TIRE nvarchar(max), TIRE_FAIL_TYPE nvarchar(max), ORIG_EQUIP_YN nvarchar(max), MANUF_DT date, SEAT_TYPE nvarchar(max), RESTRAINT_TYPE nvarchar(max),

DEALER_TEL nvarchar(max),
DEALER_CITY nvarchar(max),
DEALER_STATE nvarchar(max),
DEALER_ZIP nvarchar(max),
PROD_TYPE nvarchar(max),
REPAIRED_YN nvarchar(max),
MEDICAL_ATTN nvarchar(max),
VEHICLES_TOWED_YN nvarchar(max)

The decision to utilize nvarchar(max) instead of adhering to the reference file is due to the presence of human error in the data.

DEALER_NAME nvarchar(max),

1.4 Check your 'complaints' Table









Transform + Stage





Copy Data 1 - Source

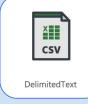




(Output from HW 2)

complaints.txt

1.2GB



Source

Azure Blob Storage

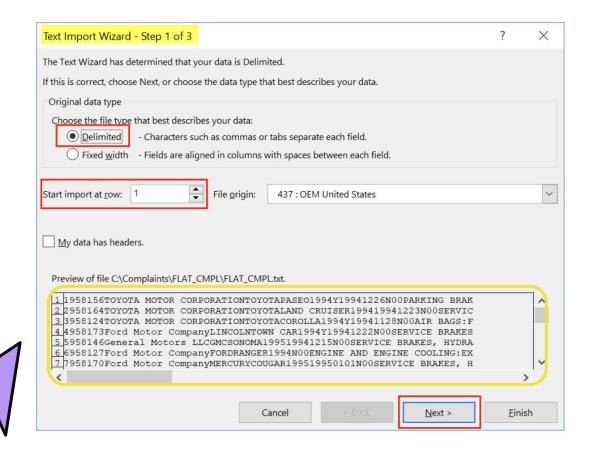
Delimited Text

complaints.txt

*Import Schema - 49 Columns







| is screen lets you s | set the delimiters yo | our data contain | s. You can | see how your text is | s affect | ed in t | the preview be | low. | |
|--|--|-------------------------------------|------------------|--|--|---------|--|------------------|--|
| Delimiters Iab Semicolon Comma Space | Treat cons | secutive delimit | ers as one | | | | | | |
| Other: | | | | | | | | | |
| Data preview 1 958156 1 2 958164 1 3 958124 1 4 958146 5 958146 6 | TOYOTA MOTOR C TOYOTA MOTOR C TOYOTA MOTOR C TOYOTA MOTOR C TOYOTA MOTOR Com Ford Motor Com | ORPORATION ORPORATION pany LLC pany | TOYOTA TOYOTA | PASEO LAND CRUISER COROLLA TOWN CAR SONOMA RANGER COUGAR | 1994 1994 1994 1994 1995 1994 | Y | 19941226 19941223 19941128 19941222 19941215 19950101 | N N N N | |



Copy Data 1 - Sink



Complaints Table

*Authentication: Back to Slide 'Authentication' Sink

SQL Database

Table dbo.complaints

*Import Schema - 49 Columns





Copy Data Overview

Source

(Output from HW 2)

Azure Blob Storage

Delimited Text

.txt



COPY DATA 1



Transform + Stage

Sink

SQL Database

Table dbo.Complaints











'DateA' Column

From 'String' to 'Date' Type

*Type Conversion Setting:

Datetime Format: yyyyMMdd

COPY DATA 1

Transform

* From <u>Reference</u> File

16 DATEA CHAR (8)

(YYYYMMDD) DATE ADDED TO FILE







Create Table

CMPLID INT PRIMARY

- Optional: Add 'NOT NULL'
- Do not add IDENTITY(1,1)

Copy Data 1 'Settings'

*Navigate to **'Fault Tolerance'**

- Skip Incompatible Rows
- ■ Logging
- This is to skip rows affected by human errors

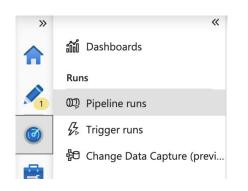
Running Pipeline

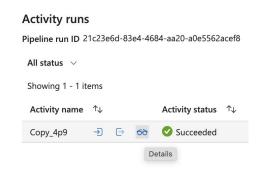
- Pipeline runs on **Azure Cloud**
 - Turning off your computer will not affect the run
- Approx. **2.0 hours** for completion
- **Monitor** to ensure the data is being read & written properly

*RUNNING A PIPELINE

- Publish & Add Trigger >> Cheaper than 'Debug' Action
- Monitor pipeline runs through the **'monitor'** section
 - o Go to details (eyeglasses icon) for details on data read & written)

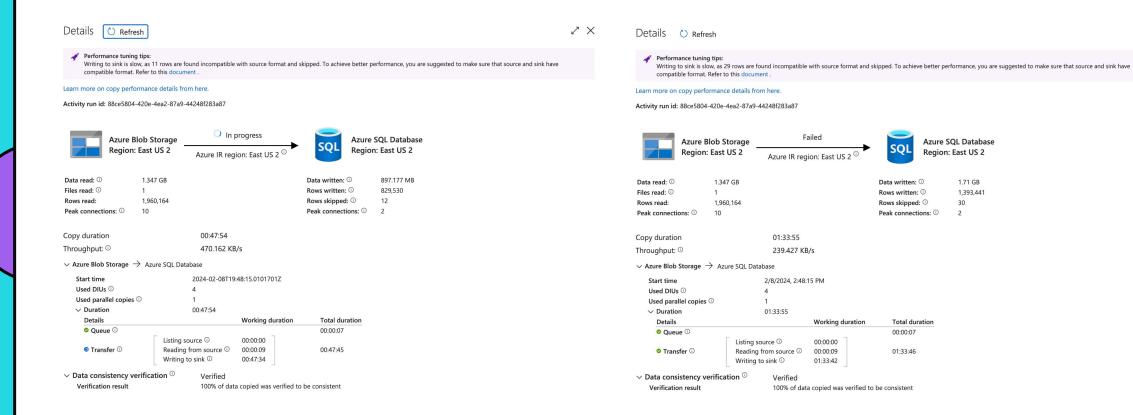












desmania.com





3. Go back to Azure SQL Database

| db47fdf4- | SQL server | East US |
|-----------|--------------|---------|
| dba09cc5- | SQL database | East US |



Azure SQL Database



Azure SQL Server







Final Query:

SELECT *

FROM Complaints

WHERE DATEA = CONVERT(Date, GETDATE()-1)

- ^^ Get Data From Yesterday?
- But... No Data From Yesterday?
- **Change the 1** in GETDATE()-1 to find the most recent date in your data

Find Today's Date

(Azure Time Zone)

SELECT GETDATE()

Azure's Time Zone is not EST

Find Dates in Data:

SELECT DISTINCT CONVERT (VARCHAR, DATEA)

FROM Complaints

ORDER BY CONVERT (VARCHAR, DATEA) DESC

- List the dates in 'DateA' column
- Descending order (Most recent on Top)





HW1c Screenshot

Transforming & Staging

