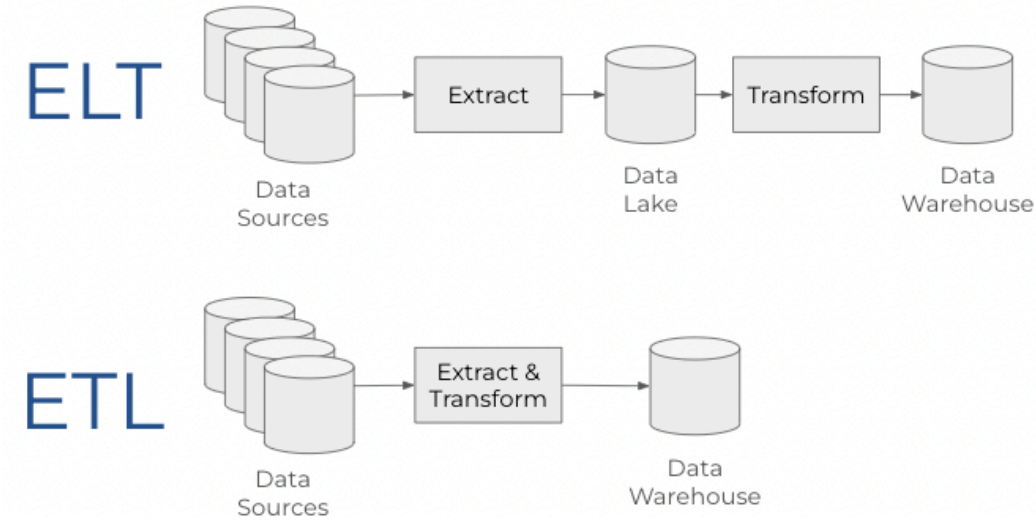
CDS DS 594
Lec 07

Data Transformation

1. Data Transformation
    a. To make data usable for analysis and visualization
2. Data Transformation Tools
    a. 

3. ETL VS ELT
    a. 

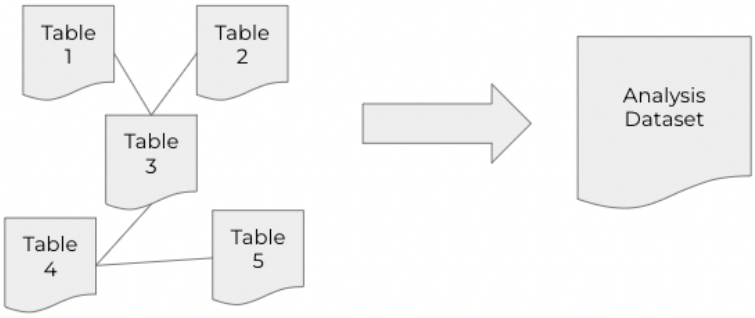
4. Types of Transformation
    a. Parsing
    b. Denormalization
    c. Unpivoting
    d. Enrichment
    e. Imputation

f. Filtering
g. Aggregation
h. Anonymization
i. Typecasting
j. Standardization

5. Parsing

a.

6. Denormalization



a.

7. Unpivot

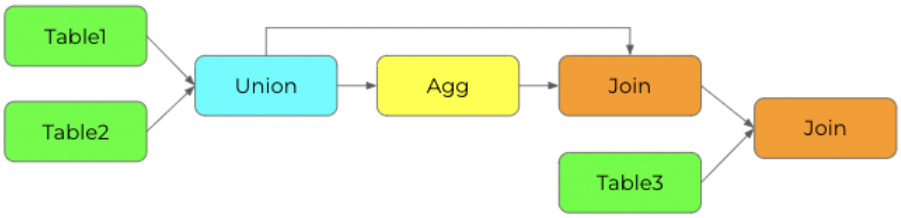
City	Morning temperature	Evening temperature
Austin	62	90.7
Boston	41	48.0
Chicago	51	57.2
Denver	45	52.5

City	Time	Temperature
Austin	Morning	62.0
Austin	Evening	90.7
Boston	Morning	41.0
Boston	Evening	48.0
Chicago	Morning	51.0
Chicago	Evening	57.2
Denver	Morning	45.0
Denver	Evening	52.5

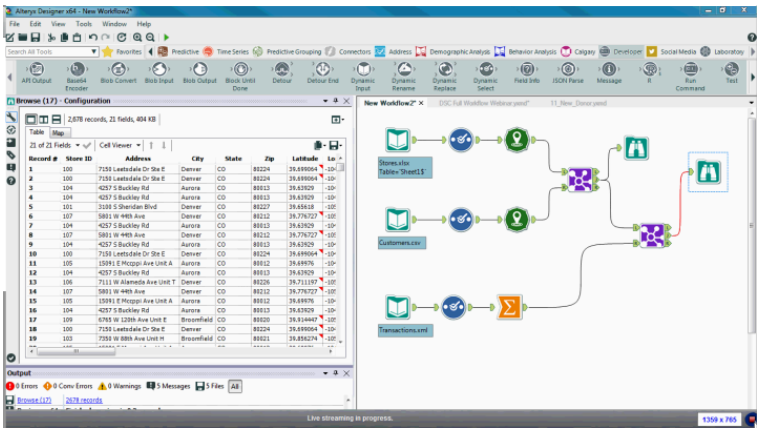
a.

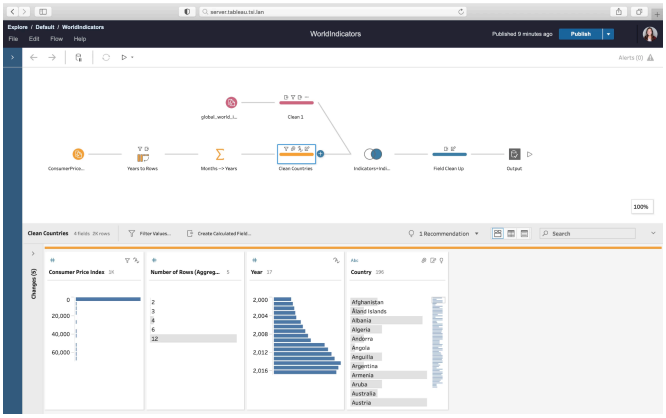
8. Enrichment
    a. Blending data from other sources
        i. Lookups
        ii. Metrics
    b. Adding calculations from existing columns
    c. Appending rows (UNION)

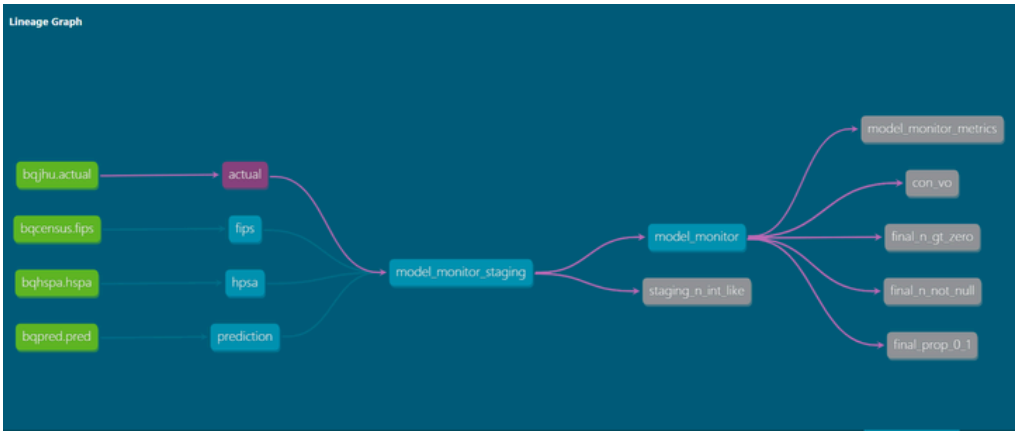
9. Imputation
    a. Filling in missing values
        i. Averages
        ii. Models
        iii. Zeros
10. Other
    a. Filtering: removing unnecessary records or duplicate records
    b. Aggregation: Totals across a group added as a new column
    c. Anonymization: Masking sensitive data
    d. Typecasting: Changing data types (i.e. date saved as a string to a date type)
    e. Standardization: Applying a common value to data that mean the same thing (i.e. MA, Mass, Ma, Massachusetts)
11. Data Lineage
    a.



    b.






Lineage Graph
bqjhu.actual
actual
bqcensus.fips
fips
bqhspa.hspa
hpsa
bqpred.pred
prediction
model_monitor_staging
model_monitor
staging_n_int_like
model_monitor_metrics
con_vo
final_n_gt_zero
final_n_not_null
final_prop_0_1