**Table Creation Queries**

**Create table PRCP:**

CREATE TABLE ghcnblog.CA005020881\_PRCP

WITH(

format = 'PARQUET'

) AS

SELECT

ID,

ELEMENT,

CAST(DATA\_VALUE AS REAL)/10 AS VALUE,

YEAR\_DATE,

SUBSTR(YEAR\_DATE,1,4) AS YEAR,

SUBSTR(YEAR\_DATE,5,2)||'-' || SUBSTR(YEAR\_DATE,7,2) AS MONTH\_DAY,

SUBSTR(YEAR\_DATE,5,2) AS MONTH,

SUBSTR(YEAR\_DATE,7,2) AS DAY

FROM ghcnblog.tblallyears\_qa

WHERE ID LIKE 'CA005020881'

AND ELEMENT IN ('PRCP')

ORDER BY YEAR\_DATE, ELEMENT

**Create Table TMIN:**

CREATE TABLE ghcnblog.CA005020881\_TMIN

WITH(

format = 'PARQUET'

) AS

SELECT

ID,

ELEMENT,

CAST(DATA\_VALUE AS REAL)/10 AS T\_MIN\_VALUE,

YEAR\_DATE,

SUBSTR(YEAR\_DATE,1,4) AS YEAR,

SUBSTR(YEAR\_DATE,5,2)||'-' || SUBSTR(YEAR\_DATE,7,2) AS MONTH\_DAY,

SUBSTR(YEAR\_DATE,5,2) AS MONTH,

SUBSTR(YEAR\_DATE,7,2) AS DAY

FROM ghcnblog.tblallyears\_qa

WHERE ID LIKE 'CA005020881'

AND ELEMENT IN ('TMIN')

ORDER BY YEAR\_DATE, ELEMENT

**Create Table TMAX:**

CREATE TABLE ghcnblog.CA005020881\_TMAX

WITH(

format = 'PARQUET'

) AS

SELECT

ID,

ELEMENT,

CAST(DATA\_VALUE AS REAL)/10 AS T\_MAX\_VALUE,

YEAR\_DATE,

SUBSTR(YEAR\_DATE,1,4) AS YEAR,

SUBSTR(YEAR\_DATE,5,2)||'-' || SUBSTR(YEAR\_DATE,7,2) AS MONTH\_DAY,

SUBSTR(YEAR\_DATE,5,2) AS MONTH,

SUBSTR(YEAR\_DATE,7,2) AS DAY

FROM ghcnblog.tblallyears\_qa

WHERE ID LIKE 'CA005020881'

AND ELEMENT IN ('TMAX')

ORDER BY YEAR\_DATE, ELEMENT

**Create Table Growing Degree Days (GDD):**

CREATE TABLE ghcnblog.CA005020881\_GDD

WITH(

format = 'PARQUET'

) AS

SELECT TMAX.ID,

TMAX.ELEMENT,

TMAX.YEAR\_DATE,

TMAX.YEAR,

TMAX.MONTH\_DAY,

TMAX.MONTH,

TMAX.DAY,

TMAX.T\_MAX\_VALUE,

TMIN.T\_MIN\_VALUE,

TMAX.T\_MAX\_VALUE-TMIN.T\_MIN\_VALUE/2 AS AVG\_TEMP,

IF((TMAX.T\_MAX\_VALUE-TMIN.T\_MIN\_VALUE/2>0) AND (TMIN.T\_MIN\_VALUE>0), 1, 0) AS GDD

FROM ghcnblog.CA005020881\_TMAX AS TMAX,

ghcnblog.CA005020881\_TMIN AS TMIN

WHERE TMAX.YEAR\_DATE = TMIN.YEAR\_DATE;

**Create 10 Year Analysis Table:**

CREATE TABLE ghcnblog.tbl1831every10thyear

WITH (

format='PARQUET',

external\_location='s3://my228bucket/ghcnblog/1831every10years/'

) AS

SELECT TA.id as id, substr(TA.year\_date,1,4) as year, TS.state, CAST(TS.longitude as real) as longitde, CAST(TS.latitude as real) as latitude, element, CAST(data\_value as real) as data\_value

FROM "ghcnblog".tblallyears\_qa as TA, "ghcnblog".tblghcnd\_stations\_qa as TS

WHERE substr(TA.year\_date,1,4) IN ('1831','1841','1851','1861', '1871','1881','1891','1901', '1911','1921','1931','1941', '1951','1961','1971', '1981','1991','2001', '2011', '2021')

AND substr(TA.id,1,2) = 'US'

AND state <> 'PI'

AND TRIM(TA.id) = TRIM(TS.id)

GROUP BY TA.id, substr(TA.year\_date,1,4), state, longitude, latitude, element, data\_value;

**EDA Queries**

**Total No. of Stations:**

SELECT count(\*) AS Total\_Number\_of\_Stations

FROM ghcnblog.tblghcnd\_stations\_qa;

**Element Value:**

SELECT element,

round(avg(CAST(data\_value AS real)/10),2) AS value

FROM ghcnblog.tblallyears\_qa

WHERE element IN ('TMIN', 'TMAX', 'PRCP')

GROUP BY element;

**Number of Stations:**

SELECT DISTINCT id AS numberofstations,

substr(year\_date,1,4) as year

FROM ghcnblog.tblallyears\_qa

GROUP BY substr(year\_date,1,4), id

ORDER BY substr(year\_date,1,4)

**numberofstations year**

**ITE00100554 1763**

**ITE00100554 1764**

**ITE00100554 1765**

**ITE00100554 1766**

**ITE00100554 1767**

**ITE00100554 1768**

**ITE00100554 1769**

**ITE00100554 1770**

**ITE00100554 1771**

**ITE00100554 1772**

**ITE00100554 1773**

**ITE00100554 1774**

**EZE00100082 1775**

**ITE00100554 1775**

**ITE00100554 1776**

**EZE00100082 1776**

**ITE00100554 1777**

**EZE00100082 1777**

**ITE00100554 1778**

**EZE00100082 1778**

**EZE00100082 1779**

**ITE00100554 1779**

**EZE00100082 1780**

**ITE00100554 1780**

**ITE00100554 1781**

**GM000010962 1781**

**EZE00100082 1781**

**ITE00100554 1782**

**EZE00100082 1782**

**GM000010962 1782**

**GM000010962 1783**

**ITE00100554 1783**

**EZE00100082 1783**

**ITE00100554 1784**

**GM000010962 1784**

**EZE00100082 1784**

**ITE00100554 1785**

**EZE00100082 1785**

**GM000010962 1785**

**ITE00100554 1786**

**GM000010962 1786**

**EZE00100082 1786**

**ITE00100554 1787**

**EZE00100082 1787**

**ITE00100554 1788**

**EZE00100082 1788**

**ITE00100554 1789**

**GM000010962 1789**

**EZE00100082 1789**

**GM000010962 1790**

**ITE00100554 1790**

**EZE00100082 1790**

**ITE00100554 1791**

**EZE00100082 1791**

**GM000010962 1791**

**GM000010962 1792**

**EZE00100082 1792**

**ITE00100554 1792**

**EZE00100082 1793**

**ITE00100554 1793**

**EZE00100082 1794**

**GM000010962 1794**

**ITE00100554 1794**

**GM000010962 1795**

**ITE00100554 1795**

**EZE00100082 1795**

**GM000010962 1796**

**EZE00100082 1796**

**ITE00100554 1796**

**ITE00105250 1797**

**ITE00100554 1797**

**GM000010962 1797**

**EZE00100082 1797**

**ITE00100554 1798**

**ITE00105250 1798**

**GM000010962 1798**

**EZE00100082 1798**

**ITE00100554 1799**

**EZE00100082 1799**

**ITE00100554 1800**

**EZE00100082 1800**

**GM000010962 1800**

**GM000010962 1801**

**EZE00100082 1801**

**ITE00100554 1801**

**ITE00105250 1802**

**ITE00100554 1802**

**GM000010962 1802**

**EZE00100082 1802**

**GM000010962 1803**

**EZE00100082 1803**

**ITE00100554 1803**

**GM000010962 1804**

**EZE00100082 1804**

**ITE00100554 1804**

**ITE00100554 1805**

**EZE00100082 1805**

**GM000010962 1805**

**EZE00100082 1806**

**ITE00100554 1806**