ETL – Project for Week 13

Team: Eve Barkley, John McIntyre, Rupesh Patel

Our team decides to look at the National School Lunch Program. We had an unrealistic assumption that the reporting of student participation was managed and stored at the National Level because the program ties into the SNAP program, which uses US Census records and the USDA program for child nutrition. The School Lunch Program was in signed into law in 1946, by Harry Truman. It was originally established as a way to prop up food prices by absorbing farm surpluses, that would provide food to school age children who qualified to get free or reduced lunches at school. The plan was started as subsidy program that gave most of its support as cash reimbursement for meals served. In addition, schools and school districts were strongly encouraged to receive and purchase surpluses of agricultural commodities in their locales. Up until the 1960's counts of participating students and cash payments were managed at a National level.

In 1962 Congress modified the NSLP to adjust fund distribution based on the budget allocated for each state to determine actual need for assistance. Now state boards of education certify the enrollment numbers and the numbers of students who are identified for NLSP and the numbers of students receiving NLSP support.

To access the best possible raw data, our team found two sources:

The USDA Food and Nutrition Service NLSP for annual State totals: https://www.fns.usda.gov/nslp

- This data was collected from Excel spreadsheets.

National Center for Education Statistics for annual National totals for enrollment, testing scores and demographic information on students https://nces.ed.gov/

- This data was collected from Excel spreadsheets.

USDA – for Direct Certification of each States' NLSP over a two-year period
http://www.fns.usda.gov/ops/reseach-and-analysis

- This data was collected from a report pdf.

The data collected had limits that eventually defined the scope of what information over time we could collect. We settled on collecting Annual information by State participation from 2016 to 2019.

We collected National data from 2016 to 2019 as it was available to match to the State collections of data and for student test scores, we were able to find complete datasets for 4th, 8th and 12th grade students. We chose the 8th grade test score set because middle school scores are found to be good indicators of future student performance.

Transform

Data was transformed and saved into CSV files; dates data type was changed to integer using Postgres. All numeric counts data was cast as BigInt to maintain precision. The cash payments data was cast a Double Precision numbers to allow for decimals if there were any.

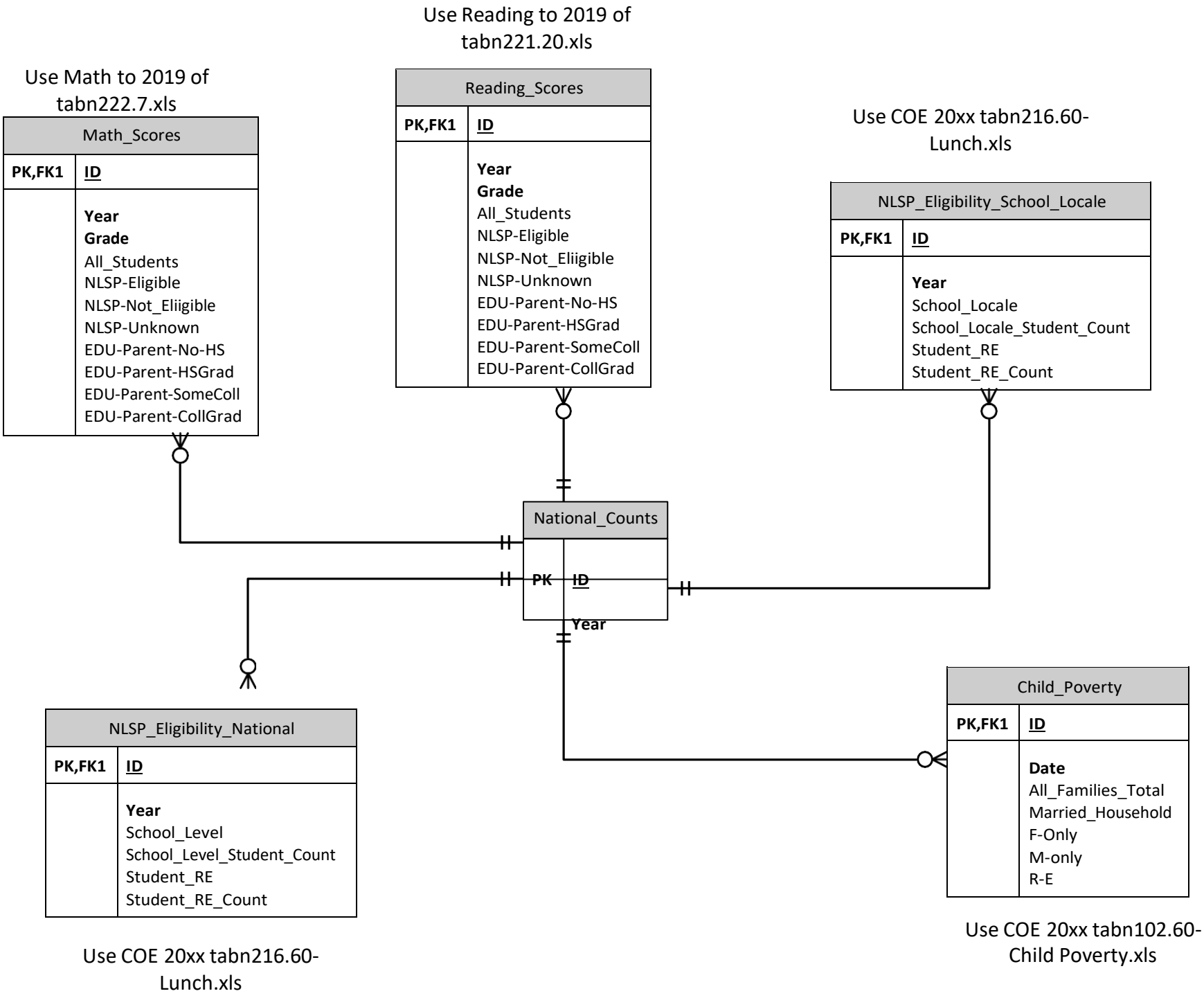
Data was rearranged to decide how the column layout will be for loading into that database. As Postgres needs data entered in a specific way to read.

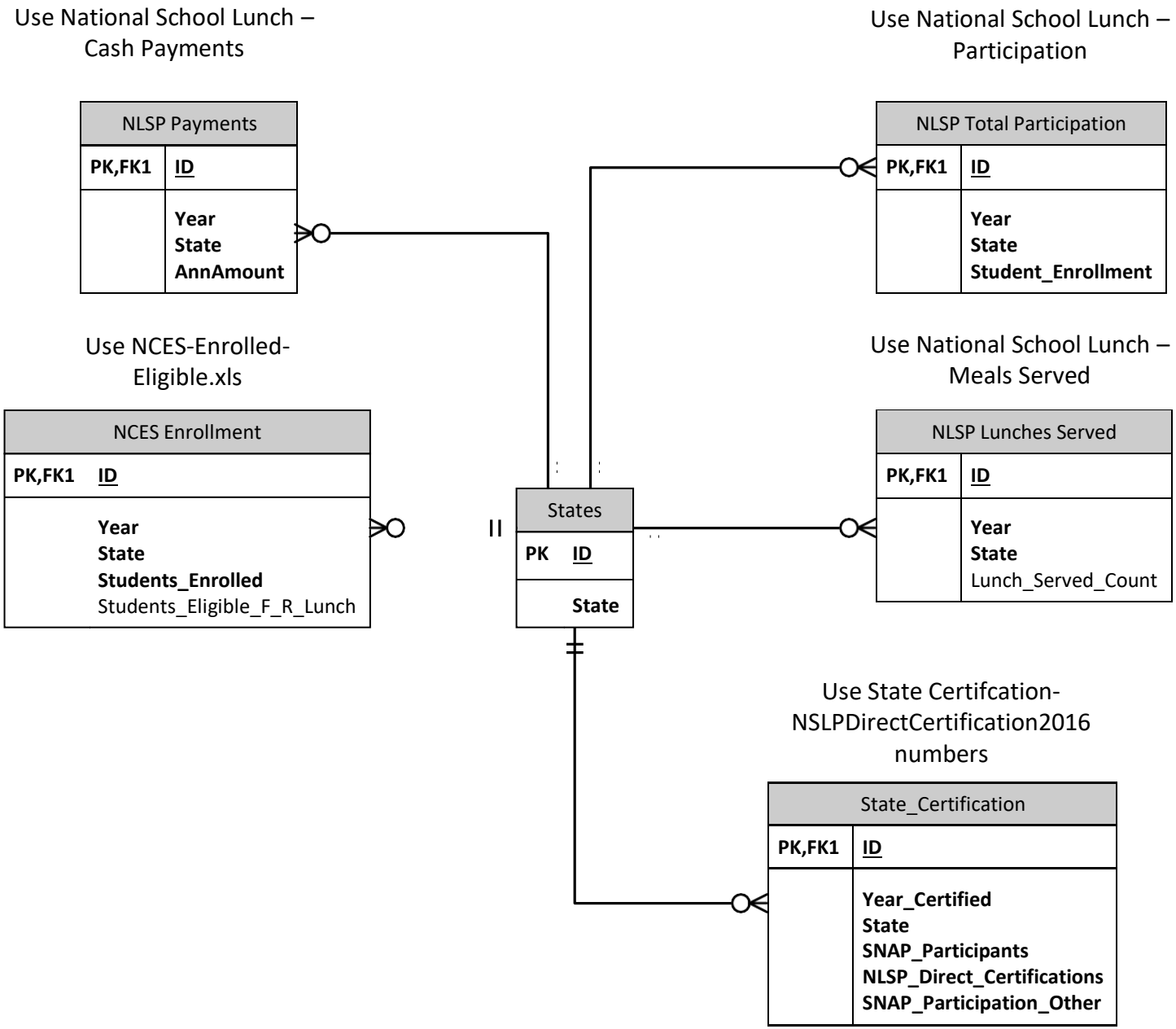
2 tables were created as join tables, one as National Counts and one for State. For the National counts table all data sets were joined based on year. For the States table all data sets were joined by State column. The data was cleaned to take out columns not being used because of redundance or not relevant to our criteria for data collection.

In the Git Repo, the data sets loaded to PostgreSQL database are:
Joined table Lunch Cleaned.csv
Math to 2019 of Cleaned.csv
National_count.csv
Reading to 2019 Cleaned.csv
National School Lunch – Cash Payments Cleaned.csv
National School Lunch – Meals Served Cleaned.csv
National School Lunch – Participation Cleaned.csv
NCES – Enrolled – Eligible Cleaned.csv
SNAP Participation School Year 2015-2016 Cleaned.csv
SNAP Participation School Year 2016-2017 Cleaned.csv
States.csv

The final database design was created for upload to Postgres. The datasets were not granular enough to create a normalized database architecture [Boyce Codd Normal Form]. We found we had two relatable structures using a star pattern. This design still allows for fast queries and will show activity and enrollment numbers in a year or over 5 years. The two-star designs are maintained within the same database to allow some cross-relationships to show student test performance and student poverty across states.

The following designs show the entities to be created and have the raw data files collected identified to be used for loading the datasets.


Use Reading to 2019 of tabn221.20.xls
Use Math to 2019 of tabn222.7.xls
Use COE 20xx tabn216.60-Lunch.xls
Reading_Scores
PK,FK1
ID
Year
Grade
All_Students
NLSP-Eligible
NLSP-Not_Eliigible
NLSP-Unknown
EDU-Parent-No-HS
EDU-Parent-HSGrad
EDU-Parent-SomeColl
EDU-Parent-CollGrad
Math_Scores
PK,FK1
ID
Year
Grade
All_Students
NLSP-Eligible
NLSP-Not_Eliigible
NLSP-Unknown
EDU-Parent-No-HS
EDU-Parent-HSGrad
EDU-Parent-SomeColl
EDU-Parent-CollGrad
NLSP_Eligibility_School_Locale
PK,FK1
ID
Year
School_Locale
School_Locale_Student_Count
Student_RE
Student_RE_Count
National_Counts
PK
ID
Year
NLSP_Eligibility_National
PK,FK1
ID
Year
School_Level
School_Level_Student_Count
Student_RE
Student_RE_Count
Child_Poverty
PK,FK1
ID
Date
All_Families_Total
Married_Household
F-Only
M-only
R-E
Use COE 20xx tabn216.60-Lunch.xls
Use COE 20xx tabn102.60-Child Poverty.xls


Use National School Lunch – Cash Payments
NLSP Payments
PK,FK1 ID
Year
State
AnnAmount
Use National School Lunch – Participation
NLSP Total Participation
PK,FK1 ID
Year
State
Student_Enrollment
Use NCES-Enrolled-Eligible.xls
NCES Enrollment
PK,FK1 ID
Year
State
Students_Enrolled
Students_Eligible_F_R_Lunch
States
PK ID
State
Use National School Lunch – Meals Served
NLSP Lunches Served
PK,FK1 ID
Year
State
Lunch_Served_Count
Use State Certifcation-NSLPDirectCertification2016 numbers
State_Certification
PK,FK1 ID
Year_Certified
State
SNAP_Participants
NLSP_Direct_Certifications
SNAP_Participation_Other

PostgreSQL DDL script:

```
-- Database: ETL_DB
-- DROP DATABASE "ETL_DB";
CREATE DATABASE "ETL_DB"
    WITH
    OWNER = postgres
    ENCODING = 'UTF8'
    LC_COLLATE = 'English_United States.1252'
    LC_CTYPE = 'English_United States.1252'
    TABLESPACE = pg_default
    CONNECTION LIMIT = -1;
-- Table: public.Child_Poverty
-- DROP TABLE public."Child_Poverty";
CREATE TABLE public."Child_Poverty"
(
    "ID" integer NOT NULL GENERATED ALWAYS AS IDENTITY ( INCREMENT 1 START 100 MINVALUE 100
MAXVALUE 1000 CACHE 1 ),
    "Date" date NOT NULL,
    "All_Families_Total" bigint NOT NULL,
    "Married_Household" bigint NOT NULL,
    "F-Only" bigint NOT NULL,
    "M-Only" bigint NOT NULL,
    "Race/Ethnicity" "char" NOT NULL,
    CONSTRAINT "Child_Poverty_pkey" PRIMARY KEY ("ID"),
    CONSTRAINT "National_Counts_Child_Poverty_FK" FOREIGN KEY ("ID")
        REFERENCES public."National_Counts" ("ID") MATCH FULL
        ON UPDATE NO ACTION
        ON DELETE NO ACTION
)

TABLESPACE pg_default;
ALTER TABLE public."Child_Poverty"
    OWNER to postgres;
-- Table: public.Math_Scores
-- DROP TABLE public."Math_Scores";
CREATE TABLE public."Math_Scores"
(
    "ID" bigint NOT NULL GENERATED ALWAYS AS IDENTITY ( INCREMENT 1 START 100 MINVALUE 100
MAXVALUE 1000 CACHE 1 ),
    "Year" date NOT NULL,
    "Grade" "char"[] NOT NULL,
    "All_Students" bigint NOT NULL,
    "NLSP_Eligible" bigint NOT NULL,
    "NLSP_Not_Eligible" bigint NOT NULL,
    "NLSP_Unknown" bigint NOT NULL,
    "EDU_Parent_No_HS" bigint NOT NULL,
    "EDU_Parent_HS_Grad" bigint NOT NULL,
```

```
    "EDU_Parent_Some_College" bigint NOT NULL,
    "EDU_Parent_CollGrad" bigint NOT NULL,
    CONSTRAINT "Math_Scores_pkey" PRIMARY KEY ("ID"),
    CONSTRAINT "National_Counts_Math_Scores_FK" FOREIGN KEY ("ID")
        REFERENCES public."National_Counts" ("ID") MATCH SIMPLE
        ON UPDATE NO ACTION
        ON DELETE NO ACTION
)

TABLESPACE pg_default;
ALTER TABLE public."Math_Scores"
    OWNER to postgres;
-- Table: public.NLSP_Eligibility_School_Locale
-- DROP TABLE public."NLSP_Eligibility_School_Locale";
CREATE TABLE public."NLSP_Eligibility_School_Locale"
(
    "ID" bigint NOT NULL GENERATED ALWAYS AS IDENTITY ( INCREMENT 1 START 100 MINVALUE 100
MAXVALUE 1000 CACHE 1 ),
    "Year" date NOT NULL,
    "School_Locale" "char"[] NOT NULL,
    "School_Locale_Student_Count" bigint NOT NULL,
    "Student_R_E" "char"[] NOT NULL,
    "Student_R_E_Count" bigint NOT NULL,
    CONSTRAINT "NLSP_Eligibility_School_Locale_pkey" PRIMARY KEY ("ID"),
    CONSTRAINT "National_Counts_NLSP_School_Locale_FK" FOREIGN KEY ("ID")
        REFERENCES public."National_Counts" ("ID") MATCH SIMPLE
        ON UPDATE NO ACTION
        ON DELETE NO ACTION
)
TABLESPACE pg_default;
ALTER TABLE public."NLSP_Eligibility_School_Locale"
    OWNER to postgres;
-- Table: public.NLSP_Eligibility_School_National
-- DROP TABLE public."NLSP_Eligibility_School_National";
CREATE TABLE public."NLSP_Eligibility_School_National"
(
    "ID" bigint NOT NULL GENERATED ALWAYS AS IDENTITY ( INCREMENT 1 START 100 MINVALUE 100
MAXVALUE 1000 CACHE 1 ),
    "Year" date NOT NULL,
    "School_Level" "char"[] NOT NULL,
    "School_Level_Student_Count" bigint NOT NULL,
    "Student_RE" character(1)[] COLLATE pg_catalog."default" NOT NULL,
    "Student_RE_Count" bigint NOT NULL,
    CONSTRAINT "NLSP_Eligibility_School_National_pkey" PRIMARY KEY ("ID"),
    CONSTRAINT "National_Counts_NLSP_National_FK" FOREIGN KEY ("ID")
        REFERENCES public."National_Counts" ("ID") MATCH SIMPLE
        ON UPDATE NO ACTION
        ON DELETE NO ACTION
```

```
)
TABLESPACE pg_default;
ALTER TABLE public."NLSP_Eligibility_School_National"
    OWNER to postgres;
-- Table: public.National_Counts
-- DROP TABLE public."National_Counts";
CREATE  TABLE  public."National_Counts"
(
    "ID" bigint NOT NULL GENERATED ALWAYS AS IDENTITY ( INCREMENT 1 START 100 MINVALUE 100
MAXVALUE 1000 CACHE 1 ),
    "Year" date NOT NULL,
    CONSTRAINT "National_Counts_pkey" PRIMARY KEY ("ID")
)

TABLESPACE pg_default;
ALTER TABLE public."National_Counts"
    OWNER to postgres;
CREATE TABLE public."Reading_Scores"
(
    "Year" date NOT NULL,
    "All Students" bigint NOT NULL,
    "NLSP_Eligible" bigint NOT NULL,
    "NLSP_Not_Eligible" bigint NOT NULL,
    "NLSP-Unknown" bigint NOT NULL,
    "EDU-Parent-No HS" bigint NOT NULL,
    "EDU-Parent-HSGrad" bigint NOT NULL,
    "EDU-Parent-Some-College" bigint NOT NULL,
    "EDU-Parent-CollGrad" bigint[] NOT NULL,
    "ID" bigint NOT NULL GENERATED ALWAYS AS IDENTITY ( INCREMENT 1 START 100 MINVALUE 100
MAXVALUE 1000 CACHE 1 ),
    CONSTRAINT "Reading_Scores_pkey" PRIMARY KEY ("ID"),
    CONSTRAINT "National_Counts_Reading_Scores_FK" FOREIGN KEY ("ID")
        REFERENCES public."National_Counts" ("ID") MATCH SIMPLE
        ON UPDATE NO ACTION
        ON DELETE NO ACTION
)
TABLESPACE pg_default;
ALTER TABLE public."Reading_Scores"
    OWNER to postgres;
-- Table: public.S_NCES_Enrollment
-- DROP TABLE public."S_NCES_Enrollment";
CREATE  TABLE  public."S_NCES_Enrollment"
(
    "ID" bigint NOT NULL,
    "Year" date NOT NULL,
    "State" "char"[] NOT NULL,
    "Students_Enrolled" bigint NOT NULL,
    "Students_Eligible_FR_Lunch" bigint NOT NULL,
```

```
    CONSTRAINT "S_NCES_Enrollment_pkey" PRIMARY KEY ("ID"),
    CONSTRAINT "States_NCES_Enrollment" FOREIGN KEY ("ID")
        REFERENCES public."States" ("ID") MATCH SIMPLE
        ON UPDATE NO ACTION
        ON DELETE NO ACTION
)

TABLESPACE pg_default;

ALTER TABLE public."S_NCES_Enrollment"
    OWNER to postgres;

-- Table: public.S_NLSP_Lunches_Served

-- DROP TABLE public."S_NLSP_Lunches_Served";

CREATE TABLE public."S_NLSP_Lunches_Served"
(
    "ID" bigint NOT NULL,
    "Year" date NOT NULL,
    "State" "char"[] NOT NULL,
    "Lunches_Served_Count" bigint NOT NULL,
    CONSTRAINT "S_NLSP_Lunches_Served_pkey" PRIMARY KEY ("ID"),
    CONSTRAINT "States_Lunches_FK" FOREIGN KEY ("ID")
        REFERENCES public."States" ("ID") MATCH SIMPLE
        ON UPDATE NO ACTION
        ON DELETE NO ACTION
)

TABLESPACE pg_default;

ALTER TABLE public."S_NLSP_Lunches_Served"
    OWNER to postgres;
-- Table: public.S_NLSP_Participation
-- DROP TABLE public."S_NLSP_Participation";
CREATE TABLE public."S_NLSP_Participation"
(
    "ID" bigint NOT NULL,
    "Year" date NOT NULL,
    "State" "char"[] NOT NULL,
    "Student_Enrollment" bigint NOT NULL,
    CONSTRAINT "S_NLSP_Participation_pkey" PRIMARY KEY ("ID"),
    CONSTRAINT "States_Participation_FK" FOREIGN KEY ("ID")
        REFERENCES public."States" ("ID") MATCH SIMPLE
        ON UPDATE NO ACTION
        ON DELETE NO ACTION
)
```

```
TABLESPACE pg_default;
ALTER TABLE public."S_NLSP_Participation"
    OWNER to postgres;
-- Table: public.S_NLSP_Payments
-- DROP TABLE public."S_NLSP_Payments";
CREATE TABLE public."S_NLSP_Payments"
(
    "ID" bigint NOT NULL,
    "State" "char"[] NOT NULL,
    "AnnAmount" double precision NOT NULL,
    CONSTRAINT "S_NLSP_Payments_pkey" PRIMARY KEY ("ID"),
    CONSTRAINT "States_Payments_FK" FOREIGN KEY ("ID")
        REFERENCES public."States" ("ID") MATCH SIMPLE
        ON UPDATE NO ACTION
        ON DELETE NO ACTION
)
TABLESPACE pg_default;
ALTER TABLE public."S_NLSP_Payments"
    OWNER to postgres;
-- Table: public.State_Certification
-- DROP TABLE public."State_Certification";
CREATE TABLE public."State_Certification"
(
    "ID" bigint NOT NULL,
    "Year_Certified" date NOT NULL,
    "State" "char"[] NOT NULL,
    "SNAP_Participants" bigint NOT NULL,
    "NLSP_Direct_Certifications" bigint NOT NULL,
    CONSTRAINT "State_Certification_pkey" PRIMARY KEY ("ID"),
    CONSTRAINT "States_State_Certification_FK" FOREIGN KEY ("ID")
        REFERENCES public."States" ("ID") MATCH SIMPLE
        ON UPDATE NO ACTION
        ON DELETE NO ACTION
)
TABLESPACE pg_default;
ALTER TABLE public."State_Certification"
    OWNER to postgres;
-- Table: public.States
-- DROP TABLE public."States";
CREATE TABLE public."States"
(
    "ID" bigint NOT NULL,
    "State" "char"[] NOT NULL,
    CONSTRAINT "States_pkey" PRIMARY KEY ("ID")
)
TABLESPACE pg_default;
ALTER TABLE public."States"
    OWNER to postgres;
```

References:

National School Lunch Program | USDA-FNS

In FY 2019, schools served over 4.8 billion lunches to children nationwide. In FY 2019, schools served over 4.8 billion lunches to children nationwide. The National School Lunch Program (NSLP) is a federally assisted meal program operating in public and nonprofit private schools and residential ...
https://www.fns.usda.gov/nslp

National Center for Education Statistics (NCES) Home Page, part of the U.S. Department of Education
https://nces.ed.gov/
Digest of Education Statistics,
https://nces.ed.gov/programs/digest/

How the Census Bureau Measures Poverty
Bureau
https://www.census.gov/topics/income-poverty/poverty/guidance/poverty-measures.html

US Open Data Portal, data.gov's datasets
https://data.world/datagov-us

The President & First Lady on Child Nutrition Bill: "The Basic Nutrition They Need to Learn and Grow and to Pursue Their Dreams"
https://obamawhitehouse.archives.gov/healthykids

Healthy, Hunger-Free Kids Act of 2010 - Wikipedia
The Healthy, Hunger-Free Kids Act of 2010 (Pub.L. 111–296 (text)) is a federal statute signed into law by President Barack Obama on December 13, 2010. The law is part of the reauthorization of funding for child nutrition (see the original Child Nutrition Act).It funded child nutrition programs and free lunch programs in schools for 5 years. In addition, the law set new nutrition standards ...
https://en.wikipedia.org/wiki/Healthy,_Hunger-Free_Kids_Act_of_2010

National School Lunch Act - Wikipedia
The Richard B. Russell National School Lunch Act (79 P.L. 396, 60 Stat. 230) is a 1946 United States federal law that created the National School Lunch Program (NSLP) to provide low-cost or free school lunch meals to qualified students through subsidies to schools. The program was established as a way to prop up food prices by absorbing farm surpluses, while at the same time providing food to ...
https://en.wikipedia.org/wiki/National_School_Lunch_Act