Graduation Rates Data Pipeline

Name: Cheryl Machingura

Course: Big Data Final Project

Date: May 2025

Introduction

This project demonstrates a complete big data pipeline using NiFi, HDFS, Hive, Spark and Hbase. The goal was to ingest, store, transform, and load graduation rate data through four major components of the big data ecosystem.

Dataset

The dataset used for this project is `states_all.csv`, which contains U.S. state-level education data. It was sourced from Kaggle and the added to my GitHub repository:

https://github.com/CherylMachingura/hdfs/blob/main/states_all.csv

This is a link to the raw file:
https://raw.githubusercontent.com/CherylMachingura/hdfs/main/states_all.csv

Pipeline Overview

The following technologies and components were used:

- Apache NiFi: for ingestion via InvokeHTTP
- HDFS: for storing raw and cleaned files
- Apache Hive: for schema definition and querying
- Apache Spark: for transformation of CSV data
- Apache HBase: for storing the cleaned data in columnar NoSQL format

Issues Encountered & Resolutions

Several issues were encountered during this pipeline:
a) Port Conflicts – Docker containers could not start due to port 8080 already being in use. This was resolved by stopping other containers and reassigning ports.
b) Stuck Spark Jobs – Some Spark jobs hung indefinitely. Resolution involved restarting the YARN ResourceManager and ensuring proper memory allocation.
c) NiFi to HDFS Path Validation – NiFi's PutHDFS processor failed until correct core-site.xml and hdfs-site.xml paths were provided inside the container.
d) Data Not Visible in Hive – Table loaded but data did not appear; the issue was due to a mismatch in schema format and missing headers.

Step 1: Ingest Data Using Apache NiFi

1.1 Access the NiFi Web UI

Open your browser and go to: http://localhost:8080/nifi

You should see the NiFi canvas workspace where you can drag and drop processors.

1.2 Add Required Processors

You'll use two main processors:

- GetHTTP – to download your CSV file from GitHub.
- PutHDFS – to write the downloaded file into HDFS.

Instructions:

1. Click the Processor icon (sixth icon from left) in the top menu.
2. In the dialog box:
   - Type `GetHTTP`, select it, and click Add.
   - Repeat to add `PutHDFS`.

1.3 Configure GetHTTP Processor

1. Double-click the GetHTTP processor.
2. Under Settings tab:
   - Change the name to something meaningful like `Fetch Education CSV`.
3. Under the Properties tab:
   - URL:
     https://raw.githubusercontent.com/CherylMachingura/hdfs/main/states_all.csv
   - (Optional) Set Scheduling to run every 5 minutes or on demand for testing.
4. Click Apply.

1.4 Configure PutHDFS Processor

1. Double-click the PutHDFS processor.
2. Under the Settings tab, name it `Write to HDFS`.
3. Under the Properties tab, set:
   - Directory:
     /data/education/raw/
   - Hadoop Configuration Resources: Point this to your core-site.xml and hdfs-site.xml if not

preloaded.

4. Click Apply.

1.5 Connect the Processors

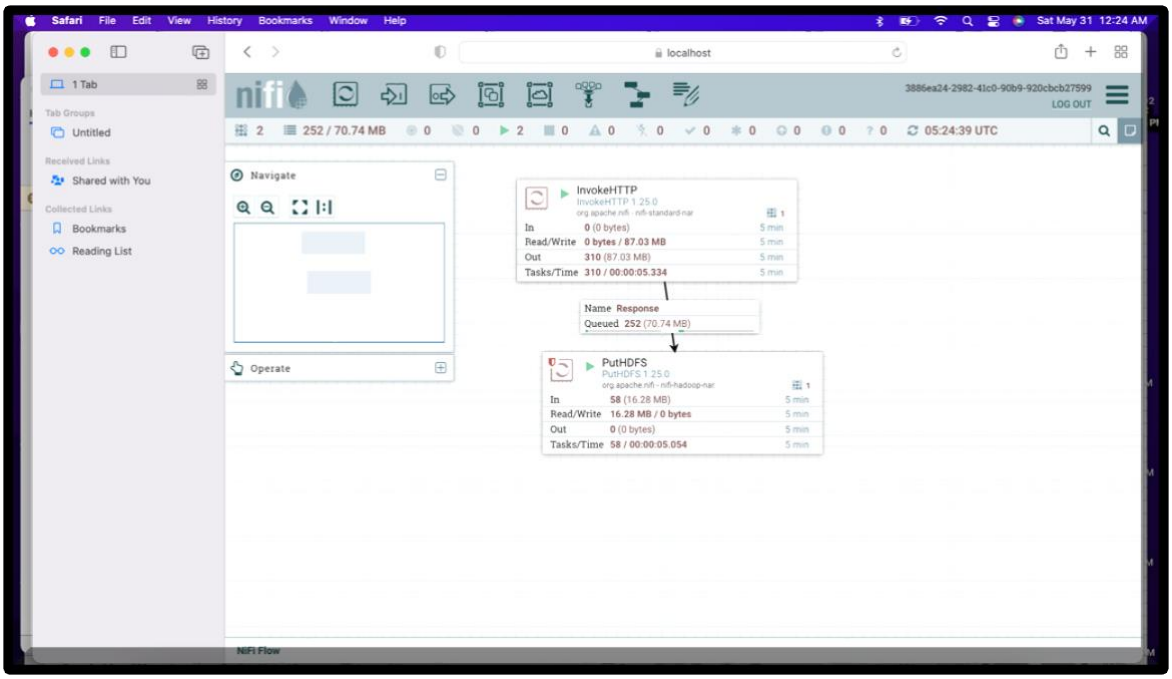
1. Hover over GetHTTP, click and drag the arrow to PutHDFS.

2. Choose `success` as the relationship.

1.6 Start the Flow

1. Right-click on each processor and choose Start.

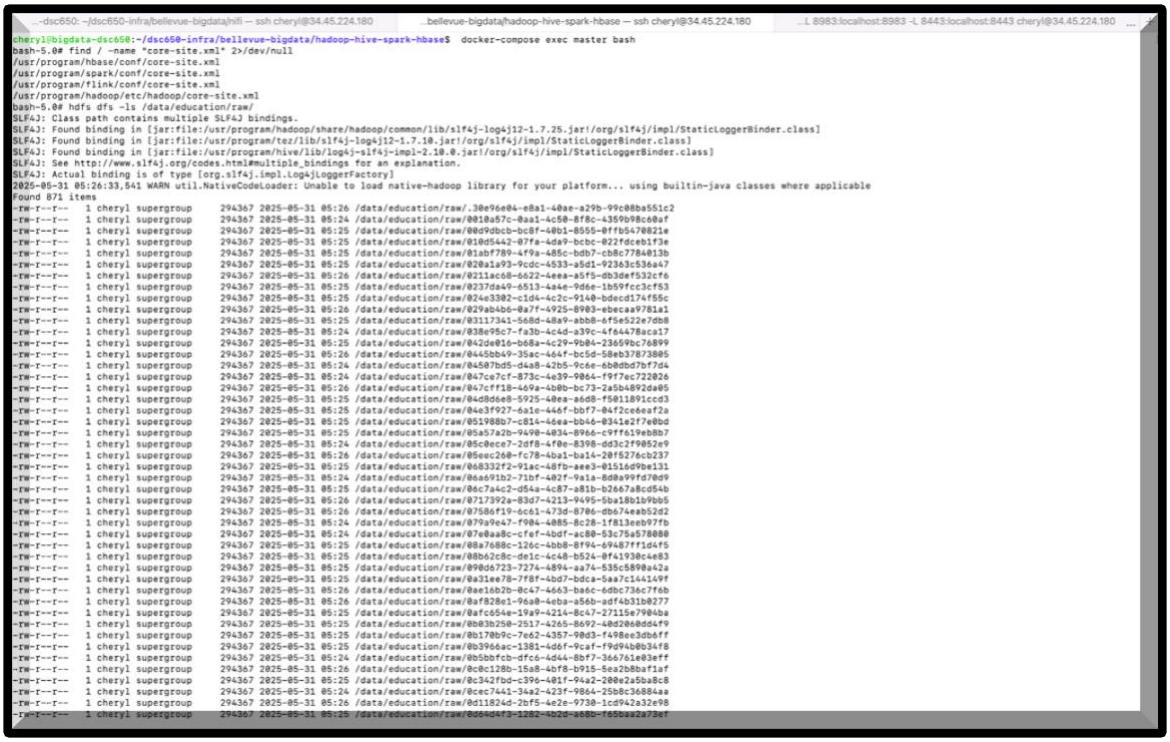
2. Watch the data flow through the queue — it should move from GetHTTP → PutHDFS.



1.7 Confirm HDFS Output

Run the following command in your terminal to verify the file is in HDFS:

hdfs dfs -ls /data/education/raw/


..-dsc650: ~/dsc650-infra/bellevue-bigdata/nifi — ssh cheryl@34.45.224.180
..bellevue-bigdata/hadoop-hive-spark-hbase — ssh cheryl@34.45.224.180
...L 8983:localhost:8983 -L 8443:localhost:8443 cheryl@34.45.224.180
cheryl@bigdata-dsc650:~/dsc650-infra/bellevue-bigdata/hadoop-hive-spark-hbase$ docker-compose exec master bash
bash-5.0# find / -name "core-site.xml" 2>/dev/null
/usr/program/hbase/conf/core-site.xml
/usr/program/spark/conf/core-site.xml
/usr/program/flink/conf/core-site.xml
/usr/program/hadoop/etc/hadoop/core-site.xml
bash-5.0# hdfs dfs -ls /data/education/raw/
SLF4J: Class path contains multiple SLF4J bindings.
SLF4J: Found binding in [jar:file:/usr/program/hadoop/share/hadoop/common/lib/slf4j-log4j12-1.7.25.jar!/org/slf4j/impl/StaticLoggerBinder.class]
SLF4J: Found binding in [jar:file:/usr/program/tez/lib/slf4j-log4j12-1.7.10.jar!/org/slf4j/impl/StaticLoggerBinder.class]
SLF4J: Found binding in [jar:file:/usr/program/hive/lib/log4j-slf4j-impl-2.10.0.jar!/org/slf4j/impl/StaticLoggerBinder.class]
SLF4J: See http://www.slf4j.org/codes.html#multiple_bindings for an explanation.
SLF4J: Actual binding is of type [org.slf4j.impl.Log4jLoggerFactory]
2025-05-31 05:26:33,541 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
Found 871 items
-rw-r--r-- 1 cheryl supergroup 294367 2025-05-31 05:26 /data/education/raw/.30e96e04-e8a1-40ae-a29b-99c08ba551c2
-rw-r--r-- 1 cheryl supergroup 294367 2025-05-31 05:24 /data/education/raw/0010a57c-0aa1-4c50-8f8c-4359b98c60af
-rw-r--r-- 1 cheryl supergroup 294367 2025-05-31 05:25 /data/education/raw/00d9dbcb-bc8f-40b1-8555-0ffb5470821e
-rw-r--r-- 1 cheryl supergroup 294367 2025-05-31 05:25 /data/education/raw/010d5442-07fa-4da9-bcbc-022fdceb1f3e
-rw-r--r-- 1 cheryl supergroup 294367 2025-05-31 05:25 /data/education/raw/01abf789-4f9a-485c-bdb7-cb8c7784013b
-rw-r--r-- 1 cheryl supergroup 294367 2025-05-31 05:25 /data/education/raw/020a1a93-9cdc-4533-a5d1-923b3c536a47
-rw-r--r-- 1 cheryl supergroup 294367 2025-05-31 05:26 /data/education/raw/0211ac68-6622-4eea-a5f5-db3def532cf6
-rw-r--r-- 1 cheryl supergroup 294367 2025-05-31 05:25 /data/education/raw/0237da49-6513-4a4e-9d6e-1b59fcc3cf53
-rw-r--r-- 1 cheryl supergroup 294367 2025-05-31 05:25 /data/education/raw/024e3302-c1d4-4c2c-9140-bdecd174f55c
-rw-r--r-- 1 cheryl supergroup 294367 2025-05-31 05:26 /data/education/raw/029ab4b6-0a7f-4925-8903-ebecaa9781a1
-rw-r--r-- 1 cheryl supergroup 294367 2025-05-31 05:25 /data/education/raw/03117341-568d-48a9-abb8-6f5e522e7db8
-rw-r--r-- 1 cheryl supergroup 294367 2025-05-31 05:24 /data/education/raw/038e95c7-fa3b-4c4d-a39c-4f64478aca17
-rw-r--r-- 1 cheryl supergroup 294367 2025-05-31 05:25 /data/education/raw/042de016-b68a-4c29-9b04-23659bc76899
-rw-r--r-- 1 cheryl supergroup 294367 2025-05-31 05:26 /data/education/raw/0445bb49-35ac-464f-bc5d-58eb37873805
-rw-r--r-- 1 cheryl supergroup 294367 2025-05-31 05:24 /data/education/raw/04507bd5-d4a8-42b5-9c6e-6b0dbd7bf7d4
-rw-r--r-- 1 cheryl supergroup 294367 2025-05-31 05:24 /data/education/raw/047ce7cf-873c-4e39-9864-f9f7ec722826
-rw-r--r-- 1 cheryl supergroup 294367 2025-05-31 05:26 /data/education/raw/047cff18-469a-4b0b-bc73-2a5b4892da05
-rw-r--r-- 1 cheryl supergroup 294367 2025-05-31 05:25 /data/education/raw/04d8d6e8-5925-40ea-a6d8-f5011891ccd3
-rw-r--r-- 1 cheryl supergroup 294367 2025-05-31 05:25 /data/education/raw/04e3f927-6a1e-446f-bbf7-04f2ce6eaf2a
-rw-r--r-- 1 cheryl supergroup 294367 2025-05-31 05:25 /data/education/raw/051988b7-c814-46ea-bb46-0341e2f7e0bd
-rw-r--r-- 1 cheryl supergroup 294367 2025-05-31 05:25 /data/education/raw/05a57a2b-9498-4034-8966-c9ffb19eb8b7
-rw-r--r-- 1 cheryl supergroup 294367 2025-05-31 05:24 /data/education/raw/05c0ece7-2df8-4f0e-8398-dd3c2f9052e9
-rw-r--r-- 1 cheryl supergroup 294367 2025-05-31 05:26 /data/education/raw/05eec260-fc78-4ba1-ba14-20f5276cb237
-rw-r--r-- 1 cheryl supergroup 294367 2025-05-31 05:25 /data/education/raw/068332f2-91ac-48fb-aee3-01516d9be131
-rw-r--r-- 1 cheryl supergroup 294367 2025-05-31 05:24 /data/education/raw/06aa91b2-71bf-482f-9a1a-8d0a99fd70d9
-rw-r--r-- 1 cheryl supergroup 294367 2025-05-31 05:25 /data/education/raw/06c7a4c2-d54a-4c87-a81b-b2667a8cd54b
-rw-r--r-- 1 cheryl supergroup 294367 2025-05-31 05:26 /data/education/raw/0717392a-83d7-4213-9495-5ba18b1b9bb5
-rw-r--r-- 1 cheryl supergroup 294367 2025-05-31 05:26 /data/education/raw/07586f19-6c61-473d-8706-db674eab52d2
-rw-r--r-- 1 cheryl supergroup 294367 2025-05-31 05:24 /data/education/raw/079a9e47-f904-4085-8c28-1f813eeb97fb
-rw-r--r-- 1 cheryl supergroup 294367 2025-05-31 05:24 /data/education/raw/07e0aa8c-cfef-4bdf-ac80-53c75a578080
-rw-r--r-- 1 cheryl supergroup 294367 2025-05-31 05:25 /data/education/raw/08a7688c-126c-4bb8-8f94-694877f1d4f5
-rw-r--r-- 1 cheryl supergroup 294367 2025-05-31 05:25 /data/education/raw/08b62c8c-de1c-4c48-b524-0f41930c4e83
-rw-r--r-- 1 cheryl supergroup 294367 2025-05-31 05:25 /data/education/raw/090d6723-7274-4894-aa74-535c5898a42a
-rw-r--r-- 1 cheryl supergroup 294367 2025-05-31 05:25 /data/education/raw/0a31ee78-7f8f-4bd7-bdca-5aa7c144149f
-rw-r--r-- 1 cheryl supergroup 294367 2025-05-31 05:26 /data/education/raw/0ae16b2b-0c47-4663-ba6c-6dbc736c7f6b
-rw-r--r-- 1 cheryl supergroup 294367 2025-05-31 05:26 /data/education/raw/0af828e1-96a0-4eba-a56b-adf4b31b0277
-rw-r--r-- 1 cheryl supergroup 294367 2025-05-31 05:25 /data/education/raw/0afc654e-19a9-4214-8c47-27115e790dba
-rw-r--r-- 1 cheryl supergroup 294367 2025-05-31 05:25 /data/education/raw/0b03b250-2517-4265-8692-40d206ddd4f9
-rw-r--r-- 1 cheryl supergroup 294367 2025-05-31 05:25 /data/education/raw/0b170b9c-7e62-4357-98d3-f498ee3db6ff
-rw-r--r-- 1 cheryl supergroup 294367 2025-05-31 05:25 /data/education/raw/0b3966ac-1381-4d6f-9caf-f9d94b0b34f8
-rw-r--r-- 1 cheryl supergroup 294367 2025-05-31 05:24 /data/education/raw/0b5bbfcb-dfc6-4d44-8bf7-36e761e03eff
-rw-r--r-- 1 cheryl supergroup 294367 2025-05-31 05:26 /data/education/raw/0c0c128b-15a8-4bf8-b915-5ea2b8baf1af
-rw-r--r-- 1 cheryl supergroup 294367 2025-05-31 05:25 /data/education/raw/0c342fbd-c396-401f-94a2-200e2a5ba8c8
-rw-r--r-- 1 cheryl supergroup 294367 2025-05-31 05:24 /data/education/raw/0cec7441-34a2-423f-9864-25b8c36884aa
-rw-r--r-- 1 cheryl supergroup 294367 2025-05-31 05:26 /data/education/raw/0d11824d-2bf5-4e2e-9730-1cd942a32e98
-rw-r--r-- 1 cheryl supergroup 294367 2025-05-31 05:25 /data/education/raw/0d64d4f3-1282-4b2d-a68b-f65baa2a73ef

Step 2: Create Hive External and Cleaned Tables

2.1 Access Hive Command Line

1. SSH into your Hadoop master container:

```
docker-compose exec master bash
```

2. Start Hive shell:

```
hive
```

2.2 Create External Table from Raw CSV

This table will read data from the HDFS location populated by NiFi.

```
CREATE EXTERNAL TABLE education_raw (
  year STRING,
  state STRING,
  total_expenditure STRING,
  enrollment STRING,
  expenditure_per_student STRING
)
ROW FORMAT DELIMITED
FIELDS TERMINATED BY ','
STORED AS TEXTFILE
LOCATION '/data/education/raw/'
TBLPROPERTIES ("skip.header.line.count"="1");
```

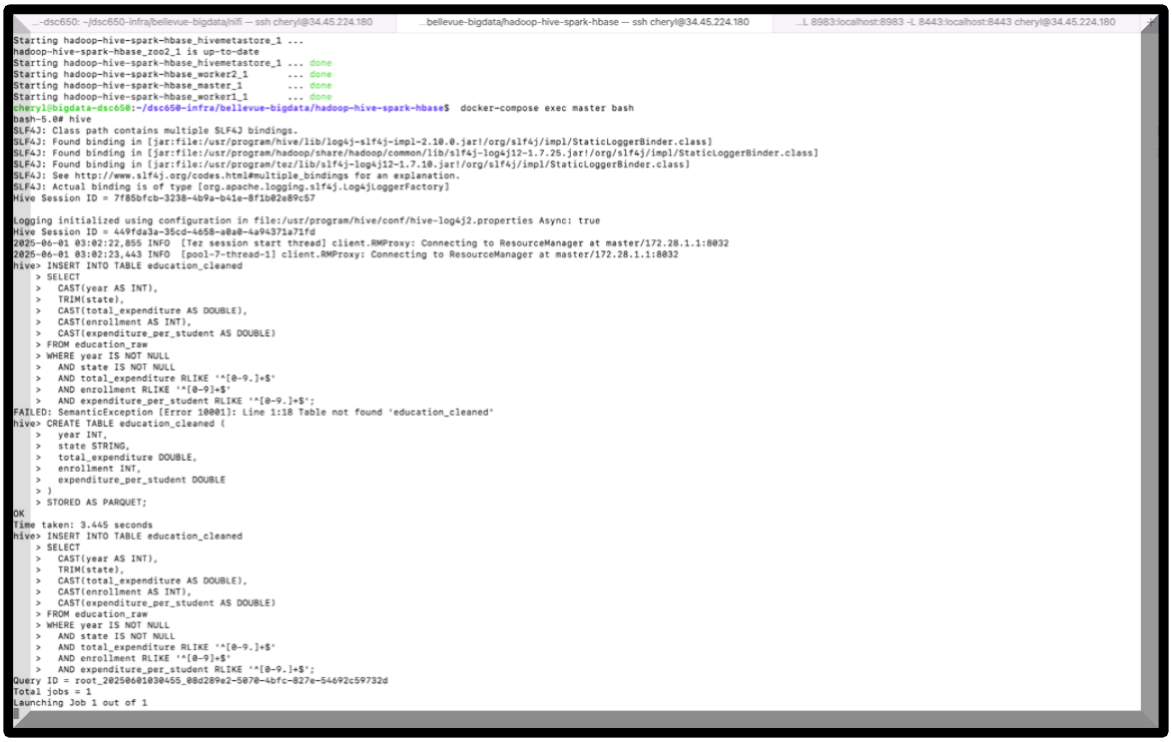
Note: All fields are initially loaded as STRING for cleansing flexibility.

2.3 Verify the Raw Table

Run a basic query to ensure the table is reading data:

SELECT * FROM education_raw LIMIT 5;

You should see rows from states_all.csv printed.



2.4 Create a Cleaned and Typed Table

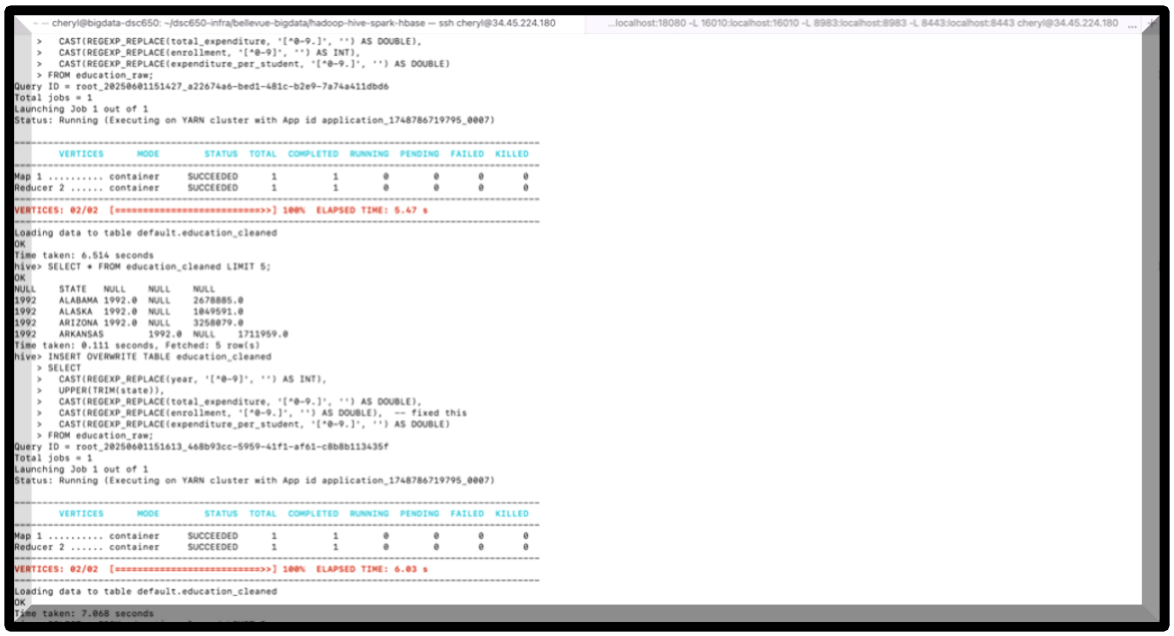
This table stores the cleaned data with correct types in Parquet format:

```
CREATE TABLE education_cleaned (
  year INT,
  state STRING,
  total_expenditure DOUBLE,
  enrollment INT,
  expenditure_per_student DOUBLE
)
STORED AS PARQUET;
```

2.5 Insert Cleaned Data

Transform the raw strings into usable types and filter out malformed entries:

```
INSERT INTO TABLE education_cleaned
SELECT
  CAST(year AS INT),
  TRIM(state),
  CAST(total_expenditure AS DOUBLE),
  CAST(enrollment AS INT),
  CAST(expenditure_per_student AS DOUBLE)
FROM education_raw
WHERE year IS NOT NULL
  AND state IS NOT NULL
  AND total_expenditure RLIKE '^[0-9.]+$'
  AND enrollment RLIKE '^[0-9]+$'
  AND expenditure_per_student RLIKE '^[0-9.]+$';
```



This step may take time as it launches a TEZ job. You can monitor progress via:

http://localhost:8088/cluster

2.6 Confirm Cleaned Data

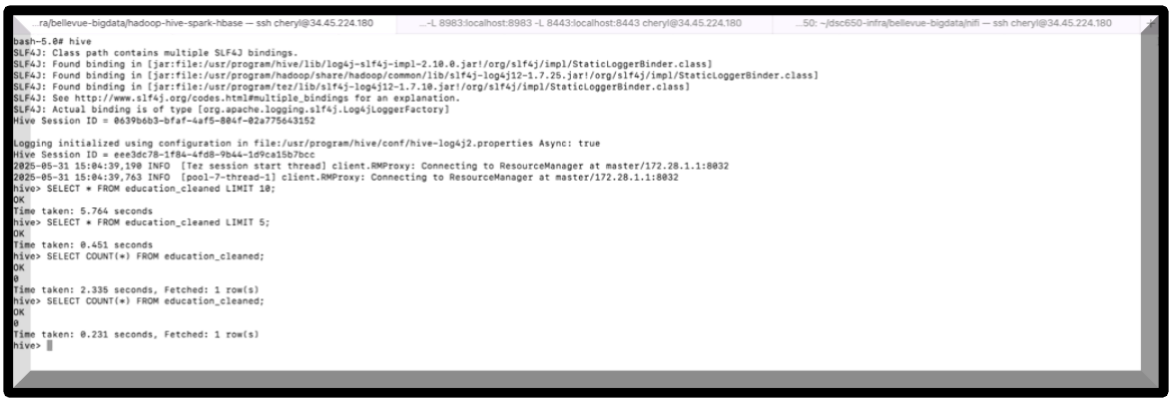
Run:

SELECT * FROM education_cleaned LIMIT 10;

And check how many valid rows were inserted:

SELECT COUNT(*) FROM education_cleaned;

You should see a count > 0 if everything processed correctly.



Step 3: Query & Analyze the Data Using Hive

3.1 Launch Hive Shell

1. SSH into your Hadoop master container :

docker exec -it master bash

2. Start the Hive shell:

hive

3.2 Preview Data in the Cleaned Table

Once inside the Hive CLI:

1. Show existing tables:

SHOW TABLES;

2. Preview the first few rows of your cleaned table:

SELECT * FROM education_cleaned LIMIT 5;

```
Loading data to table default.education_cleaned
OK
Time taken: 6.514 seconds
hive> SELECT * FROM education_cleaned LIMIT 5;
OK
NULL    STATE   NULL    NULL    NULL
1992    ALABAMA 1992.0  NULL    2678885.0
1992    ALASKA  1992.0  NULL    1049591.0
1992    ARIZONA 1992.0  NULL    3258079.0
1992    ARKANSAS        1992.0  NULL    1711959.0
Time taken: 0.111 seconds, Fetched: 5 row(s)
hive>
```

```
hive> SELECT * FROM education_cleaned LIMIT 5;
OK
NULL    PRIMARY_KEY     NULL    NULL    NULL
NULL    1992_ALABAMA    1992.0  NULL    2678885.0
NULL    1992_ALASKA     1992.0  NULL    1049591.0
NULL    1992_ARIZONA    1992.0  NULL    3258079.0
NULL    1992_ARKANSAS   1992.0  NULL    1711959.0
Time taken: 0.141 seconds, Fetched: 5 row(s)
hive> SELECT enrollment, expenditure_per_student FROM education_raw LIMIT 10;
OK
ENROLL  TOTAL_REVENUE
        2678885.0
        1049591.0
        3258079.0
        1711959.0
        26260025.0
        3185173.0
        3834302.0
        645233.0
        789480.0
Time taken: 0.197 seconds, Fetched: 10 row(s)
hive>
```

3. Count total number of rows to validate data:

SELECT COUNT(*) FROM education_cleaned;

3.3 Run Analytical Queries

Now that your table is ready, run queries to analyze trends and gain insights.

Queries:

Average Expenditure Per Student by State

```
SELECT
  state,
  ROUND(AVG(expenditure_per_student), 2) AS avg_expenditure
FROM
  education_cleaned
GROUP BY
  state
ORDER BY
  avg_expenditure DESC;
```

Total Expenditure Over Time (Nationwide)

```
SELECT
  year,
  SUM(total_expenditure) AS total_spending
FROM
  education_cleaned
GROUP BY
  year
ORDER BY
  year;
```

Top 5 States with Highest Spending each year

```
SELECT
  state,
  total_expenditure
```

```
FROM
  education_cleaned
WHERE
  year = 2010
ORDER BY
  total_expenditure DESC
LIMIT 5;
```

3.4 Export Query Results

If you want to save the result to HDFS:

1. Set a Hive configuration for output format:

```
SET hive.cli.print.header=true;
```

---

2. Run and save the result:

```
INSERT OVERWRITE DIRECTORY
'/data/education/output/high_spending_states'
ROW FORMAT DELIMITED
FIELDS TERMINATED BY ','
SELECT
  state,
  total_expenditure
FROM
  education_cleaned
WHERE
  year = 2010
ORDER BY
  total_expenditure DESC
LIMIT 5;
```

---

3. Verify in HDFS:

hdfs dfs -ls /data/education/output/high_spending_states

Step 4: Load Cleaned Data to HBase

4.1 Start HBase Shell

Run this in your Hadoop terminal:

hbase shell

4.2 Create an HBase Table

Run the following command in the HBase shell:

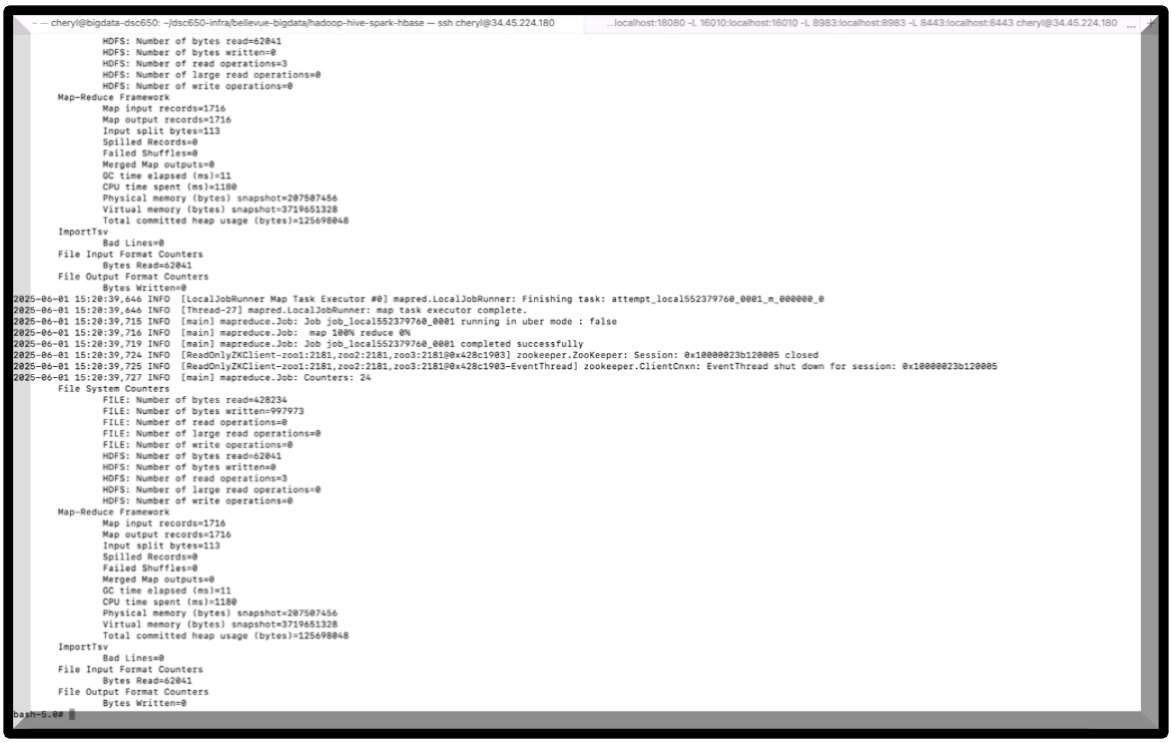
create 'education_data', 'info'

This creates a table named 'education_data' with column family 'info'.

4.3 Load Data into HBase Using ImportTsv

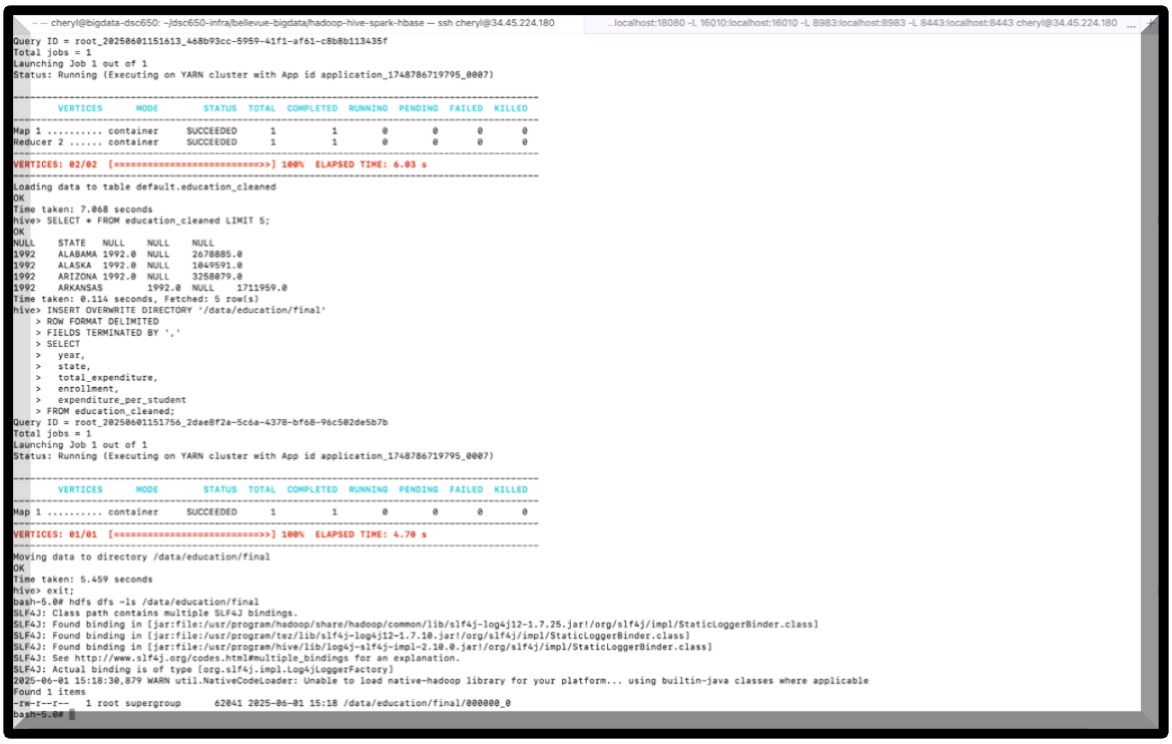
Assuming you have exported a CSV file like cleaned_education.csv, move it to HDFS:

hdfs dfs -put cleaned_education.csv /data/education/final/



Then use the following command to import it into HBase:

```
hbase org.apache.hadoop.hbase.mapreduce.ImportTsv \
-Dimporttsv.separator="," \
-Dimporttsv.columns="HBASE_ROW_KEY,info:year,info:state,info:enrollment,info:total_expenditure,info:expenditure_per_student"\
education_data \
/data/education/final/cleaned_education.csv
```



Replace HBASE_ROW_KEY with a unique field (e.g., concatenate year+state).

4.4 Verify Data in HBase

Back in the HBase shell, run:

```
scan 'education_data'
```

This will display the rows inserted into your HBase table.

```
hbase(main):001:0> scan 'education_data'
ROW                                                   COLUMN+CELL
 1986                                                 column=info:enrollment, timestamp=2025-06-01T15:20:35.831, value=\x5CN
 1986                                                 column=info:state, timestamp=2025-06-01T15:20:35.831, value=1986.0
 1986                                                 column=info:total_expenditure, timestamp=2025-06-01T15:20:35.831, value=\x5CN
 1986                                                 column=info:year, timestamp=2025-06-01T15:20:35.831, value=WYOMING
 1987                                                 column=info:enrollment, timestamp=2025-06-01T15:20:35.831, value=\x5CN
 1987                                                 column=info:state, timestamp=2025-06-01T15:20:35.831, value=1987.0
 1987                                                 column=info:total_expenditure, timestamp=2025-06-01T15:20:35.831, value=\x5CN
 1987                                                 column=info:year, timestamp=2025-06-01T15:20:35.831, value=WYOMING
 1988                                                 column=info:enrollment, timestamp=2025-06-01T15:20:35.831, value=\x5CN
 1988                                                 column=info:state, timestamp=2025-06-01T15:20:35.831, value=1988.0
 1988                                                 column=info:total_expenditure, timestamp=2025-06-01T15:20:35.831, value=\x5CN
 1988                                                 column=info:year, timestamp=2025-06-01T15:20:35.831, value=WYOMING
 1989                                                 column=info:enrollment, timestamp=2025-06-01T15:20:35.831, value=\x5CN
 1989                                                 column=info:state, timestamp=2025-06-01T15:20:35.831, value=1989.0
 1989                                                 column=info:total_expenditure, timestamp=2025-06-01T15:20:35.831, value=\x5CN
 1989                                                 column=info:year, timestamp=2025-06-01T15:20:35.831, value=WYOMING
 1990                                                 column=info:enrollment, timestamp=2025-06-01T15:20:35.831, value=\x5CN
 1990                                                 column=info:state, timestamp=2025-06-01T15:20:35.831, value=1990.0
 1990                                                 column=info:total_expenditure, timestamp=2025-06-01T15:20:35.831, value=\x5CN
 1990                                                 column=info:year, timestamp=2025-06-01T15:20:35.831, value=NATIONAL
 1991                                                 column=info:enrollment, timestamp=2025-06-01T15:20:35.831, value=\x5CN
 1991                                                 column=info:state, timestamp=2025-06-01T15:20:35.831, value=1991.0
 1991                                                 column=info:total_expenditure, timestamp=2025-06-01T15:20:35.831, value=\x5CN
 1991                                                 column=info:year, timestamp=2025-06-01T15:20:35.831, value=WYOMING
 1992                                                 column=info:enrollment, timestamp=2025-06-01T15:20:35.831, value=\x5CN
 1992                                                 column=info:state, timestamp=2025-06-01T15:20:35.831, value=1992.0
 1992                                                 column=info:total_expenditure, timestamp=2025-06-01T15:20:35.831, value=\x5CN
 1992                                                 column=info:year, timestamp=2025-06-01T15:20:35.831, value=NATIONAL
 1993                                                 column=info:enrollment, timestamp=2025-06-01T15:20:35.831, value=615130.0
 1993                                                 column=info:state, timestamp=2025-06-01T15:20:35.831, value=1993.0
 1993                                                 column=info:total_expenditure, timestamp=2025-06-01T15:20:35.831, value=100313
 1993                                                 column=info:year, timestamp=2025-06-01T15:20:35.831, value=WYOMING
 1994                                                 column=info:enrollment, timestamp=2025-06-01T15:20:35.831, value=\x5CN
 1994                                                 column=info:state, timestamp=2025-06-01T15:20:35.831, value=1994.0
 1994                                                 column=info:total_expenditure, timestamp=2025-06-01T15:20:35.831, value=\x5CN
 1994                                                 column=info:year, timestamp=2025-06-01T15:20:35.831, value=NATIONAL
 1995                                                 column=info:enrollment, timestamp=2025-06-01T15:20:35.831, value=634237.0
 1995                                                 column=info:state, timestamp=2025-06-01T15:20:35.831, value=1995.0
 1995                                                 column=info:total_expenditure, timestamp=2025-06-01T15:20:35.831, value=100314
 1995                                                 column=info:year, timestamp=2025-06-01T15:20:35.831, value=WYOMING
 1996                                                 column=info:enrollment, timestamp=2025-06-01T15:20:35.831, value=\x5CN
 1996                                                 column=info:state, timestamp=2025-06-01T15:20:35.831, value=1996.0
 1996                                                 column=info:total_expenditure, timestamp=2025-06-01T15:20:35.831, value=\x5CN
 1996                                                 column=info:year, timestamp=2025-06-01T15:20:35.831, value=NATIONAL
 1997                                                 column=info:enrollment, timestamp=2025-06-01T15:20:35.831, value=660079.0
```

```
~ — cheryl@bigdata-dsc650: ~/dsc650-infra/bellevue-bigdata/hadoop-hive-spark-hbase — ssh cheryl@34.45.224.180   ...localhost:18080 -L 16010:localhost:16010 -L 8983:localhost:8983 -L 8443:localhost:8443 cheryl@34.45.224.180 ...
1996                                                 column=info:total_expenditure, timestamp=2025-06-01T15:20:35.831, value=\x5CN
1996                                                 column=info:year, timestamp=2025-06-01T15:20:35.831, value=NATIONAL
1997                                                 column=info:enrollment, timestamp=2025-06-01T15:20:35.831, value=660079.0
1997                                                 column=info:state, timestamp=2025-06-01T15:20:35.831, value=1997.0
1997                                                 column=info:total_expenditure, timestamp=2025-06-01T15:20:35.831, value=98777
1997                                                 column=info:year, timestamp=2025-06-01T15:20:35.831, value=WYOMING
1998                                                 column=info:enrollment, timestamp=2025-06-01T15:20:35.831, value=\x5CN
1998                                                 column=info:state, timestamp=2025-06-01T15:20:35.831, value=1998.0
1998                                                 column=info:total_expenditure, timestamp=2025-06-01T15:20:35.831, value=\x5CN
1998                                                 column=info:year, timestamp=2025-06-01T15:20:35.831, value=NATIONAL
1999                                                 column=info:enrollment, timestamp=2025-06-01T15:20:35.831, value=780875.0
1999                                                 column=info:state, timestamp=2025-06-01T15:20:35.831, value=1999.0
1999                                                 column=info:total_expenditure, timestamp=2025-06-01T15:20:35.831, value=94420
1999                                                 column=info:year, timestamp=2025-06-01T15:20:35.831, value=WYOMING
2000                                                 column=info:enrollment, timestamp=2025-06-01T15:20:35.831, value=\x5CN
2000                                                 column=info:state, timestamp=2025-06-01T15:20:35.831, value=2000.0
2000                                                 column=info:total_expenditure, timestamp=2025-06-01T15:20:35.831, value=\x5CN
2000                                                 column=info:year, timestamp=2025-06-01T15:20:35.831, value=NATIONAL
2001                                                 column=info:enrollment, timestamp=2025-06-01T15:20:35.831, value=804297.0
2001                                                 column=info:state, timestamp=2025-06-01T15:20:35.831, value=2001.0
2001                                                 column=info:total_expenditure, timestamp=2025-06-01T15:20:35.831, value=89711
2001                                                 column=info:year, timestamp=2025-06-01T15:20:35.831, value=WYOMING
2002                                                 column=info:enrollment, timestamp=2025-06-01T15:20:35.831, value=\x5CN
2002                                                 column=info:state, timestamp=2025-06-01T15:20:35.831, value=2002.0
2002                                                 column=info:total_expenditure, timestamp=2025-06-01T15:20:35.831, value=\x5CN
2002                                                 column=info:year, timestamp=2025-06-01T15:20:35.831, value=NATIONAL
2003                                                 column=info:enrollment, timestamp=2025-06-01T15:20:35.831, value=\x5CN
2003                                                 column=info:state, timestamp=2025-06-01T15:20:35.831, value=2003.0
2003                                                 column=info:total_expenditure, timestamp=2025-06-01T15:20:35.831, value=\x5CN
2003                                                 column=info:year, timestamp=2025-06-01T15:20:35.831, value=NATIONAL
2004                                                 column=info:enrollment, timestamp=2025-06-01T15:20:35.831, value=974888.0
2004                                                 column=info:state, timestamp=2025-06-01T15:20:35.831, value=2004.0
2004                                                 column=info:total_expenditure, timestamp=2025-06-01T15:20:35.831, value=86925
2004                                                 column=info:year, timestamp=2025-06-01T15:20:35.831, value=WYOMING
2005                                                 column=info:enrollment, timestamp=2025-06-01T15:20:35.831, value=\x5CN
2005                                                 column=info:state, timestamp=2025-06-01T15:20:35.831, value=2005.0
2005                                                 column=info:total_expenditure, timestamp=2025-06-01T15:20:35.831, value=\x5CN
2005                                                 column=info:year, timestamp=2025-06-01T15:20:35.831, value=NATIONAL
2006                                                 column=info:enrollment, timestamp=2025-06-01T15:20:35.831, value=1150345.0
2006                                                 column=info:state, timestamp=2025-06-01T15:20:35.831, value=2006.0
2006                                                 column=info:total_expenditure, timestamp=2025-06-01T15:20:35.831, value=86155
2006                                                 column=info:year, timestamp=2025-06-01T15:20:35.831, value=WYOMING
2007                                                 column=info:enrollment, timestamp=2025-06-01T15:20:35.831, value=\x5CN
2007                                                 column=info:state, timestamp=2025-06-01T15:20:35.831, value=2007.0
2007                                                 column=info:total_expenditure, timestamp=2025-06-01T15:20:35.831, value=\x5CN
2007                                                 column=info:year, timestamp=2025-06-01T15:20:35.831, value=NATIONAL
2008                                                 column=info:enrollment, timestamp=2025-06-01T15:20:35.831, value=1602514.0
2008                                                 column=info:state, timestamp=2025-06-01T15:20:35.831, value=2008.0
2008                                                 column=info:total_expenditure, timestamp=2025-06-01T15:20:35.831, value=85991
2008                                                 column=info:year, timestamp=2025-06-01T15:20:35.831, value=WYOMING
2009                                                 column=info:enrollment, timestamp=2025-06-01T15:20:35.831, value=\x5CN
2009                                                 column=info:state, timestamp=2025-06-01T15:20:35.831, value=2009.0
2009                                                 column=info:total_expenditure, timestamp=2025-06-01T15:20:35.831, value=\x5CN
2009                                                 column=info:year, timestamp=2025-06-01T15:20:35.831, value=NATIONAL
2010                                                 column=info:enrollment, timestamp=2025-06-01T15:20:35.831, value=1718605.0
2010                                                 column=info:state, timestamp=2025-06-01T15:20:35.831, value=2010.0
2010                                                 column=info:total_expenditure, timestamp=2025-06-01T15:20:35.831, value=87379
2010                                                 column=info:year, timestamp=2025-06-01T15:20:35.831, value=WYOMING
2011                                                 column=info:enrollment, timestamp=2025-06-01T15:20:35.831, value=\x5CN
2011                                                 column=info:state, timestamp=2025-06-01T15:20:35.831, value=2011.0
2011                                                 column=info:total_expenditure, timestamp=2025-06-01T15:20:35.831, value=\x5CN
```


~ — cheryl@bigdata-dsc650: ~/dsc650-infra/bellevue-bigdata/hadoop-hive-spark-hbase — ssh cheryl@34.45.224.180
...localhost:18080 -L 16010:localhost:16010 -L 8983:localhost:8983 -L 8443:localhost:8443 cheryl@34.45.224.180
05 column=info:total_expenditure, timestamp=2025-06-01T15:20:35.831, value=\x5CN
05 column=info:year, timestamp=2025-06-01T15:20:35.831, value=NATIONAL
06 column=info:enrollment, timestamp=2025-06-01T15:20:35.831, value=1158345.0
06 column=info:state, timestamp=2025-06-01T15:20:35.831, value=2006.0
06 column=info:total_expenditure, timestamp=2025-06-01T15:20:35.831, value=86155
06 column=info:year, timestamp=2025-06-01T15:20:35.831, value=WYOMING
07 column=info:enrollment, timestamp=2025-06-01T15:20:35.831, value=\x5CN
07 column=info:state, timestamp=2025-06-01T15:20:35.831, value=2007.0
07 column=info:total_expenditure, timestamp=2025-06-01T15:20:35.831, value=\x5CN
07 column=info:year, timestamp=2025-06-01T15:20:35.831, value=NATIONAL
08 column=info:enrollment, timestamp=2025-06-01T15:20:35.831, value=1602514.0
08 column=info:state, timestamp=2025-06-01T15:20:35.831, value=2008.0
08 column=info:total_expenditure, timestamp=2025-06-01T15:20:35.831, value=85991
08 column=info:year, timestamp=2025-06-01T15:20:35.831, value=WYOMING
09 column=info:enrollment, timestamp=2025-06-01T15:20:35.831, value=\x5CN
09 column=info:state, timestamp=2025-06-01T15:20:35.831, value=2009.0
09 column=info:total_expenditure, timestamp=2025-06-01T15:20:35.831, value=\x5CN
09 column=info:year, timestamp=2025-06-01T15:20:35.831, value=NATIONAL
10 column=info:enrollment, timestamp=2025-06-01T15:20:35.831, value=1718605.0
10 column=info:state, timestamp=2025-06-01T15:20:35.831, value=2010.0
10 column=info:total_expenditure, timestamp=2025-06-01T15:20:35.831, value=87379
10 column=info:year, timestamp=2025-06-01T15:20:35.831, value=WYOMING
11 column=info:enrollment, timestamp=2025-06-01T15:20:35.831, value=\x5CN
11 column=info:state, timestamp=2025-06-01T15:20:35.831, value=2011.0
11 column=info:total_expenditure, timestamp=2025-06-01T15:20:35.831, value=\x5CN
11 column=info:year, timestamp=2025-06-01T15:20:35.831, value=NATIONAL
12 column=info:enrollment, timestamp=2025-06-01T15:20:35.831, value=1664983.0
12 column=info:state, timestamp=2025-06-01T15:20:35.831, value=2012.0
12 column=info:total_expenditure, timestamp=2025-06-01T15:20:35.831, value=89994
12 column=info:year, timestamp=2025-06-01T15:20:35.831, value=WYOMING
13 column=info:enrollment, timestamp=2025-06-01T15:20:35.831, value=\x5CN
13 column=info:state, timestamp=2025-06-01T15:20:35.831, value=2013.0
13 column=info:total_expenditure, timestamp=2025-06-01T15:20:35.831, value=\x5CN
13 column=info:year, timestamp=2025-06-01T15:20:35.831, value=NATIONAL
14 column=info:enrollment, timestamp=2025-06-01T15:20:35.831, value=1772633.0
14 column=info:state, timestamp=2025-06-01T15:20:35.831, value=2014.0
14 column=info:total_expenditure, timestamp=2025-06-01T15:20:35.831, value=92732
14 column=info:year, timestamp=2025-06-01T15:20:35.831, value=WYOMING
15 column=info:enrollment, timestamp=2025-06-01T15:20:35.831, value=\x5CN
15 column=info:state, timestamp=2025-06-01T15:20:35.831, value=2015.0
15 column=info:total_expenditure, timestamp=2025-06-01T15:20:35.831, value=\x5CN
15 column=info:year, timestamp=2025-06-01T15:20:35.831, value=NATIONAL
16 column=info:enrollment, timestamp=2025-06-01T15:20:35.831, value=2044669.0
16 column=info:state, timestamp=2025-06-01T15:20:35.831, value=2016.0
16 column=info:total_expenditure, timestamp=2025-06-01T15:20:35.831, value=94511
16 column=info:year, timestamp=2025-06-01T15:20:35.831, value=WYOMING
17 column=info:enrollment, timestamp=2025-06-01T15:20:35.831, value=\x5CN
17 column=info:state, timestamp=2025-06-01T15:20:35.831, value=2017.0
17 column=info:total_expenditure, timestamp=2025-06-01T15:20:35.831, value=\x5CN
17 column=info:year, timestamp=2025-06-01T15:20:35.831, value=NATIONAL
19 column=info:enrollment, timestamp=2025-06-01T15:20:35.831, value=\x5CN
19 column=info:state, timestamp=2025-06-01T15:20:35.831, value=2019.0
19 column=info:total_expenditure, timestamp=2025-06-01T15:20:35.831, value=\x5CN
19 column=info:year, timestamp=2025-06-01T15:20:35.831, value=WYOMING
5CN column=info:enrollment, timestamp=2025-06-01T15:20:35.831, value=\x5CN
5CN column=info:state, timestamp=2025-06-01T15:20:35.831, value=\x5CN
5CN column=info:total_expenditure, timestamp=2025-06-01T15:20:35.831, value=\x5CN
5CN column=info:year, timestamp=2025-06-01T15:20:35.831, value=STATE
row(s)
k 0.9592 seconds
ase(main):002:0>