Data Layer Analysis of ClearCampaign An App For Transparent Political Ad Data

A Project Report Submitted to fulfill requirements of Database Systems I, Spring 2024, CMPE180B/138

by Group 3

Ravi Teja Gattu[*] & **Poorna Ravuri**[†] & **Jacob Culaton**[‡]

Supervised By Our Esteemed Professor Dr. Gheorghi Guzun

Submitted on 4-May-2024
Due Date 6-May-2024

[*]Student ID: 017503746, Email: raviteja.gattu@sjsu.edu
[†]Student ID: 017510883, Email: poorna.ravuri@sjsu.edu
[‡]Student ID: 014706172, Email: jacob.culaton@sjsu.edu

Abstract

Dataset from the bigquery-public-data repository: *google_political_ads*.

The project is about an app that allows users to see the political ad campaign data. This involves the spending of political parties that are contesting in elections and have used Google as an advertising platform to campaign. Its vision is to create transparency in the arena of politics for a better world and a better future. The data layer of the app consists of the dataset: *google_political_ads* from bigquery-public-data repository that holds all required data to cater to user personas and add value to them by serving to their motivations.

Important tables within *google_political_ads* dataset are :
creative_stats : This table contains the information for election ads that have appeared on Google Ads Services.
advertiser_stats : The data regarding advertisers that have placed election ads on Google Ads Services with at least one impression is shown in this table.
advertiser_geo_spend : incorporates the total amount spent by US advertisers on political advertisements for each US state as well as the DC,
advertiser_declared_stats : Additional information by marketers in California and New Zealand) etc.,

The project focuses on exploring data and crafting queries that enable the features of the application. The project will include a comparison of existing applications in the market, a comprehensive description of the datasets, an analysis of features provided in the application and queries that enable and optimize them.

User groups involve multiple stakeholders such as voters, contesting political parties, media officials, government employees, judicial system, and law enforcement. Also research enthusiasts who have their own motivations to access political ad campaign data like knowing which party has invested how much into advertising, how much have the other parties set the bar in marketing budget, information to disseminate via media outlets, to check if political parties contesting are adhering to the campaigning rules set, to check if any means of marketing by political parties is objectionable to intervene and stop etc.

Table of Contents

List of Figures

List of Tables

1 Introduction, Motivation / Problem Definition

1.1 Selection Of Dataset and Problem Aimed At Solving

The Dataset *google_political_ads* (1) is selected as it is a unique data which is not usually found on the internet. This information has the potential to be used in multiple ways by many sections of the society. And there is a huge barrier to gain information in the arena of political ad spending as it is the neo-age means to reach and to market to the new generation people. Any field that has information barrier can be disrupted with the introduction of enablers that help key stakeholders of that field. In this case, it is the political ad-spending data that is crucial.

1.2 Schema Description

Dataset : google_political_ads
Description : This dataset contains details about the expenditure of verified advertisers on political advertising through Google Ad Services. It includes insights into demographic targeting strategies employed in political ad campaigns by these advertisers. Additionally, it provides links to the actual political ads in the Google Transparency Report (`https://adstransparency.google.com`). Data related to an election is retained for 7 years post the election, after which it is removed from the dataset and becomes inaccessible.(1)
List of Tables:
bigquery-public-data.google_political_ads.advertiser_declared_stats: Advertiser declared stats for Google Political Ads.

bigquery-public-data.google_political_ads.advertiser_geo_spend: Geographical spending data for Google Political Ads by advertisers.

bigquery-public-data.google_political_ads.advertiser_stats: Statistics related to advertisers' activities in Google Political Ads.

bigquery-public-data.google_political_ads.advertiser_weekly_spend: Weekly spending data for Google Political Ads by advertisers.

bigquery-public-data.google_political_ads.campaign_targeting: Campaign targeting information for Google Political Ads.

bigquery-public-data.google_political_ads.creative_stats: Statistics on the creativity of Google Political Ads.

bigquery-public-data.google_political_ads.geo_spend: Geographical spending data for Google Political Ads.

1.3 ER Diagram

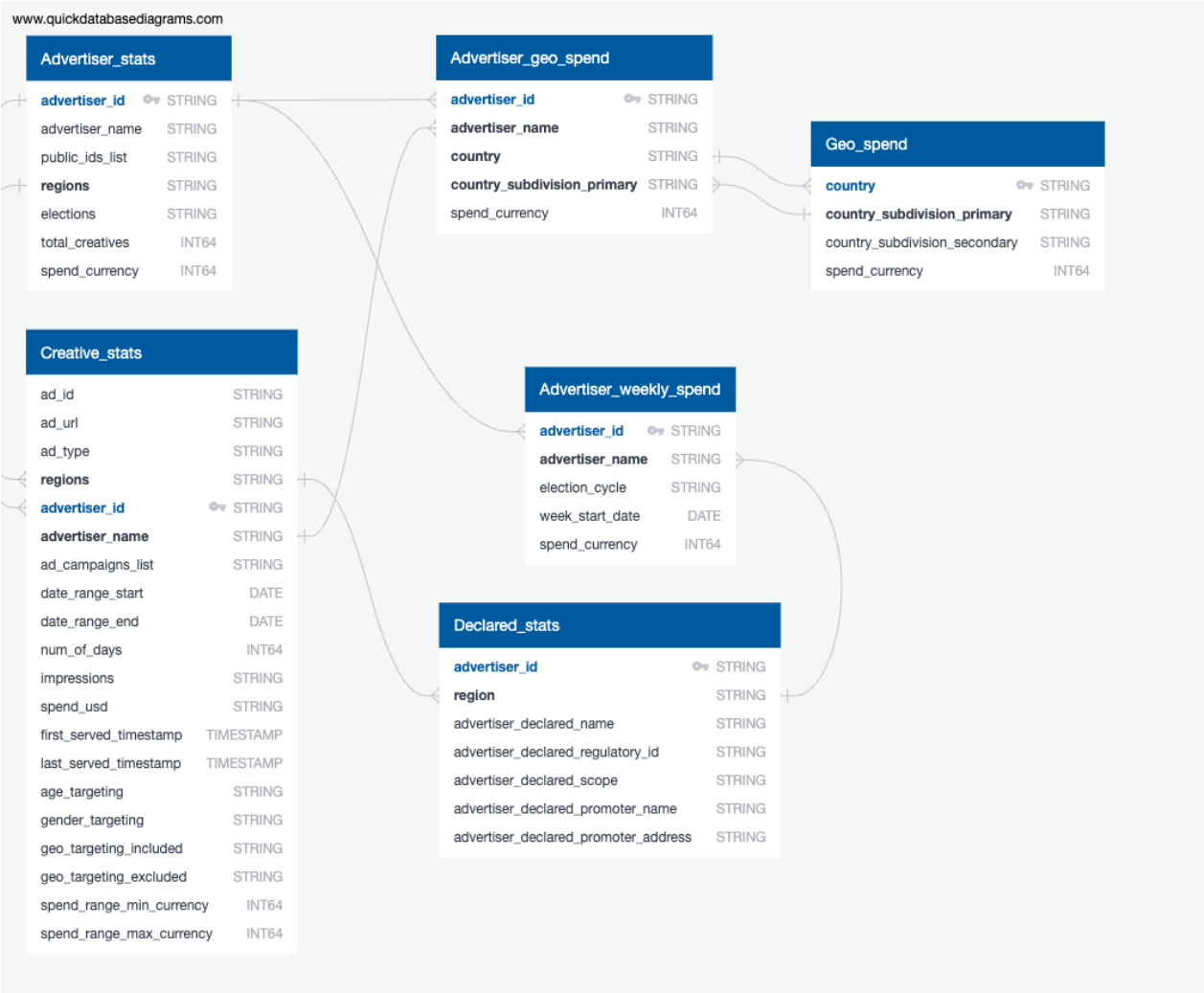

Figure 1: ER Diagram of Dataset: google_political_ads

1.4 Application Designed To Leverage The Dataset Selected

Application : **ClearCampaign** is the application designed to leverage dataset selected (*google_political_ads*). This ensures to bridge the gap between multiple sections of society and the information of political advertisement spending data. This removes the barriers to accessing information with regard to the revolutionary means to campaign for political impact.

1.5 Reasons For Selecting And Why This Is Interesting

- To introduce transparency in political ad spending.
- To enable access political ad spending data.
- To enable access of information of political ad campaign data like knowing which party has invested how much into advertising.
- To understand how much have the political parties set the bar in marketing budget.
- For enabling a monitor, if political parties contesting are adhering to the campaigning rules set
- For monitoring, if any means/groups targeted of/by marketing by political parties is objectionable to intervene and stop etc.
- To collect information and disseminate via media outlets

For all the above reasons, the dataset was selected to be leveraged via app with mentioned potential serving vision.
This is also interesting as it would enable groundbreaking information disruption in aspects for transparency in political ad campaign spending.
It is also interesting due to the fact that this information was never that readily available for the society at the disposal of their hands.
With advent of this app, ClearCampaign, information will be available at a citizen's fingertips, considered the data usage will be responsible.

1.6 Application's Users

- Vote Casting Citizens
- Political Stakeholders
- Researchers
- Media
- National Election Commissions / Ombudsmen
- Law Enforcement
- Judicial Entities

1.7 Application's Usage by User Groups

- Citizens access political ad spending data and help them making better decisions.

- Research enthusiasts access information, who have their own motivations to access political ad campaign data like knowing which party has invested how much into advertising.
- Political stakeholders use application to understand how much have the other parties set the bar in marketing budget, information to disseminate via media outlets
- Political, Law Enforcing Ombudsmen can check if political parties contesting are adhering to the campaigning rules set
- Judicial / Law Enforcement can monitor if any means/groups targetted of/by marketing by political parties is objectionable to intervene and stop etc.
- Media entities can collect information and disseminate via media outlets

1.8 Why Is It Challenging

- Impact this data carries is enormous, ie., crucial in making decisions that drive who leads a country's politics and its impact in growth of economies.
- Transparency in such scenarios is of utmost requirement.
- Advertisement providing entity becomes the neutral, natural, and an unbiased source of information.
- It also acts as one stop shop to access mutliple competing political entities.

1.9 Shortcomings And Limitations Of Existing Work

- Despite there are applications existing to serve the same purpose as ClearCampaign, the existing solutions are not as well designed as ClearCampaign.
- They are not simplified to the level of user experience we have planned which is multi-leveled, user-empathetic and user-friendly unlike existing solutions that are rigid from usage perspective. Example : ClearCampaign has multiple options for users to choose from, so that they have choice of wide range of filters on data they can narrow down to. Existing solutions do not possess the same when compared to that of ClearCampaign.
- The existing solutions either need the user to be verified by the application platform or need intense sign up procedures to access data. This acts as entry barrier to users trying to access data required, as sign-up involved also has users to produce documentation.
- On the other hand few solutions provide data similar to that of ClearCampaign.
- Some solutions provide data in form of document/report unlike ClearCampaign.

2 Related Work

2.1 Relating ClearCampaign With Other Solutions

Other Existing Solutions (2) :

- AdImpact (3)
- Facebook Open Research and Transparency (FORT) Platform (4)
- Bloomberg Government (5)
- OpenSecrets (6)
- Roku Ad Platform (7)

2.2 Solutions Comparison with ClearCampaign

Despite there are applications existing to serve the same purpose as ClearCampaign, the existing solutions are not as well designed as ClearCampaign. They are not simplified to the level of app we have planned. They either need the user to be verified by the application platform or need intense sign up procedures to access data. On the other hand few solutions provide data in similar manner. Some solutions provide data in form of document/report unlike ClearCampaign.

2.3 Citations

2.3.1 List of Citations Referred, Brief Explanation

- AdImpact: `https://adimpact.com/solution/tailored-insights/`
- Facebook Open Research and Transparency (FORT) Platform: `https://www.opensecrets.org/campaign-expenditures/vendors`
- Bloomberg Government: `https://about.bgov.com/request-a-demo-government-affairs/?bbgsum=AD-PA-GGL-Nbrand-H124880-bgov&utm_medium=paidsearch&utm_source=google&gad_source=1&gclid=Cj0KCQjwir2xBhC_ARIsAMTXk87kr-txAxB6Q3VOsz6IcWdBAmNcWHZaWPo-7d8_mvnaWUL2xdWZONsaArWDEALw_wcB&gclsrc=aw.ds`
- OpenSecrets: `https://www.opensecrets.org/campaign-expenditures/vendors`
- Roku Ad Platform: `https://advertising.roku.com/learn/resources/voters-are-streaming-how-2024-political-campaigns-are-adapting-to-a-changed-tv-landscape`

2.3.2 Pros - Cons, Comparison and Improvement by ClearCampaign

- AdImpact

  - Pros : Presence of comprehensive data.
  - Cons : No dedicated UI for easy access.
  - Comparison with ClearCampaign : AdImpact only provides on-demand reports and database.
  - Improvements in ClearCampaign : ClearCampaign simplifies access to data by providing filters on multiple levels.
- Facebook Open Research and Transparency (FORT) Platform
  - Pros : Political ad campaign targetting data is also present for us to analyze its impact.
  - Cons : No direct access to data, instead API is provided.
  - Comparison with ClearCampaign : On par with ClearCampaign with filters.
  - Improvements in ClearCampaign : UI offering while FORT is an API.
- Bloomberg Government
  - Pros : Application provides customer service.
  - Cons : Not a one stop shop as this application might not host complete ad data.
  - Comparison with ClearCampaign : BG is a paid service.
  - Improvements in ClearCampaign : Free service with wide range of data.
- OpenSecrets
  - Pros : Less effort in accessing data,
  - Cons : Uses the same dataset on top of which ClearCampaign is built.
  - Comparison with ClearCampaign : Filtering is easily accessible in both.
  - Improvements in ClearCampaign : Advanced filters and UI.
- Roku Ad Platform
  - Pros : Up-to-date data provided on the platform.
  - Cons : Limited access to public users.
  - Comparison with ClearCampaign : Lack of target customer data like demographics.
  - Improvements in ClearCampaign : ClearCampaign UX is simple unlike Roku's, where data is provided in form of excel sheets.

3 Methods Description

3.0.1 Features Provided, Reasons And Queries Powering The Feature

Feature 1 : Top 10 All Time Political Ad Campaigns in USA by Spending

- Reason : To access political spending statistics / historical data by user groups of researchers, media, political entities etc and process information for their motivations
- Query (Before Optimization) :

```
SELECT
  a.advertiser_id,
  a.advertiser_name,
  SUM(g.spend_usd) AS total_spend_usd
FROM
  `bigquery-public-data.google_political_ads.
  advertiser_stats` AS a
JOIN
  `bigquery-public-data.google_political_ads
  .advertiser_geo_spend` AS g
ON
  a.advertiser_id = g.advertiser_id
WHERE
  g.country = 'US'
GROUP BY
  a.advertiser_id,
  a.advertiser_name
ORDER BY
  total_spend_usd DESC
LIMIT
  10;
```

- Github Query Link : https://github.com/RaviTejaGattu1/ClearCampaign_AppDataLayer/blob/98d49e41d183a09d47873d6851862cd5c65d522e/AllTime_Top10_USAPoliticalAdCampaigns_bySpending/AllTime_Top10_USAPoliticalAdCampaigns_bySpending.sql

Feature 2 : California Weekly Political Ad Spends in 2024

- Reason : To access weekly political ad spending in 2024 localised to California geography. Users can select other region based on their requirement.Trends of investment in political advertising spending and elections timeline can be observed in the output.
- Query :

```
SELECT
```

```
  g.country_subdivision_primary,
  w.week_start_date,
  w.advertiser_name,
  SUM(w.spend_usd) AS total_weekly_spend_usd
FROM
  `bigquery-public-
  data.google_political_ads.advertiser_weekly_spend` AS w
JOIN
  `bigquery-public-data.
  google_political_ads.advertiser_geo_spend` AS g
ON
  w.advertiser_id = g.advertiser_id
WHERE
  g.country = 'US'AND g.country_subdivision_primary='CA'
GROUP BY
  g.country_subdivision_primary,
  w.week_start_date,
  w.advertiser_name
ORDER BY
  g.country_subdivision_primary,
  w.week_start_date DESC
LIMIT
  100;
```

- Github Query Link : https://github.com/RaviTejaGattu1/ClearCampaign_AppDataLayer/blob/b3d36dabafe1bc2e0bab1a2e3410970283e8fb24/%5B2-IMP%5DCalifornia_2024_Weekly_PoliticalAdSpends/California_2024_Weekly_PoliticalAdSpends.sql

Feature 3 : Demographics Targeted and Impact by Top10 US Political Ad Spends 2024

- Reason : To find data of top 10 political ad spends, the demographics targeted by the ads and impact in terms of ad-impressions in 2024.
- Query :

```
SELECT
  a.regions,
  EXTRACT(YEAR
  FROM
    c.last_served_timestamp) AS year,
  a.advertiser_name,
  SUM(c.spend_range_max_usd) AS Spending,
  c.impressions,
  c.ad_type,
```

```
  c.age_targeting,
  c.gender_targeting,
  c.geo_targeting_included
FROM
  `bigquery-public-data
  .google_political_ads.creative_stats` AS c
JOIN
  `bigquery-public-data
  .google_political_ads.advertiser_stats` AS a
ON
  c.advertiser_id = a.advertiser_id
WHERE
  a.regions='US'
  AND a.spend_usd<>0
  AND c.age_targeting IS NOT NULL
  AND c.age_targeting IS NOT NULL
  AND EXTRACT(YEAR
  FROM
    date_range_start)=2024
GROUP BY
  a.regions,
  year,
  a.advertiser_name,
  c.ad_type,
  c.impressions,
  c.age_targeting,
  c.gender_targeting,
  c.geo_targeting_included
ORDER BY
  Spending DESC
LIMIT
  100;
```

- Github Query Link : https://github.com/RaviTejaGattu1/ClearCampaign_AppDataLayer/blob/625105795fab5e9f8109328f252ef950656a6bdd/%5B3-IMP%5DDemographics_Targetted_and_Impact_by_Top10_USPoliticalAdSpends_2024/Demographics_Targetted_and_Impact_by_Top10_USPoliticalAdSpends_2024.sql

Feature 4 : Total Political Ad Spending by Advertisers, US State Wise

- Reason : For users to access and assess political ad spending entities for the states in US.
- Query :

```sql
WITH
  FilteredAdvertiserStats AS (
  SELECT
    advertiser_id,
    advertiser_name
  FROM
    `bigquery-public-data
    .google_political_ads.advertiser_stats` ),
  FilteredGeoSpend AS (
  SELECT
    advertiser_id,
    country_subdivision_primary,
    spend_usd
  FROM
    `bigquery-public-data
    .google_political_ads.advertiser_geo_spend`
  WHERE
    country = 'US' )
SELECT
  a.advertiser_id,
  a.advertiser_name,
  g.country_subdivision_primary,
  SUM(g.spend_usd) AS total_spend_usd
FROM
  FilteredAdvertiserStats a
JOIN
  FilteredGeoSpend g
ON
  a.advertiser_id = g.advertiser_id
GROUP BY
  1,
  2,
  3
LIMIT
  100;
```

- Github Query Link : `https://github.com/RaviTejaGattu1/ClearCampaign_AppDataLayer/blob/7b70b1c8878aaf47341a6f37e1c553d0eaa4fd91/%5B4-IMP%5DUS_State_wise_Total_Spending_by_Advertiser/us_state_wise_total_spending_by_advertiser.sql`

Feature 5 : Country Wise Total Political Ad Spending

- Reason : To observe the economical aspect of political advertising campaigns worldwide.

- Query :

```
(
  SELECT
    country,
    SUM(spend_usd) AS Political_Ad_Spending
  FROM
    `bigquery-public-data
    .google_political_ads.advertiser_geo_spend`
  WHERE
    country='US'
  GROUP BY
    country
)
UNION ALL
(
  SELECT
    country,
    SUM(spend_gbp) AS Political_Ad_Spending
  FROM
    `bigquery-public-data
    .google_political_ads.advertiser_geo_spend`
  WHERE
    country='GB'
  GROUP BY
    country
)
UNION ALL
(
SELECT
  country,SUM(spend_ils) AS Political_Ad_Spending
FROM
  `bigquery-public-data
  .google_political_ads.advertiser_geo_spend`
WHERE
    country='IL'
GROUP BY
  country
LIMIT
  100
)
```

- Github Query Link : https://github.com/RaviTejaGattu1/ClearCampaign_AppDataLayer/blob/5959f5483b9a33c432aed92068c1c791ed8fb1cb/%5B5%5DCountry_

Wise_Political_Ad_Spending/country_wise_political_ad_spending.sql

Feature 6 : Weekly Total Political Ad Spending and Advertisers for 2024

- Reason : To access political ad spending by advertisers worldwide during 2024.
- Query :

```
SELECT
  EXTRACT(YEAR FROM week_start_date) as YEAR,
  week_start_date as WEEK,
  COUNT(DISTINCT advertiser_id) as No_of_Advertisers,
  SUM(spend_usd) as Total_Political_Ad_Spending
FROM
  `bigquery-public-data
  .google_political_ads.advertiser_weekly_spend`
WHERE
  EXTRACT(YEAR FROM week_start_date)=2024
GROUP BY
YEAR,
week_start_date
ORDER BY
YEAR, week_start_date
LIMIT
  50;
```

- Github Query Link : https://github.com/RaviTejaGattu1/ClearCampaign_AppDataLayer/blob/5959f5483b9a33c432aed92068c1c791ed8fb1cb/%5B6%5DWeekly_Total_Political_Ad_Spending_and_Advertisers_for_2024/Weekly_Total_Political_Ad_Spending_and_Advertisers_for_2024.sql

Feature 7 : Political Ad Spending by Genders Targetted in US

- Reason : To gain understanding of political investment structured to target genders of society.
- Query :

```
SELECT
  cs.gender_targeting, COUNT(DISTINCT cs.ad_id)
  AS gender_count,
  SUM(ags.spend_usd)
  AS total_usd
FROM
  bigquery-public-data
  .google_political_ads.creative_stats AS cs
JOIN
```

```
  bigquery-public-data
  .google_political_ads.advertiser_geo_spend AS ags
ON
  cs.advertiser_id = ags.advertiser_id
WHERE
  ags.country = 'US'
GROUP BY
  cs.gender_targeting
LIMIT
  100;
```

- Github Query Link : https://github.com/RaviTejaGattu1/ClearCampaign_AppDataLayer/blob/5959f5483b9a33c432aed92068c1c791ed8fb1cb/%5B7%5DUS_Ad_Spending_By_Gender_Targetted/us_ad_spending_by_gender.sql

Feature 8 : Advertising Entities vs Total Spending By Them

- Reason : To find the advertisers behind political campaigns and the cumulative amount spent by them.
- Query :

```
SELECT
  ast.advertiser_name,
  ast.*,
  ads.*
FROM
  `bigquery-public-data
  .google_political_ads.advertiser_stats` AS ast
INNER JOIN
  `bigquery-public-data.google_political_ads
  .advertiser_declared_stats` AS ads
ON
  ads.advertiser_id = ast.advertiser_id
WHERE
  regions IS NOT NULL
LIMIT
  50;
```

- Github Query Link : https://github.com/RaviTejaGattu1/ClearCampaign_AppDataLayer/blob/5959f5483b9a33c432aed92068c1c791ed8fb1cb/%5B8%5DAdvertiser_and_Spending/advertiser_and_spending.sql

Feature 9 : Ad Impressions

- Reason : To check types of ads and find categories, their impact in terms of ad-impressions.

- Query :

```
SELECT ad_id, ad_type, impressions
FROM bigquery-public-data
.google_political_ads.creative_stats where impressions =
    (select max(impressions) from bigquery-public-data
    .google_political_ads.creative_stats);
```

- Github Query Link : https://github.com/RaviTejaGattu1/ClearCampaign_AppDataLayer/blob/66c6bd07d2c0c44f439676fa72cc081d7ca76493/%5B91%5DImpressions_of_Ads/Impression_of_Ads.sql

Feature 10 : State Wise and Advertisement Level Spends in USA

- Reason : To get a picture of state level and ad-wise spending in USA
- Query :

```
SELECT
  gs.country_subdivision_primary AS state,
  ad.advertiser_name,
  SUM(gs.spend_usd) AS total_spend_usd
FROM
  `bigquery-public-data
  .google_political_ads.advertiser_geo_spend` gs
JOIN
  `bigquery-public-data
  .google_political_ads.advertiser_stats` ad
ON
  gs.advertiser_id = ad.advertiser_id
WHERE
  gs.country = 'US'
GROUP BY
  state,
  ad.advertiser_name
ORDER BY
  total_spend_usd DESC
```

- Github Query Link : https://github.com/RaviTejaGattu1/ClearCampaign_AppDataLayer/blob/66c6bd07d2c0c44f439676fa72cc081d7ca76493/%5B92%5DUS_State_Ad_Wise_Spending/US_State_Ad_Wise_Spending.sql

Feature 11 : State Wise Political Ad Spends in USA

- Reason : To get a picture of state level political ad spending in USA
- Query :

```
SELECT
  'US' AS country,
  country_subdivision_primary AS state,
  SUM(spend_usd) AS total_usd
FROM
  bigquery-public-data
  .google_political_ads.advertiser_geo_spend
WHERE
  country='US'
GROUP BY
  country_subdivision_primary
HAVING
  total_usd > 0
ORDER BY
  total_usd DESC;
```

- Github Query Link : https://github.com/RaviTejaGattu1/ClearCampaign_AppDataLayer/blob/66c6bd07d2c0c44f439676fa72cc081d7ca76493/%5B93%5DUS_State_wise_Ad_Spending/us_state_wise_ad_spending.sql

3.0.2 Important Query's Optimization,Algorithm,Functions/Features

Important Query 1 : Top 10 All Time Political Ad Campaigns in USA by Spending

- Query (After Optimization) :

```
WITH FilteredCreativeStats AS (
  SELECT
    advertiser_id,
    EXTRACT(YEAR FROM last_served_timestamp) AS year,
    spend_range_max_usd,
    impressions,
    ad_type,
    age_targeting,
    gender_targeting,
    geo_targeting_included,
    EXTRACT(YEAR FROM date_range_start) AS start_year
  FROM
    `bigquery-public-data
    .google_political_ads.creative_stats`
  WHERE
    age_targeting IS NOT NULL AND
    EXTRACT(YEAR FROM date_range_start) = 2024
),
```

```
FilteredAdvertiserStats AS (
  SELECT
    advertiser_id,
    regions,
    advertiser_name,
    spend_usd
  FROM
    `bigquery-public-data
    .google_political_ads.advertiser_stats`
  WHERE
    regions = 'US' AND
    spend_usd <> 0
)

SELECT
  a.regions,
  c.year,
  a.advertiser_name,
  SUM(c.spend_range_max_usd) AS Spending,
  c.impressions,
  c.ad_type,
  c.age_targeting,
  c.gender_targeting,
  c.geo_targeting_included
FROM
  FilteredCreativeStats c
JOIN
  FilteredAdvertiserStats a
  ON c.advertiser_id = a.advertiser_id
GROUP BY
  a.regions,
  c.year,
  a.advertiser_name,
  c.ad_type,
  c.impressions,
  c.age_targeting,
  c.gender_targeting,
  c.geo_targeting_included
ORDER BY
  Spending DESC
LIMIT
  100;
```

Algorithm, Functions/Features Used and Reasons :

- The WITH clause enhances query optimization by organizing code logically and potentially materializing intermediate results, improving readability and performance. It aids in understanding complex SQL queries, facilitating debugging and testing, ultimately optimizing execution efficiency and maintainability
- Filtering: The query selectively filters data, excluding rows where age targeting is NULL and focusing solely on the year 2024.
- Joining: It efficiently merges data from the creative_stats and advertiser_stats tables based on the advertiser_id column, likely benefiting from indexed columns for faster joining.
- Aggregation: Utilizing the SUM() function, the query aggregates spending data, grouping results by various dimensions such as advertiser, year, ad type, impressions, age targeting, gender targeting, and geo targeting to summarize the data effectively.
- Ordering and Limiting: The query sorts the results by spending in descending order and limits the output to the top 100 rows, aiding in quickly identifying the highest spending advertisers.
- EXTRACT(): Extracts components like the year from timestamps.
- SUM(): Aggregates spending amounts.
- JOIN: Merges tables based on the advertiser_id.
- GROUP BY: Groups results by specified dimensions.ORDER BY: Sorts results based on spending.
- LIMIT: Restricts the number of returned rows in the result set.

Important Query 2 : California Weekly Political Ad Spends in 2024

- Query After Optimization :

```
WITH FilteredGeoSpend AS (
  SELECT
    advertiser_id,
    country_subdivision_primary
  FROM
    `bigquery-public-data
    .google_political_ads.advertiser_geo_spend`
  WHERE
    country = 'US' AND
    country_subdivision_primary = 'CA'
)
SELECT
```

```
13    g.country_subdivision_primary,
14    w.week_start_date,
15    w.advertiser_name,
16    SUM(w.spend_usd) AS total_weekly_spend_usd
17  FROM
18    `bigquery-public-data.google_political_ads
19    .advertiser_weekly_spend` AS w
20  JOIN
21    FilteredGeoSpend AS g
22  ON
23    w.advertiser_id = g.advertiser_id
24  GROUP BY
25    1, 2, 3
26  ORDER BY
27    1, 2 DESC
28  LIMIT
29    100;
```

Algorithm, Functions/Features Used and Reasons :

- Algorithm involves first filtering data from the advertiser_geo_spend table using a WITH clause to create a temporary table FilteredGeoSpend, which includes only relevant data where the country is 'US' and the primary subdivision is 'CA'. Then, it performs a join operation between the advertiser_weekly_spend table and the FilteredGeoSpend subquery based on the advertiser_id column. Following the join, the data is grouped by country subdivision, week start date, and advertiser name, and spending amounts are aggregated. The results are then sorted by country subdivision and week start date in descending order and limited to the top 100 rows item SELECT: Retrieves specific columns from tables and subqueries.
- SUM(): Aggregates spending amounts.
- JOIN: Merges tables based on the advertiser_id column.
- GROUP BY: Groups results by specified dimensions.
- ORDER BY: Sorts results based on specified columns.
- LIMIT: Restricts the number of returned rows in the result set.
- WITH: Defines a temporary table (FilteredGeoSpend) that filters data before joining, optimizing the query by reducing the amount of data processed during the join operation.
- Filtering: The WITH clause filters data from the advertiser_geo_spend table, creating a temporary table (FilteredGeoSpend) with only relevant data, which improves query performance by reducing the dataset size before the join operation.

- Joining: It joins the advertiser_weekly_spend table with the filtered data from FilteredGeoSpend based on the advertiser_id column.
- Aggregation: The query aggregates spending data by country subdivision, week start date, and advertiser name using the SUM() function.
- Ordering: Results are ordered by country subdivision and week start date in descending order to provide meaningful insights.
- Limiting: Finally, the query limits the output to the top 100 rows for easier analysis and presentation of results.

Important Query 3 : Demographics Targeted and Impact by Top10 US Political Ad Spends 2024

- Query After Optimization:

```
WITH FilteredCreativeStats AS (
  SELECT
    advertiser_id,
    EXTRACT(YEAR FROM last_served_timestamp) AS year,
    spend_range_max_usd,
    impressions,
    ad_type,
    age_targeting,
    gender_targeting,
    geo_targeting_included,
    EXTRACT(YEAR FROM date_range_start) AS start_year
  FROM
    `bigquery-public-data
    .google_political_ads.creative_stats`
  WHERE
    age_targeting IS NOT NULL AND
    EXTRACT(YEAR FROM date_range_start) = 2024
),
FilteredAdvertiserStats AS (
  SELECT
    advertiser_id,
    regions,
    advertiser_name,
    spend_usd
  FROM
    `bigquery-public-data
    .google_political_ads.advertiser_stats`
  WHERE
    regions = 'US' AND
```

```
    spend_usd <> 0
)

SELECT
  a.regions,
  c.year,
  a.advertiser_name,
  SUM(c.spend_range_max_usd) AS Spending,
  c.impressions,
  c.ad_type,
  c.age_targeting,
  c.gender_targeting,
  c.geo_targeting_included
FROM
  FilteredCreativeStats c
JOIN
  FilteredAdvertiserStats a
  ON
  c.advertiser_id = a.advertiser_id
GROUP BY
  a.regions,
  c.year,
  a.advertiser_name,
  c.ad_type,
  c.impressions,
  c.age_targeting,
  c.gender_targeting,
  c.geo_targeting_included
ORDER BY
  Spending DESC
LIMIT
  100;
```

Algorithm, Functions/Features Used and Reasons :

- The algorithm involves filtering data from the creative_stats table to create a temporary table FilteredCreativeStats containing only relevant data for the year 2024 and where age targeting is not NULL. Then, it filters data from the advertiser_stats table to create a temporary table FilteredAdvertiserStats, including only data where regions are 'US' and spend_usd is not 0. Afterward, it performs an inner join between FilteredCreativeStats and FilteredAdvertiserStats on the advertiser_id column. The resulting dataset is grouped by various dimensions including regions, year, advertiser name, ad type,

impressions, age targeting, gender targeting, and geo targeting. Finally, the spending amounts are aggregated using the SUM() function, and the results are ordered by spending in descending order and limited to the top 100 rows.

- WITH: Defines temporary tables (FilteredCreativeStats and FilteredAdvertiserStats) that filter data before joining, optimizing the query by reducing the amount of data processed during the join operation.
- EXTRACT(): Extracts components such as the year from timestamps.
- SUM(): Aggregates spending amounts.
- JOIN: Merges tables based on the advertiser_id column.
- GROUP BY: Groups results by specified dimensions.
- ORDER BY: Sorts results based on specified columns.
- LIMIT: Restricts the number of returned rows in the result set.

Important Query 4 : Total Political Ad Spending by Advertisers, US State Wise

- Query After Optimization:

```
SELECT
  t1.advertiser_name,
  t2.country_subdivision_primary,
  SUM(t2.spend_usd) AS total_spending
FROM
  `bigquery-public-data
  .google_political_ads.advertiser_stats` AS t1
INNER JOIN
  `bigquery-public-
  data.google_political_ads.advertiser_geo_spend` AS t2
ON
  t1.advertiser_id = t2.advertiser_id
WHERE
  t2.country = 'US'
GROUP BY
  1,
  2
LIMIT 100;
```

Algorithm, Functions/Features Used and Reasons :

- The algorithm involves joining the advertiser_stats table (t1) with the advertiser_geo_spend table (t2) on the advertiser_id column. It then filters the joined data to include only records where the country is 'US'. The resulting dataset is grouped by advertiser name (t1.advertiser_name) and country subdivision primary

(t2.country_subdivision_primary). Finally, spending amounts are aggregated using the SUM() function, and the results are limited to the top 100 rows.

- Functions/Features Used:
- SELECT: Retrieves specific columns from tables.
- SUM(): Aggregates spending amounts.
- INNER JOIN: Merges tables based on the advertiser_id column.
- WHERE: Filters rows based on specified conditions.
- GROUP BY: Groups results by specified dimensions.
- LIMIT: Restricts the number of returned rows in the result set.

3.0.3 Data Structures Used - Reasons

1.Temporary Tables (Common Table Expressions - CTEs):

- CTEs, defined using the WITH clause, are utilized in both queries to filter and preprocess data before performing join operations.
- They help in organizing and modularizing the query,
  improving readability and maintainability.
- CTEs reduce redundancy by computing and storing intermediate results, optimizing the overall query execution.

2.Hash-Based Data Structures (for Join Operations):

- Hash-based data structures are employed during join operations to efficiently match rows from different tables based on a common key (e.g., advertiser_id).
- By hashing the join columns, the database system can quickly locate matching rows, reducing the time complexity of the join operation.
- Hash joins are particularly efficient for joining large datasets, making them suitable for optimizing queries that involve joining multiple tables, as seen in both queries provided.

3.0.4 Diagrams And Drawings

Query Optimization Results
Comparison Of IO Cost Before vs After Optimization

- Important Query 1 : Top 10 All Time Political Ad Campaigns in USA by Spending

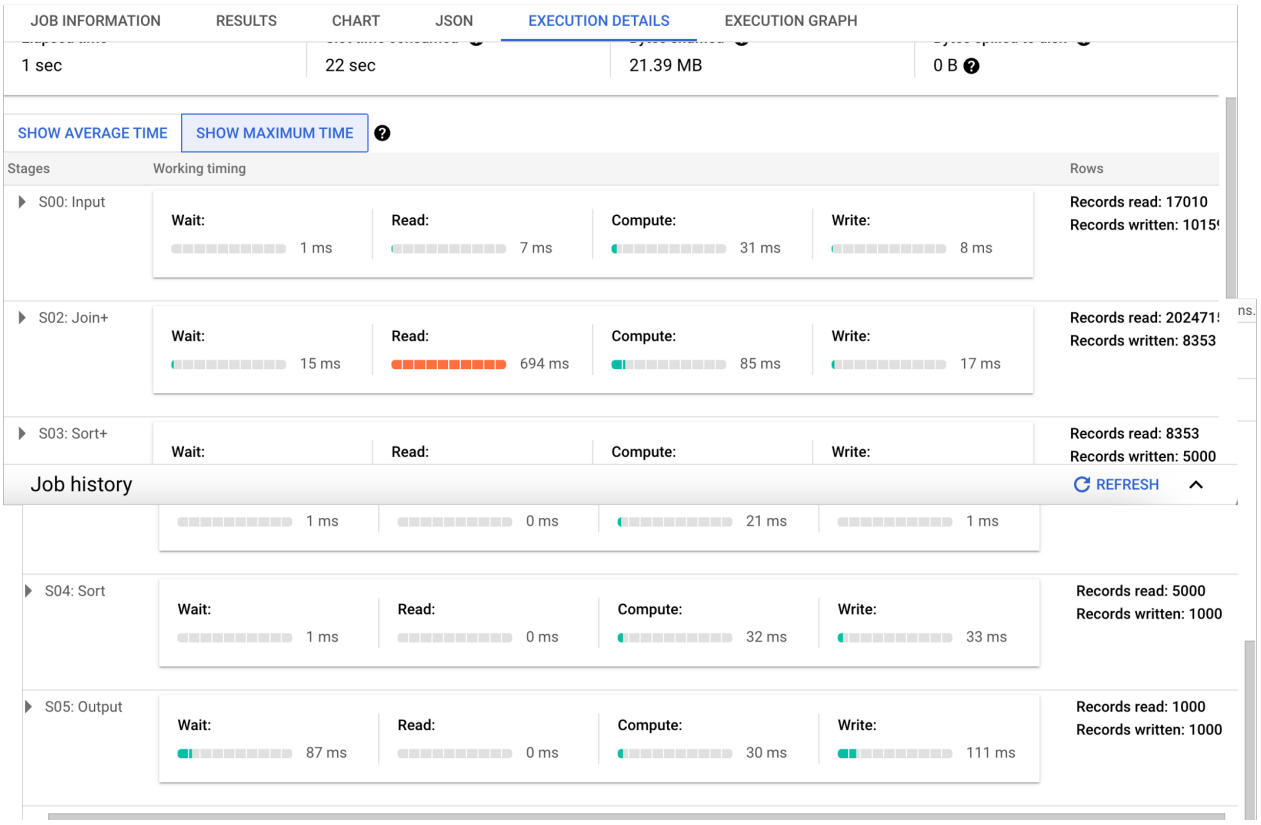

Figure 2: Query 1 IO Cost Before Optimization

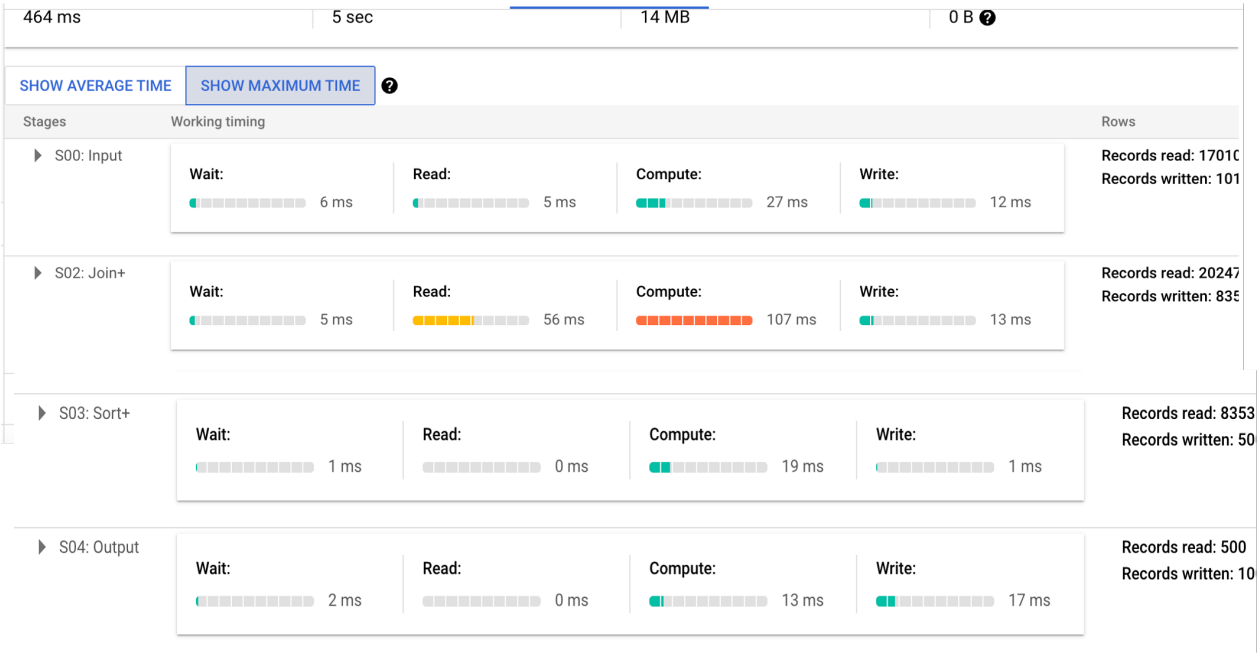

Figure 3: Query 1 IO Cost After Optimization

- Important Query 2 : California Weekly Political Ad Spends in 2024

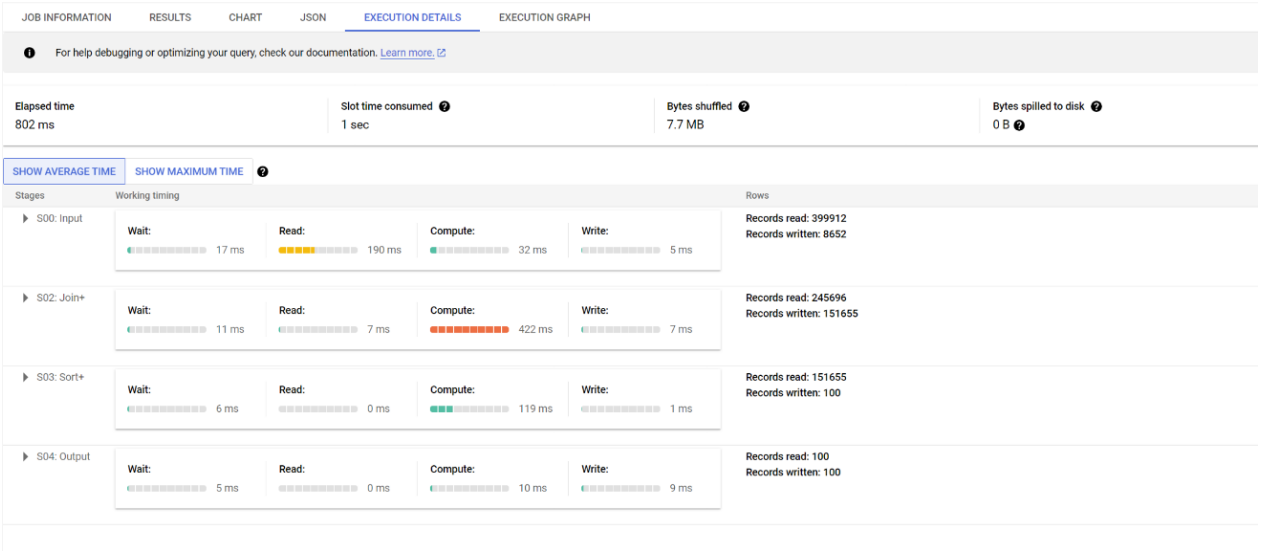

Figure 4: Query 2 IO Cost Before Optimization

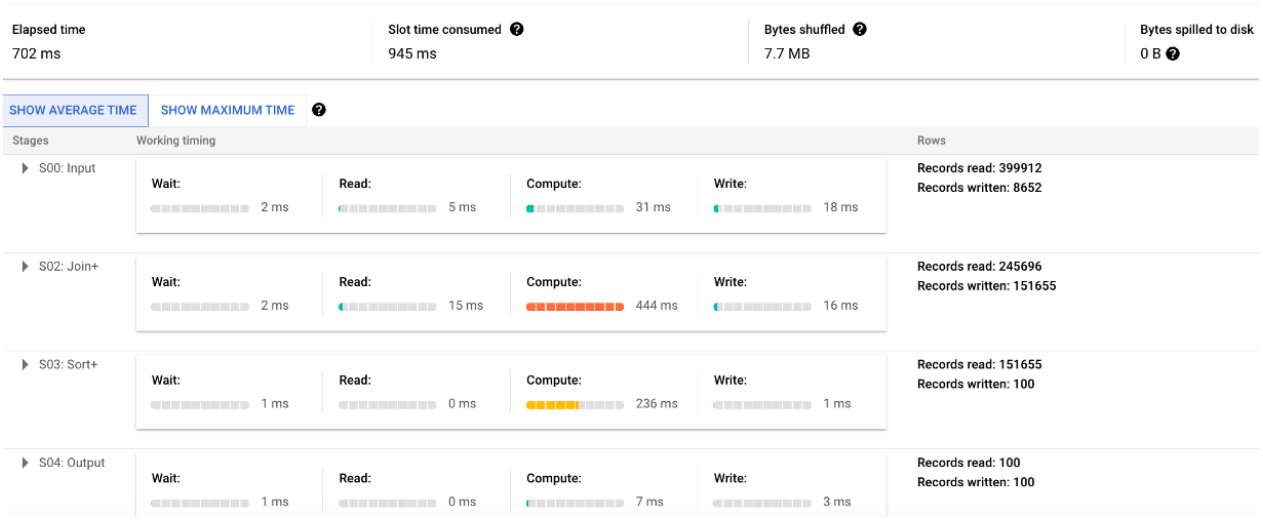

Figure 5: Query 2 IO Cost After Optimization

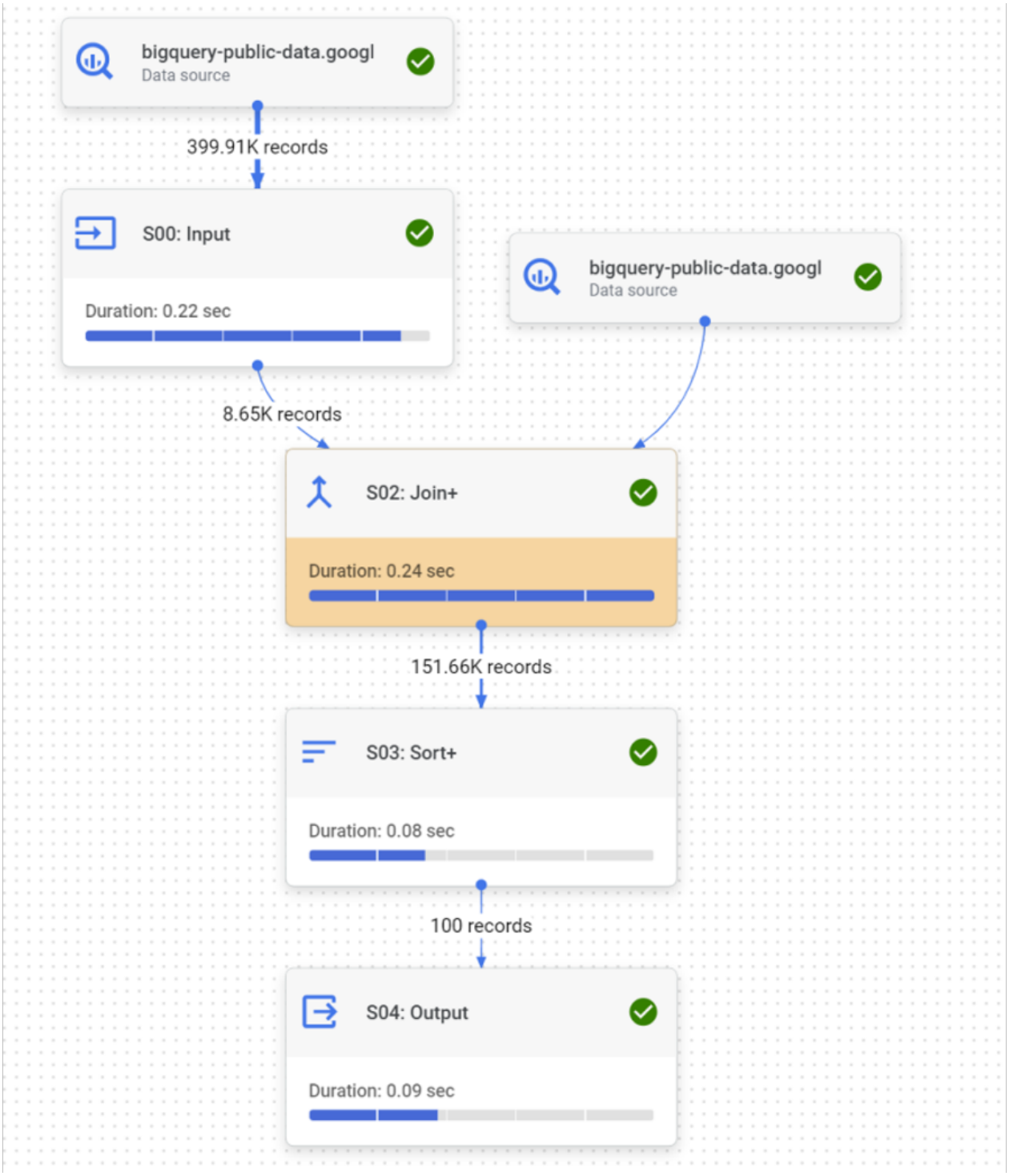

Figure 6: Query 2 Execution Plan Before Optimization

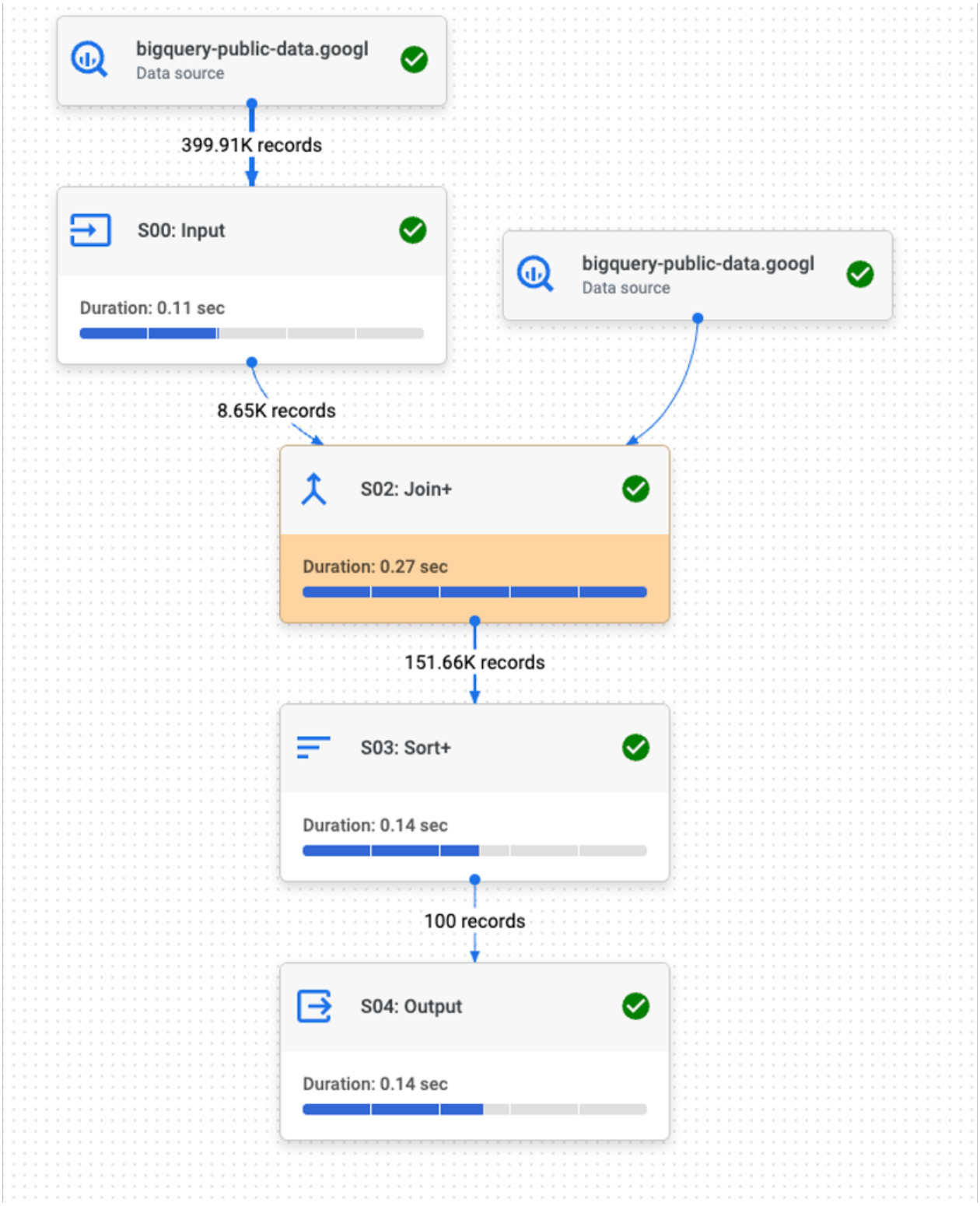

Figure 7: Query 2 Execution Plan After Optimization

- Important Query 3 : Demographics Targeted and Impact by Top10 US Political Ad Spends 2024

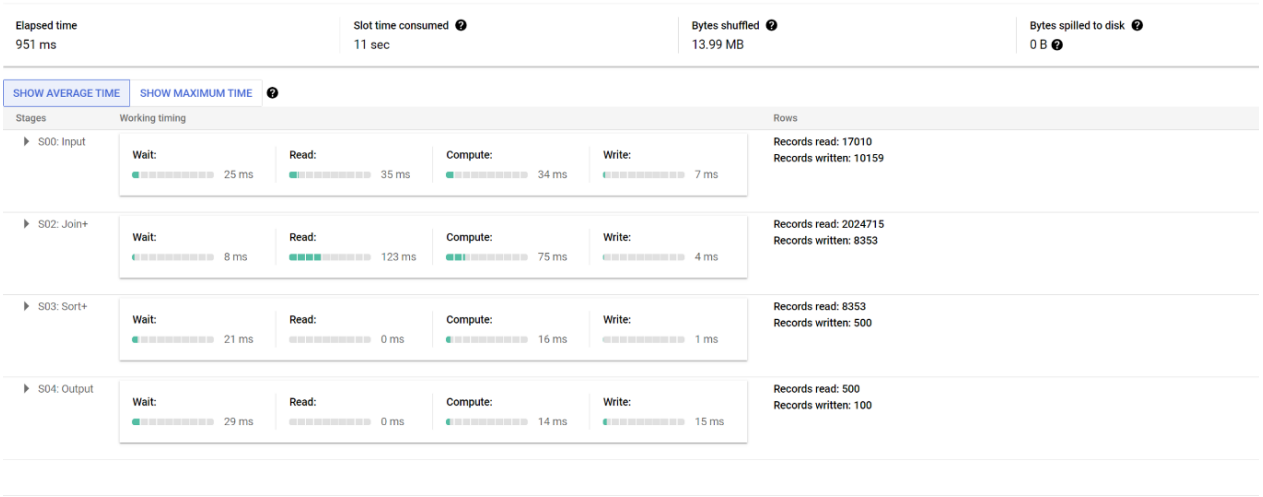

Figure 8: Query 3 IO Cost Before Optimization

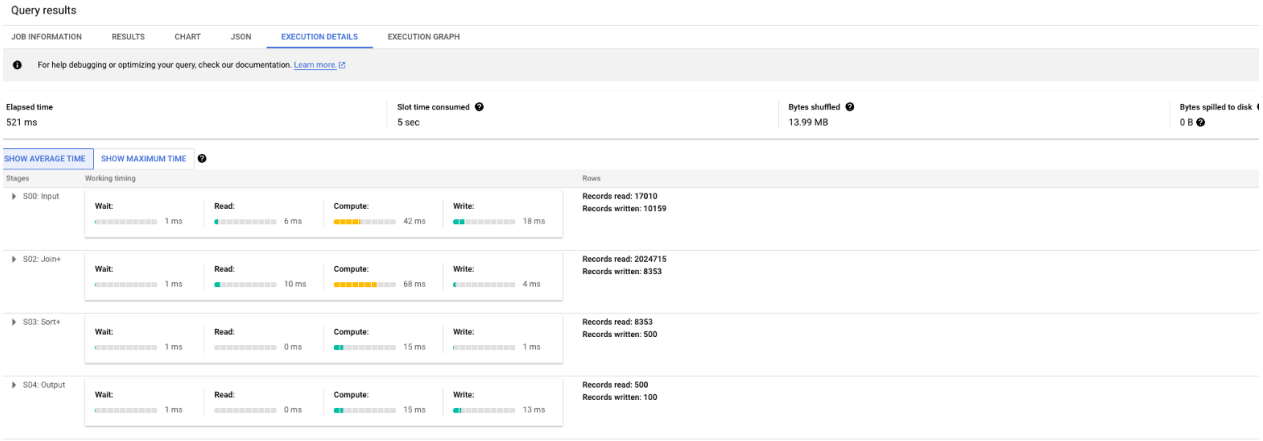

Figure 9: Query 3 IO Cost After Optimization

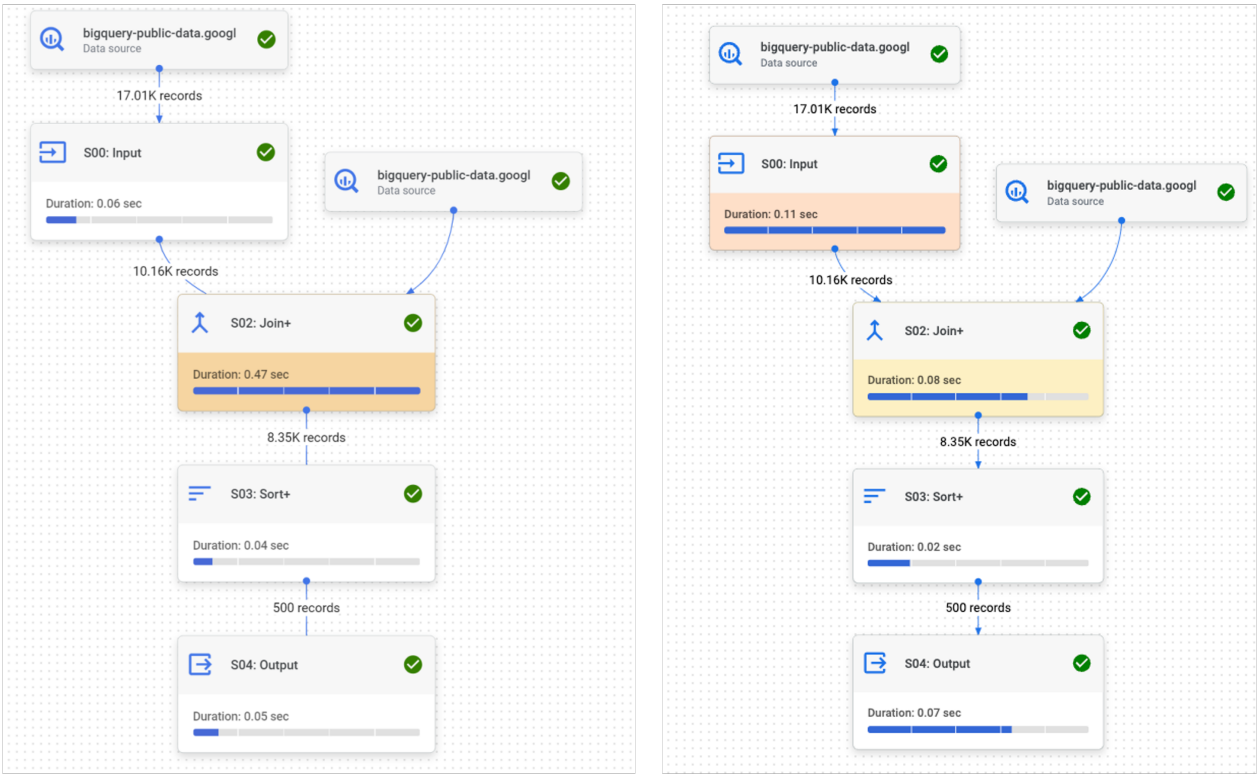

Figure 10: Query 3 Execution Plan Before and After Optimization

- Important Query 4 : Total Political Ad Spending by Advertisers, US State Wis

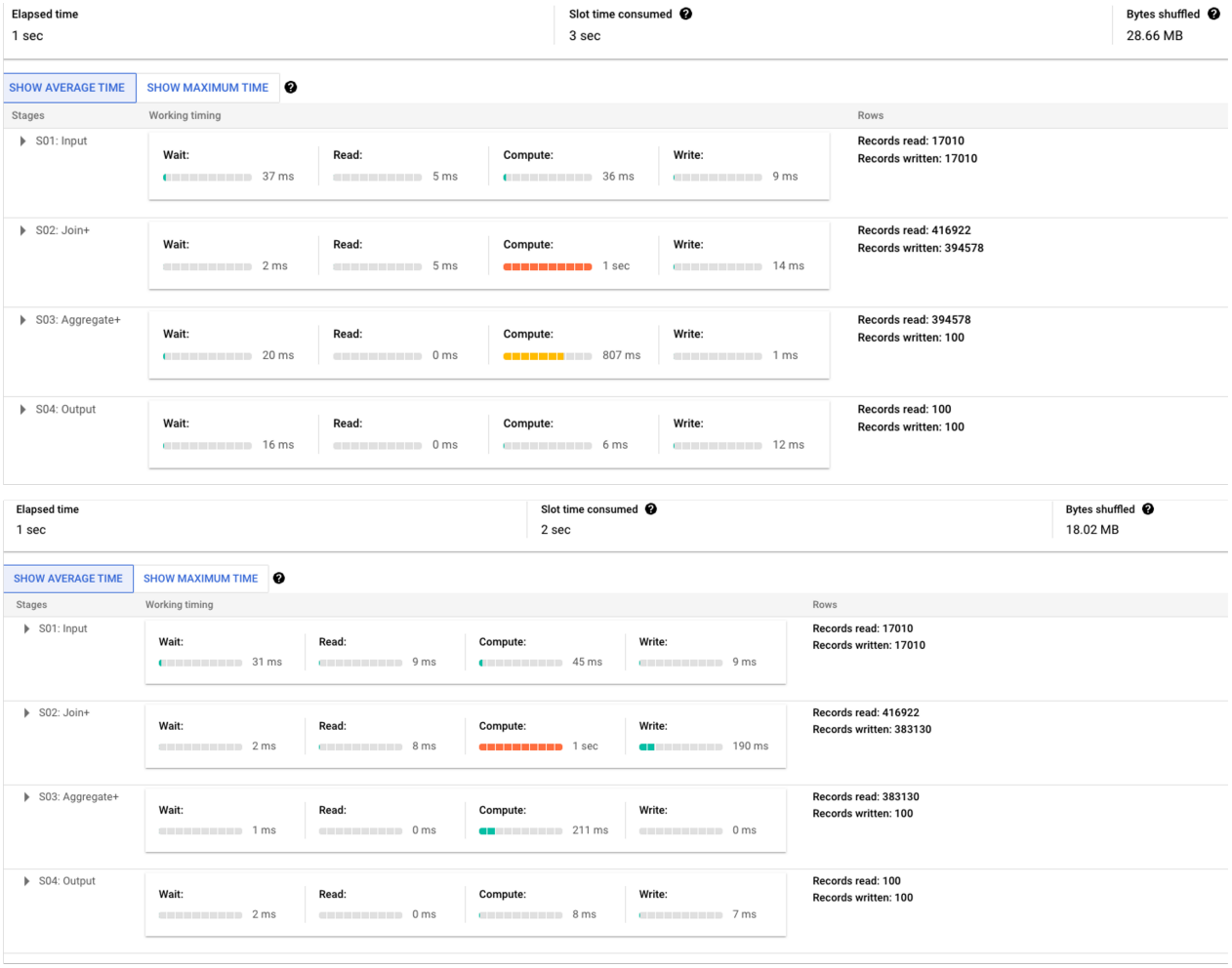

Figure 11: Query 4 IO Cost Before and After Optimization

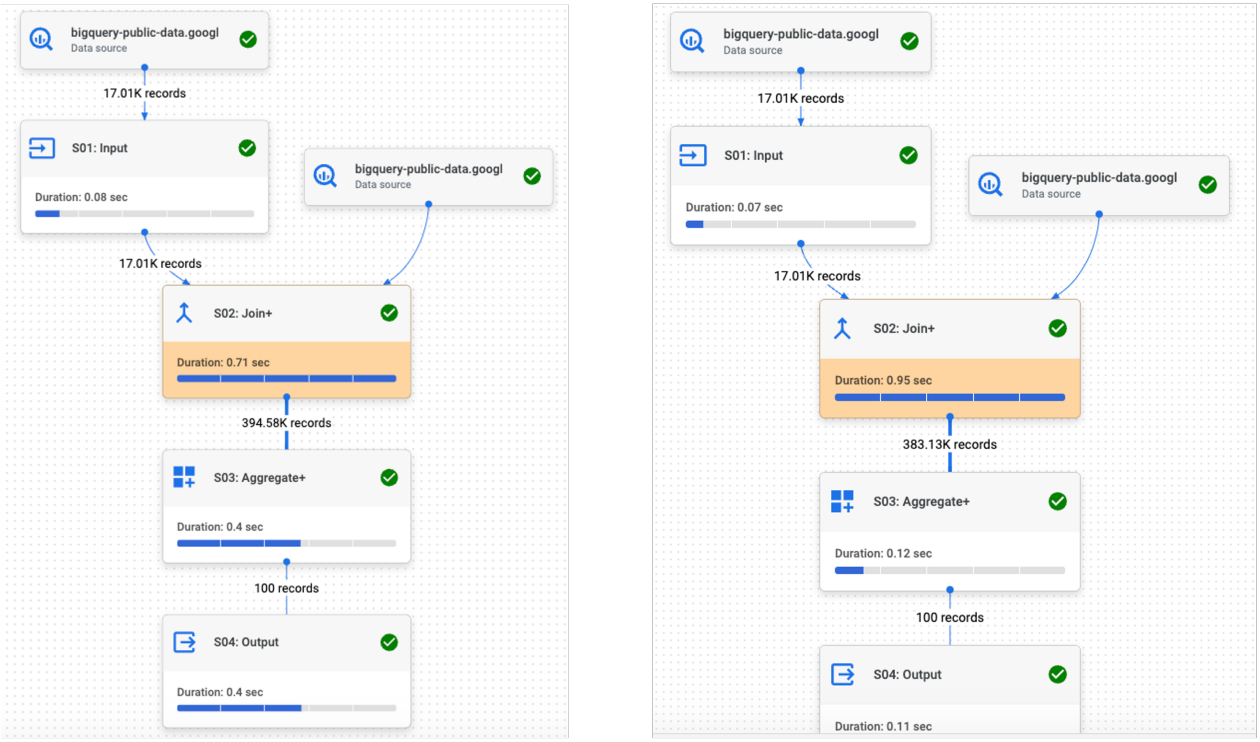

Figure 12: Query 4 Execution Plan Before and After Optimization

3.0.5 Project GitHub Repository

`https://github.com/RaviTejaGattu1/ClearCampaign_AppDataLayer/tree/main`

4 Analysis Of Algorithms

4.0.1 Most Important Queries for ClearCampaign

Important Queries For ClearCampaign:

- $AllTime\_Top10\_USAPoliticalAdCampaigns\_bySpending$
  - This query fetches data of Top 10 Political Campaigns in the decreasing order of advertisement spending of all time.
- $California\_2024\_Weekly\_PoliticalAdSpends$
  - This query provides data of spending for the political advertisements utilized in California state of the US, grouped weekly in 2024.
- $Demographics\_Targetted\_and\_Impact\_by\_Top10\_USPoliticalAdSpends\_2024$
  - This query provides the demographics targeted by the advertisements (age, gender, region etc), spending of the ads in USD and the impact of ads in terms of impressions.
- $US\_State\_wise\_Total\_Spending\_by\_Advertiser$
  - The data of political ad spending by each advertiser for the political advertisements used in US, ordered state-wise is the output.

Important Algorithms For ClearCampaign :

- SELECT
  - It chooses the data to be output from the columns of tables selected.
- FROM
  - This helps to specify the tables we are intending to select from for a particular query.
- JOIN
  - It combines the two / more tables based on the join condition which are common columns ensuring that each row in certain table matches with the corresponding row(s) in the other table/tables.
- WHERE
  - It filters the rows where the condition specified is met, focusing only on data required.
- GROUP BY

  - It groups the results by specified values which can be column data. This also ensures that the aggregation functions like (SUM()) operate on distinct groups rather than the entire dataset.

- LIMIT
  - It restricts the number of rows returned to number specified, displaying only the top n values selected.

4.0.2 Important Features, Queries, Algorithms/PseudoCode- Reasons For Choosing

Query 1: Top 10 All Time Political Ad Campaigns in USA by Spending

- **Algorithm:**
- The algorithm involves filtering data from the creative_stats and advertiser_stats tables to create temporary tables (FilteredCreativeStats and FilteredAdvertiserStats) using CTEs. It then joins the filtered data on advertiser_id, aggregates spending using SUM(), and orders the results by spending in descending order before limiting the output to the top 100 rows.
- **Functions/Features Used:**
- CTEs, Filtering, Joining, Aggregation, Ordering, Limiting.
- **Reasons:**
- CTEs enhance query readability and maintainability. Filtering improves query efficiency by reducing the dataset size. Joining efficiently combines data from different tables. Aggregation summarizes the data effectively. Ordering and limiting help in presenting the top results efficiently.

Query 2: California Weekly Political Ad Spends in 2024

- **Algorithm:**
- The algorithm involves filtering data from the advertiser_geo_spend table to create a temporary table (FilteredGeoSpend) using a CTE.
  It then joins the advertiser_weekly_spend table with FilteredGeoSpend based on advertiser_id, groups the results by country subdivision, week start date, and advertiser name, aggregates spending using SUM(), orders the results by country subdivision and week start date in descending order, and limits the output to the top 100 rows.
- **Functions/Features Used:**
- CTEs, Filtering, Joining, Aggregation, Ordering, Limiting.
- **Reasons:**

- CTEs enhance query readability and maintainability. Filtering optimizes query performance by reducing the dataset size. Joining efficiently combines relevant data. Aggregation summarizes spending data effectively. Ordering and limiting facilitate presenting the top results efficiently.

Query 3: Demographics Targeted and Impact by Top10 US Political Ad Spends 2024

  - **Algorithm:**
  - The algorithm involves filtering data from the creative_stats and advertiser_stats tables to create temporary tables (FilteredCreativeStats and FilteredAdvertiserStats) using CTEs. It then joins the filtered data on advertiser_id, aggregates spending using SUM(), and orders the results by spending in descending order before limiting the output to the top 100 rows.
  - **Functions/Features Used:**
  - CTEs, Filtering, Joining, Aggregation, Ordering, Limiting.
  - **Reasons:**
  - CTEs enhance query readability and maintainability. Filtering improves query efficiency by reducing the dataset size. Joining efficiently combines data from different tables. Aggregation summarizes the data effectively. Ordering and limiting help in presenting the top results efficiently.

Query 4: Total Political Ad Spending by Advertisers, US State Wise

- **Algorithm:**
- The algorithm involves joining the advertiser_stats table with the advertiser_geo_spend table based on advertiser_id, filtering the joined data to include only records where the country is 'US', grouping the results by advertiser name and country subdivision primary, aggregating spending using SUM(), ordering the results by advertiser name and country subdivision primary, and limiting the output to the top 100 rows.
- **Functions/Features Used:**
- Joining, Aggregation, Filtering, Ordering, Limiting.
- **Reasons:**
- Joining efficiently combines relevant data from different tables. Aggregation summarizes spending data effectively. Filtering optimizes query performance by reducing the dataset size. Ordering and limiting help in presenting the top results efficiently.

4.0.3 Inputs And Outputs Of Queries

Query 1: Top 10 All Time Political Ad Campaigns in USA by Spending
Inputs of `creative_stats`:

- advertiser_id
- last_served_timestamp
- spend_range_max_usd
- impressions
- ad_type
- age_targeting
- gender_targeting
- geo_targeting_included
- date_range_start

Inputs of `advertiser_stats`:

- advertiser_id
- regions
- advertiser_name
- spend_usd

Output :

Query results

SAVE RESULTS | EXPLORE DATA

JOB INFORMATION | RESULTS | CHART | JSON | EXECUTION DETAILS | EXECUTION GRAPH

Row	regions	year	advertiser_name	Spending	impressions	ad_type	age_targeting	gender_targeting	geo_targeting_included
1	US	2024	STANDING STRONG PAC	1500000	≥10000000	VIDEO	35-44, 45-54, 55-64, ≥65	Male, Female, Unknown gender	California,United States
2	US	2024	BIDEN VICTORY FUND	650000	300000-350000	TEXT	18-24, 25-34, 35-44, 45-54, 55-6…	Male, Female, Unknown gender	Alabama,United States, Alaska,United States, Arizona,United States, Arkansas,United States, California,United States,
3	US	2024	HOMEOWNERS AND TENANTS…	390000	≥10000000	VIDEO	18-24, 25-34, 35-44, 45-54, 55-6…	Male, Female, Unknown gender	California,United States
4	US	2024	BIDEN VICTORY FUND	350000	250000-300000	TEXT	18-24, 25-34, 35-44, 45-54, 55-6…	Male, Female, Unknown gender	Alabama,United States, Alaska,United States, Arizona,United States, Arkansas,United States, California,United States,
5	US	2024	Fairshake	350000	9000000-10000000	VIDEO	18-24, 25-34, 35-44, 45-54, 55-6…	Male, Female, Unknown gender	California,United States
6	US	2024	Safety For All Newsom Ballot …	325000	9000000-10000000	VIDEO	35-44, 45-54, 55-64, ≥65	Male, Female, Unknown gender	California,United States
7	US	2024	BIDEN VICTORY FUND	300000	50000-60000	TEXT	18-24, 25-34, 35-44, 45-54, 55-6…	Male, Female, Unknown gender	Alabama,United States, Alaska,United States, Arizona,United States, Arkansas,United States, California,United States,

Figure 13: Q1 Output

Query 2: California Weekly Political Ad Spends in 2024
Inputs of `advertiser_geo_spend`:

- advertiser_id
- country_subdivision_primary

Inputs of `advertiser_weekly_spend`:

- week_start_date
- advertiser_name
- spend_usd

Output :

Query results

JOB INFORMATION | RESULTS | CHART | JSON | EXECUTION DETAILS | EXECUTION GRAPH

Row	country_subdivision_primary	week_start_date	advertiser_name	total_weekly_spend_
1	CA	2024-04-28	AFC VICTORY FUND	24800
2	CA	2024-04-28	National Democratic Redistricti...	100
3	CA	2024-04-28	QARNI FOR CONGRESS	1800
4	CA	2024-04-28	Move Oregon's Border - Wallow...	300
5	CA	2024-04-28	FRIENDS OF ALYIA GASKINS	800
6	CA	2024-04-28	KIGGANS FOR CONGRESS	400
7	CA	2024-04-28	COMPETE DIGITAL LLC	100
8	CA	2024-04-28	Epoch Group Inc	1300
9	CA	2024-04-28	STATE DEMOCRATIC EXECUTI...	500
10	CA	2024-04-28	State Affairs Inc	400
11	CA	2024-04-28	KARI LAKE FOR SENATE	2500
12	CA	2024-04-28	TROY DOWNING FOR CONGRE...	3000
13	CA	2024-04-28	LAW FORWARD INC	100
14	CA	2024-04-28	Friends of John Barrow	16300
15	CA	2024-04-28	ROSEN FOR NEVADA	20500
16	CA	2024-04-28	Evergreen Collaborative	1600
17	CA	2024-04-28	VINCE LEACH FOR SENATE	100

Figure 14: Q2 Output

Query 3: Demographics Targeted and Impact by Top 10 US Political Ad Spends 2024

Inputs of `creative_stats`:

- advertiser_id
- last_served_timestamp
- spend_range_max_usd
- impressions
- ad_type
- age_targeting
- gender_targeting
- geo_targeting_included
- date_range_start

Inputs of `advertiser_stats`:

- advertiser_id
- regions
- advertiser_name
- spend_usd

Output :

Query results

SAVE RESULTS EXPLORE DATA

JOB INFORMATION | RESULTS | CHART | JSON | EXECUTION DETAILS | EXECUTION GRAPH

Row	regions	year	advertiser_name	Spending	impressions	ad_type	age_targeting	gender_targeting	geo_targeting_included
2	US	2024	BIDEN VICTORY FUND	650000	300000-350000	TEXT	18-24, 25-34, 35-44, 45-54, 55-6...	Male, Female, Unknown gender	Alabama,United States, Alaska,United States, Arizona,United States, Arkansas,United States, California,United States,
3	US	2024	HOMEOWNERS AND TENANTS...	390000	≥10000000	VIDEO	18-24, 25-34, 35-44, 45-54, 55-6...	Male, Female, Unknown gender	California,United States
4	US	2024	Fairshake	350000	9000000-10000000	VIDEO	18-24, 25-34, 35-44, 45-54, 55-6...	Male, Female, Unknown gender	California,United States
5	US	2024	BIDEN VICTORY FUND	350000	250000-300000	TEXT	18-24, 25-34, 35-44, 45-54, 55-6...	Male, Female, Unknown gender	Alabama,United States, Alaska,United States, Arizona,United States, Arkansas,United States, California,United States,
6	US	2024	Safety For All Newsom Ballot ...	325000	9000000-10000000	VIDEO	35-44, 45-54, 55-64, ≥65	Male, Female, Unknown gender	California,United States
7	US	2024	BIDEN VICTORY FUND	300000	50000-60000	TEXT	18-24, 25-34, 35-44, 45-54, 55-6...	Male, Female, Unknown gender	Alabama,United States, Alaska,United States,

Figure 15: Q2 Output

Query 4: Total Political Ad Spending by Advertisers, US State Wise

Inputs of `advertiser_stats`:

- advertiser_id

- advertiser_name

Inputs of `advertiser_geo_spend`:

- advertiser_id
- country_subdivision_primary
- spend_usd

Output :

JOB INFORMATION | RESULTS | CHART | JSON | EXECUTION DETAILS

Row	advertiser_name	country_subdivision_primary	total_spending
1	VINCE LEACH FOR SENATE	AZ	13500
2	TINA SMITH FOR MINNESOTA	CA	3400
3	TINA SMITH FOR MINNESOTA	WI	2300
4	TINA SMITH FOR MINNESOTA	MI	700
5	TINA SMITH FOR MINNESOTA	ND	800
6	TINA SMITH FOR MINNESOTA	WA	700
7	TINA SMITH FOR MINNESOTA	MD	700
8	TINA SMITH FOR MINNESOTA	GA	700
9	TINA SMITH FOR MINNESOTA	MA	700
10	TINA SMITH FOR MINNESOTA	CO	800
11	TINA SMITH FOR MINNESOTA	PA	1000
12	TINA SMITH FOR MINNESOTA	NC	700
13	TINA SMITH FOR MINNESOTA	TX	2400
14	TINA SMITH FOR MINNESOTA	MN	285600
15	TINA SMITH FOR MINNESOTA	NJ	1000

Figure 16: Q4 Output

4.0.4 Optimization Algorithms

Query Optimization Results
Comparison Of IO Cost Before vs After Optimization

- Important Query 1 : Top 10 All Time Political Ad Campaigns in USA by Spending

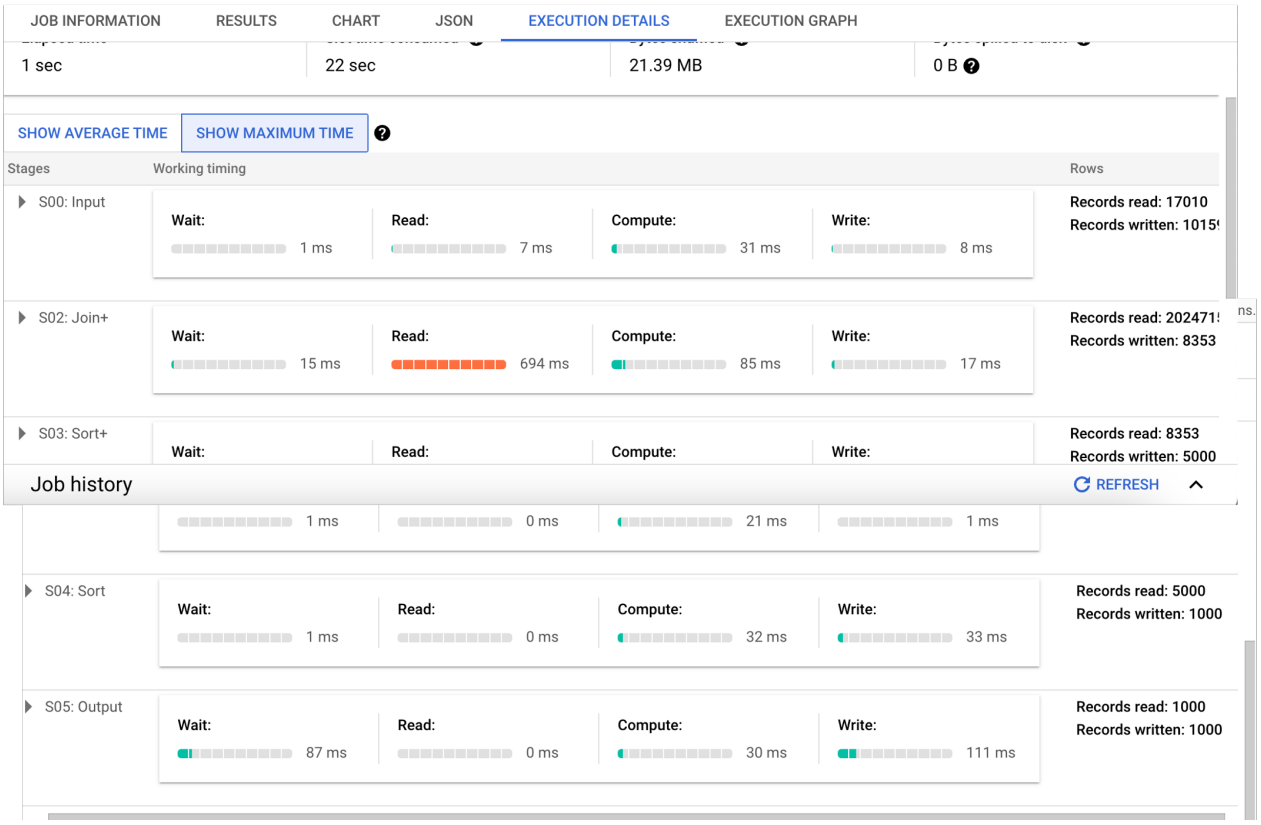

Figure 17: Query 1 IO Cost Before Optimization

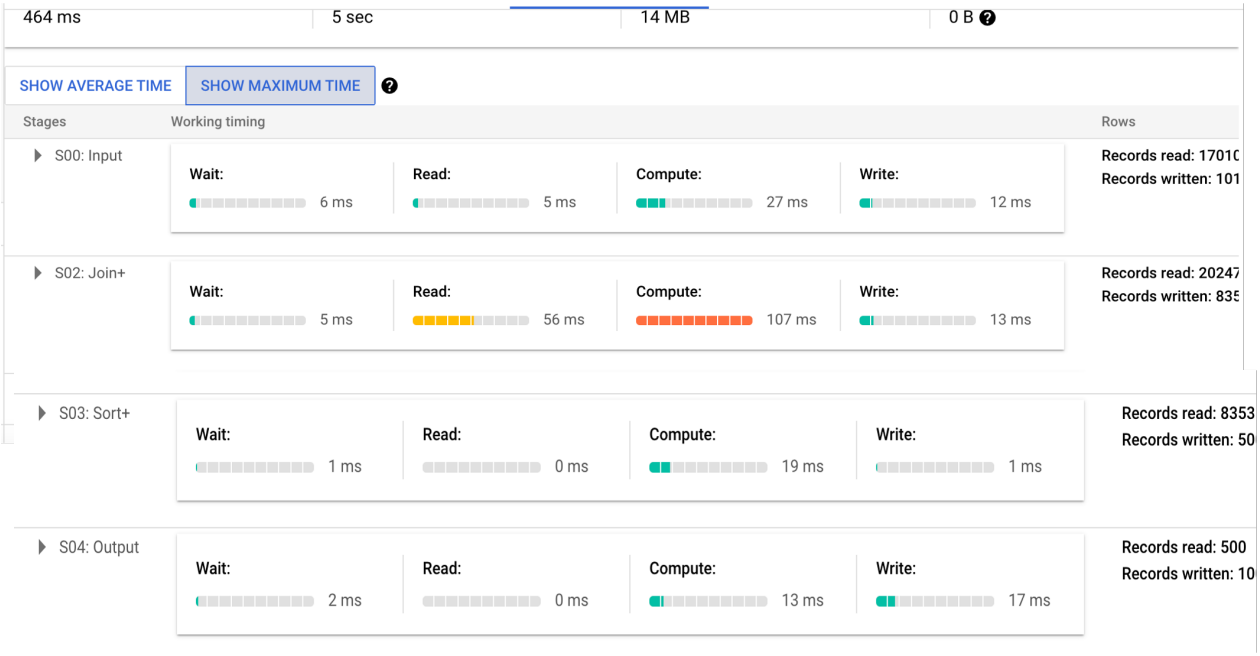

Figure 18: Query 1 IO Cost After Optimization

- Important Query 2 : California Weekly Political Ad Spends in 2024

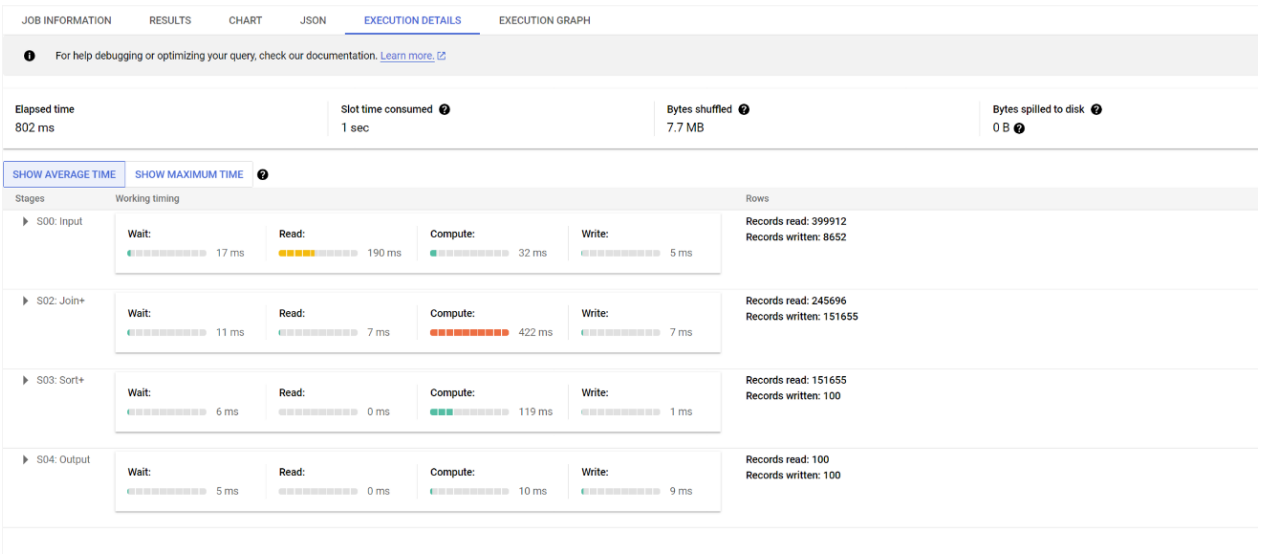

Figure 19: Query 2 IO Cost Before Optimization

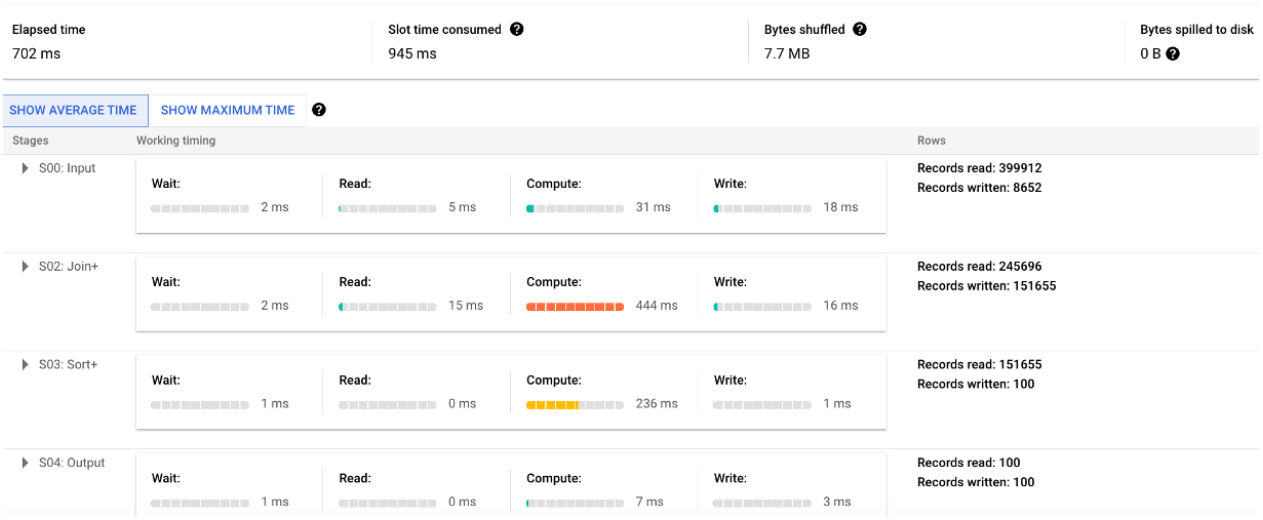

Figure 20: Query 2 IO Cost After Optimization

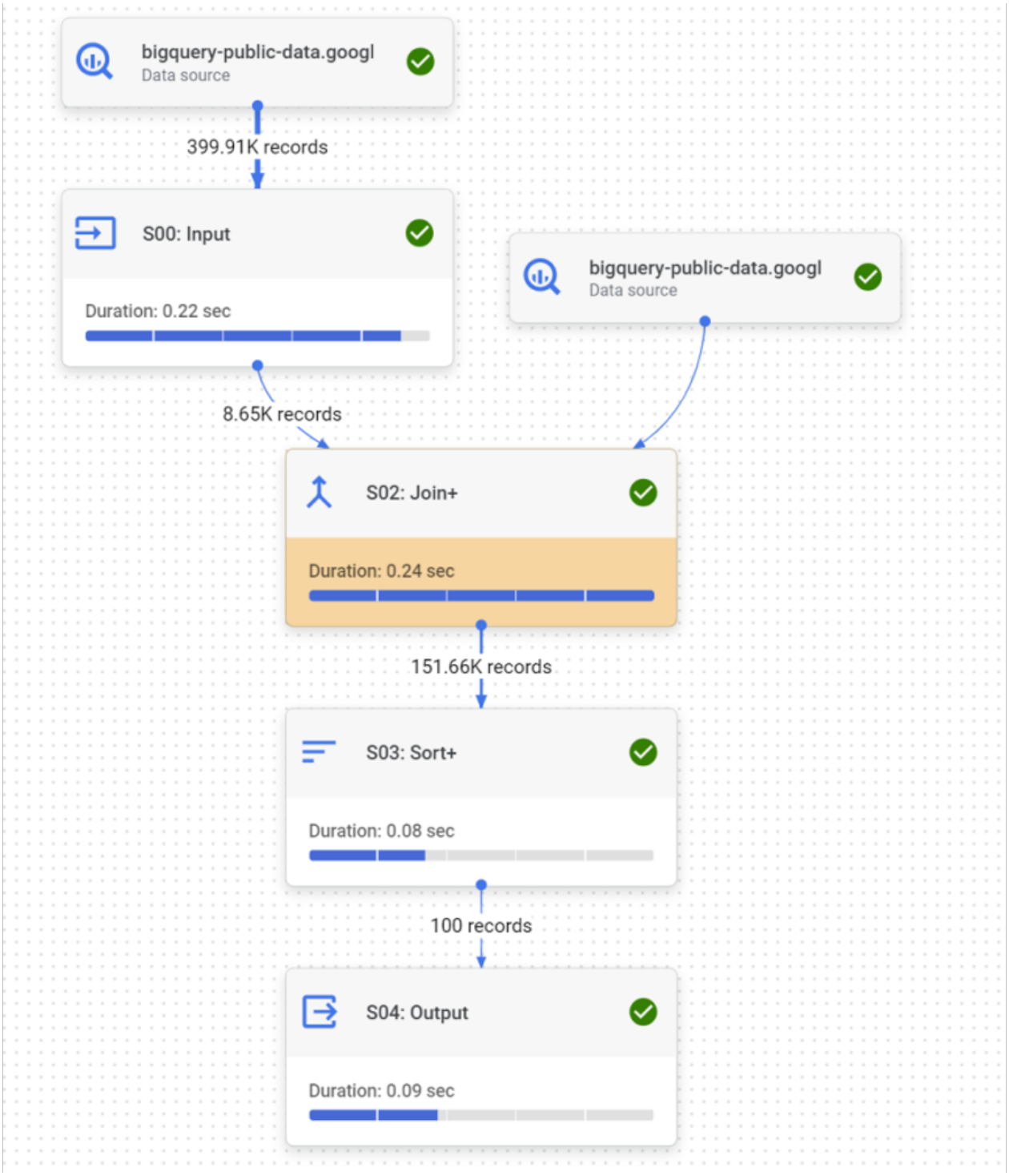

Figure 21: Query 2 Execution Plan Before Optimization

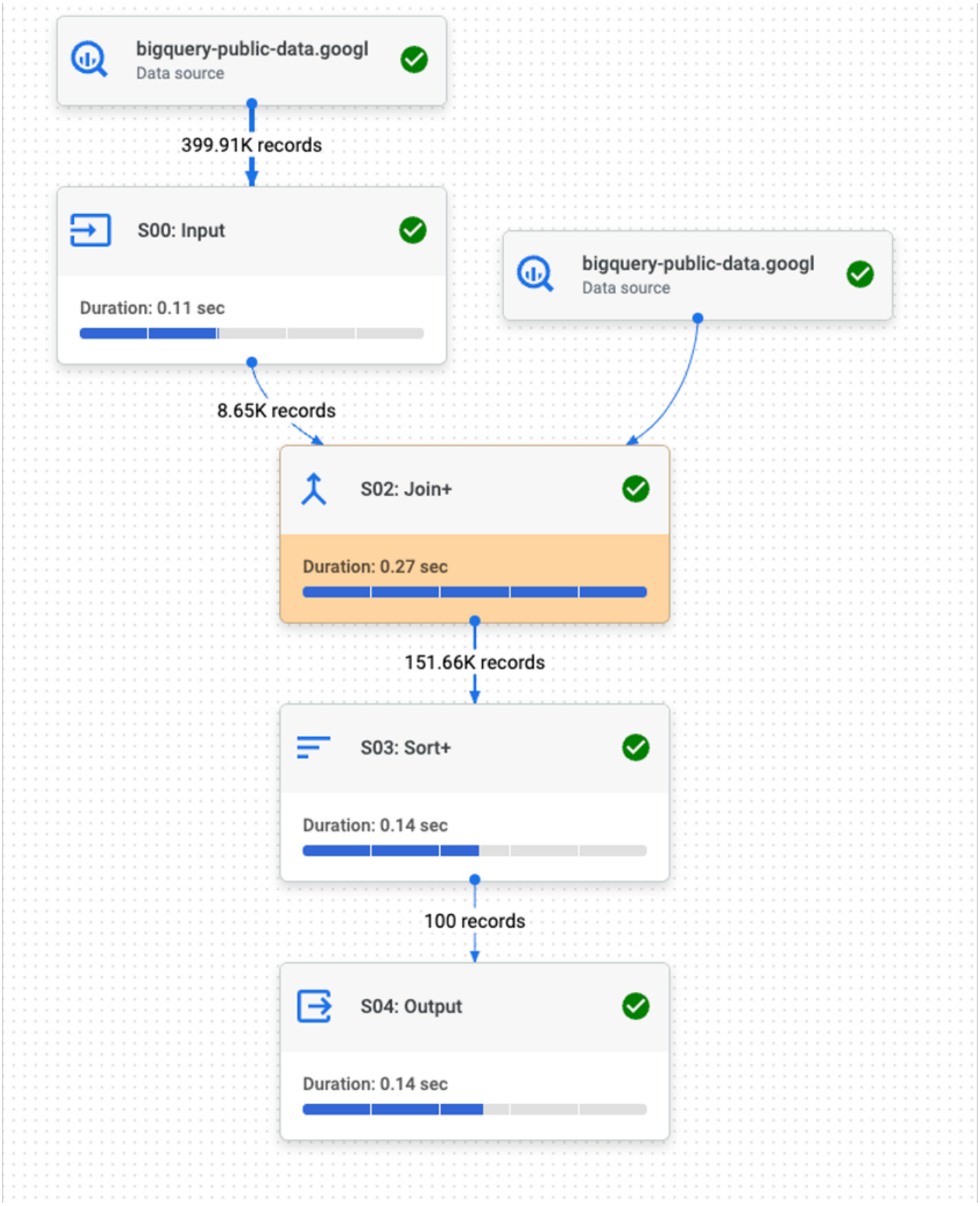

Figure 22: Query 2 Execution Plan After Optimization

- Important Query 3 : Demographics Targeted and Impact by Top10 US Political Ad Spends 2024

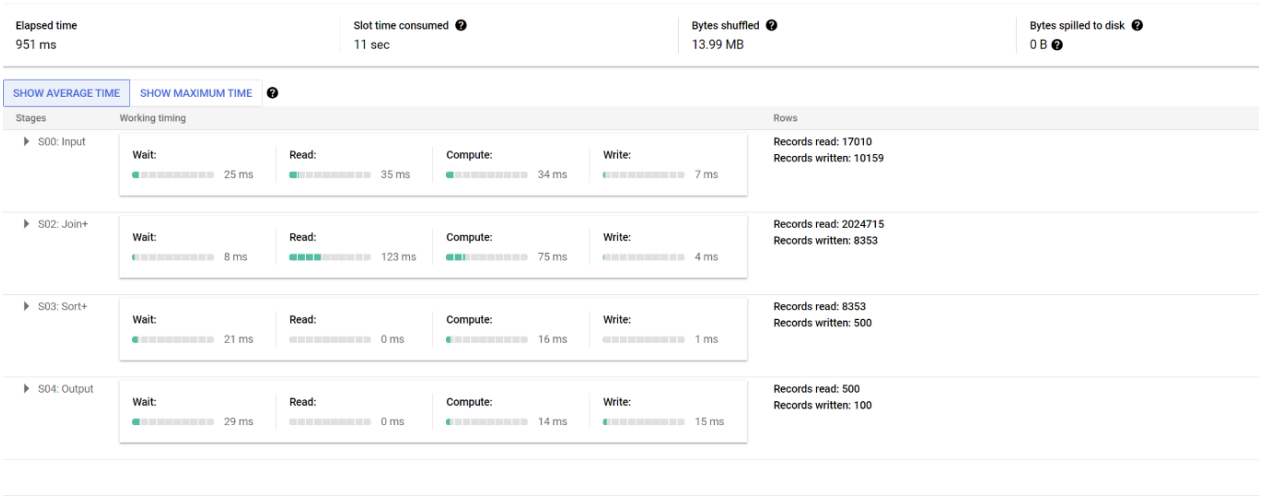

Figure 23: Query 3 IO Cost Before Optimization

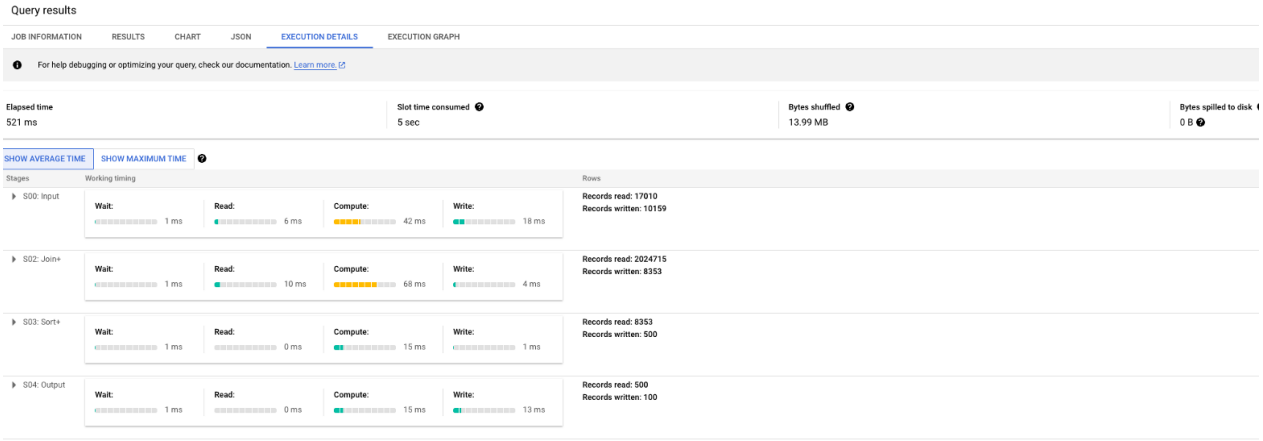

Figure 24: Query 3 IO Cost After Optimization

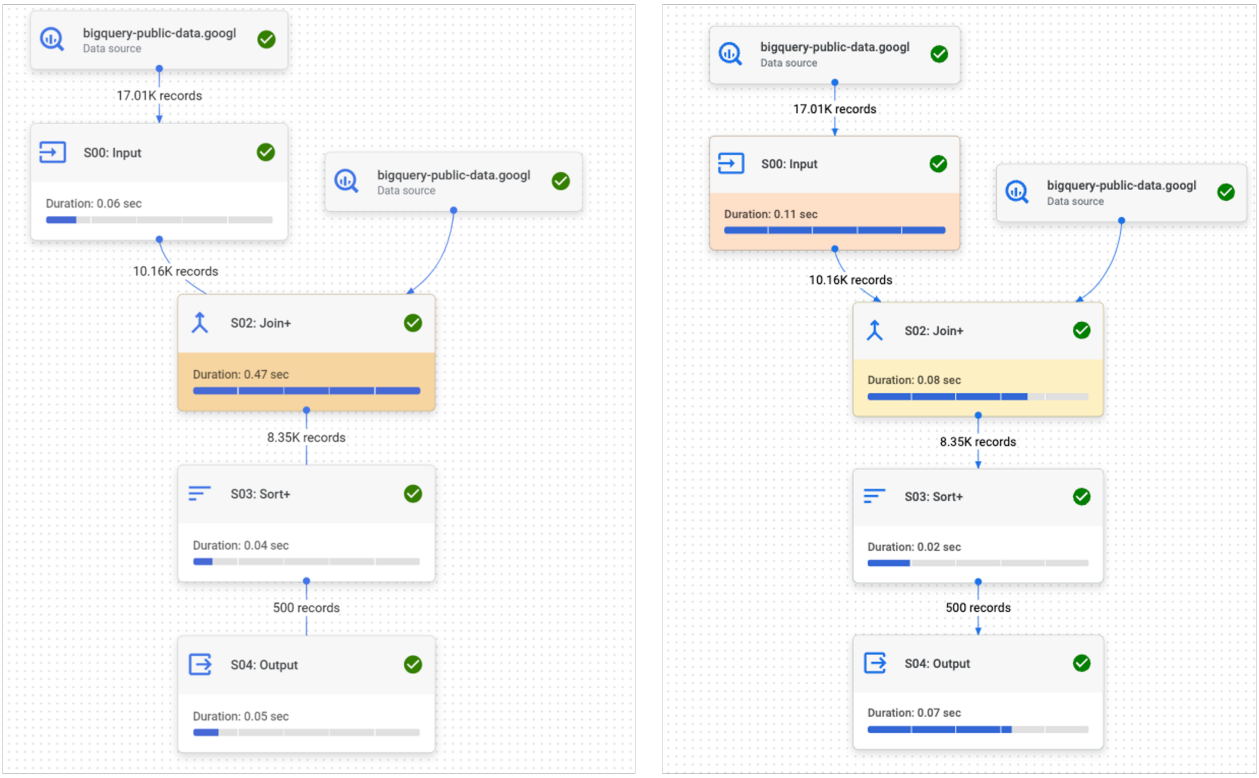

Figure 25: Query 3 Execution Plan Before and After Optimization

- Important Query 4 : Total Political Ad Spending by Advertisers, US State Wise

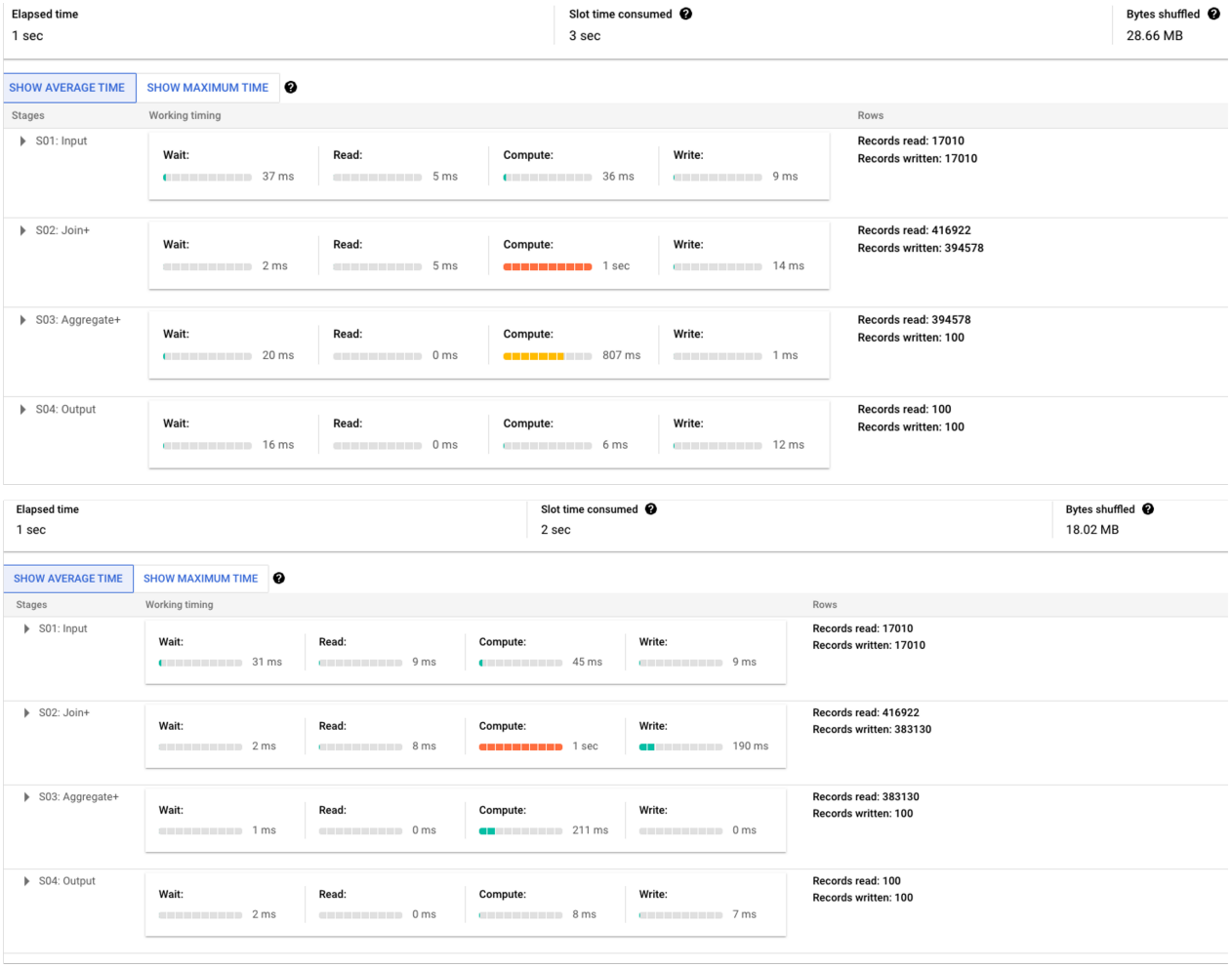

Figure 26: Query 4 IO Cost Before and After Optimization

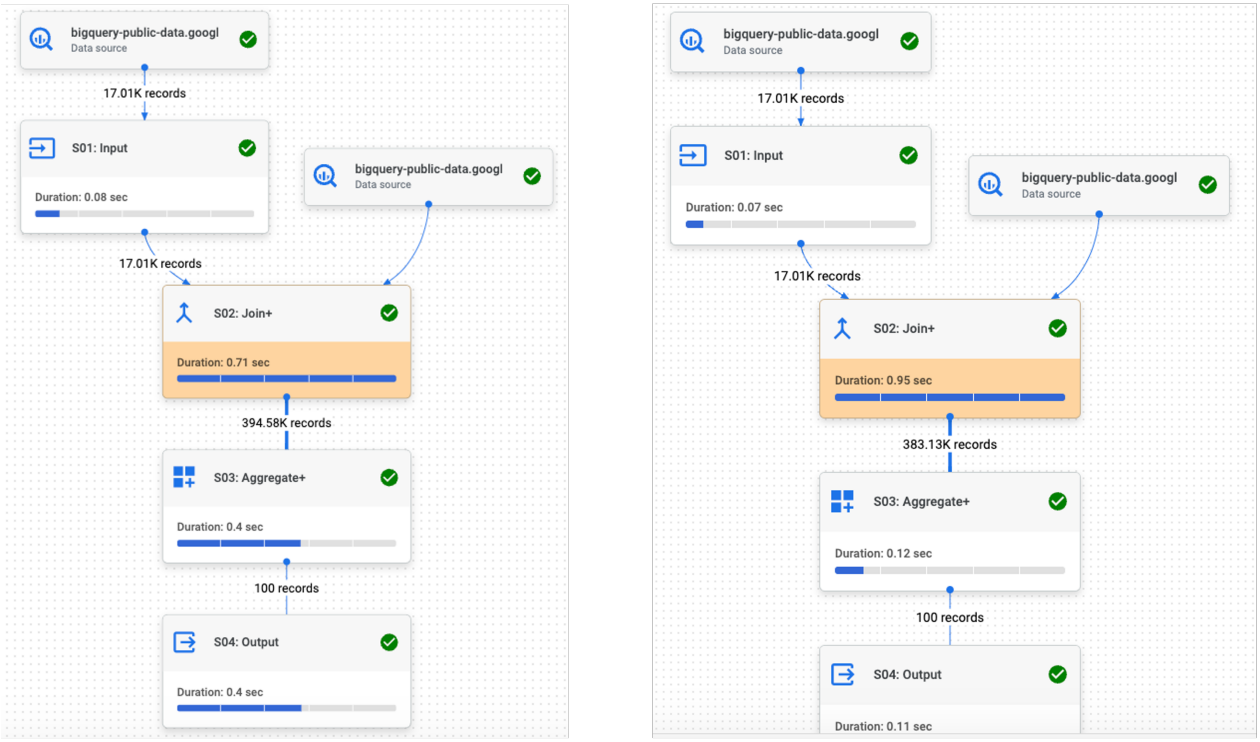

Figure 27: Query 4 Execution Plan Before and After Optimization

4.1 Estimation of Running Time in Relation to Input Size

Considering Main Query 1's Execution Details On BigQuery:

- Bytes shuffled: BigQuery shuffles data in order to support large-scale computation.
- Slot time: A slot is a unit of computational capacity required to execute SQL queries.
- Before optimization: Bytes shuffled/slot time
  = 28.83MB/5sec = 5.6MB/sec
- After optimization: Bytes shuffled/slot time
  = 28.83MB/4sec = 7.2MB/sec

5 Results And Findings

5.0.1 Solution Evaluation And Metrics Used

Metrics :

- User Empathetic UI : The application interface is designed with a focus on user experience, ensuring ease of use and accessibility for all users.
- Quantity of Features : The application offers a wide range of features and functionalities to meet diverse user needs and requirements.
- Data Recency : The data provided by the application is regularly updated to reflect the most current information and trends.
- Data Ease of Access : Users can easily access and retrieve data from the application without encountering significant barriers or limitations.
- Data Filters : The application provides robust filtering options, allowing users to refine and customize their data queries according to specific criteria or preferences.
- Global Data Presence : The application offers data coverage on a global scale, providing insights and information from various regions and countries worldwide.
- Financial Data of Countries Qty : The application includes financial data for a total of three countries, enabling users to analyze and compare economic indicators across different regions.

Solution Evaluation :

Metric	**Description**
User Empathetic UI	Yes
Quantity of Features	High
Data Recency	High
Data Ease of Access	High
Data Filters	High
Global Data Presence	Yes
Financial Data of Countries Qty	3

Table 1: ClearCampaign Features and Application Evaluation on Metrics

5.0.2 Alternative Solutions' Evaluation And Comparison with ClearCampaign

Rank	App	User Empathetic UI	Quantity of Features	Data Recency	Data Ease of Access	Data Filters	Global Data Presence	FDCQ
1	ClearCampaign	Yes	High	High	High	High	Yes	3
2	Meta's FORT	No	High	High	Low	High	Yes	NA
3	OpenSecrets	Yes	Medium	High	Low	Medium	No	1
4	AdImpact	No	Low	High	Low	Low	No	1
5	Bloomberg Government	No	Low	High	Low	Low	No	1
6	Roku Ad Platform	No	Low	High	Low	Low	No	1

Table 2: Evaluation of Apps Based on Metrics with Rank

5.0.3 Approximations and Calculations

Considering Main Query 1:

- Bytes shuffled: BigQuery shuffles data in order to support large-scale computation.
- Slot time: A slot is a unit of computational capacity required to execute SQL queries.
- Before optimization: Bytes shuffled/slot time
  = 28.83MB/5sec = 5.6MB/sec
- After optimization: Bytes shuffled/slot time
  = 28.83MB/4sec = 7.2MB/sec

Main Query 4:

- Read cost per GB: 0.01 usd
- Write cost per GB: 0.02 usd
- Total average size of each record is 100bytes
- Read cost before optimization = (17061+417711+395205)*100*0.01
- = 0.0082 usd
- Write cost before optimization =
- (17061+395205+395205)*100*0.02
  = 0.016 usd
- Total I/O cost before optimization = (Read cost + write cost)
- =(0.0082+0.016) = 0.0242 usd

5.0.4 Suggestions For Future Improvements

- More data to include other countries political ad spending data as well.
- Structure of table with large number of columns is not helpful, ie., column for each currency to represent currency of each country. A single column with a comment to denote currency type will be more productive.

- Absence of indexes for the dataset.
- When tried to add search indexes, it resulted in failure due to insufficient permissions. Capability to create dataset replica and add user defined indexes to tables will be impactful to experience indexing hands-on and benefits of it.
- Few tables in dataset were deprecated. More features could have been possible if those tables were not deprecated.

6 Conclusions and Lessons Learnt

6.1 Conclusion

The culmination of this project marks a significant stride towards realizing the vision of fostering transparency in the political landscape through technological means. By harnessing the rich dataset provided by Google Political Ads, we have developed an application that empowers various stakeholders, including voters, political parties,
media professionals, government officials, and research enthusiasts. The app serves as a comprehensive platform for accessing and analyzing political ad campaign data, offering insights into
advertising spending, geographical trends, advertiser declarations, and creative statistics.
Through the meticulous exploration of the dataset and crafting of optimized queries, we have enabled features that cater to the diverse motivations of our user personas. By providing functionalities such as tracking advertising expenditures by political parties, monitoring adherence to campaign regulations, and facilitating media dissemination of information, our application adds tangible value to the political discourse.
One of the pivotal achievements of this project lies in the thorough analysis and comparison of existing applications in the market. By identifying their strengths and weaknesses, we have been able to refine our approach and design a solution that not only addresses existing needs but also anticipates future requirements in this domain.
In conclusion, the development of this application represents a significant step
towards democratizing access to political ad campaign data and promoting accountability in the electoral process. By leveraging data-driven insights, we structured the application to contribute to a more informed electorate and foster a healthier democratic ecosystem.

6.2 Learnings

6.2.1 Data Exploration and Query Optimization:

Working with large-scale datasets such as Google Political Ads necessitated a deep dive into data exploration techniques and query optimization strategies. Through iterative refinement, we learned to craft efficient queries that extract meaningful insights while minimizing computational overhead.

6.2.2 Stakeholder-Centric Design:

Understanding the diverse motivations and requirements of our user personas was crucial in shaping the features and functionalities of our application. By adopting a stakeholder-centric design approach, we ensured that our solution aligns closely with the needs of our target audience, thereby maximizing its utility and impact.

6.2.3 Ethical Considerations

The project underscored the importance of ethical considerations in handling sensitive political data. We prioritized privacy and security measures to safeguard user information and mitigate the risk of misuse or exploitation.

6.2.4 Continuous Improvement

The dynamic nature of the political landscape necessitates a commitment to continuous improvement and iteration. Moving forward, we recognize the importance of soliciting user feedback, staying abreast of emerging trends, and adapting our solution to evolving needs and challenges.
In summary, the project served as a valuable learning experience, enabling us to hone our technical skills, deepen our domain expertise, and cultivate a mindset of innovation and adaptability. As we reflect on our journey, we are inspired to apply these learnings in future endeavors, with a steadfast commitment to driving positive change through technology.

Bibliography

References

[1] "google-bigquery-publicdata," 2024. https://cloud.google.com/bigquery/public-data.

[2] "news.berkeley.edu," 2024. https://news.berkeley.edu/2024/02/15/deep-data-and-big-money-are-driving-a-new-era-in-political-advertising.

[3] "adimpact," 2024. https://adimpact.com/solution/tailored-insights/.

[4] "Fort," 2024. https://research.facebook.com/blog/2021/2/introducing-new-election-related-ad-data-sets-for-researchers/.

[5] "bloomberggov," 2024. https://about.bgov.com/request-a-demo-government-affairs/?bbgsum=AD-PA-GGL-Nbrand-H124880-bgov&utm_medium=paidsearch&utm_source=google&gad_source=1&gclid=Cj0KCQjwir2xBhC_ARIsAMTXk87kr-txAxB6Q3VOsz6IcWdBAmNcWHZaWPo-7d8_mvnaWUL2xdWZONsaArWDEALw_wcB&gclsrc=aw.ds/.

[6] "Opensecrets," 2024. https://www.opensecrets.org/ad-data.

[7] "roku," 2024. https://roku.sharepoint.com/sites/PoliticalTransparency/Shared%20Documents/Forms/AllItems.aspx?ga=1&id=%2Fsites%2FPoliticalTransparency%2FShared%20Documents%2FPolitical%20Transparency%20Reports&viewid=e3b6c62f%2Dd2c0%2D4913%2Da962%2De55c0fc5965c.