Innovative Assignment 2

Gahan Saraiya (18MCEC10), Rushi Trivedi (18MCEC08), Raj Kothari (18MCEC07)

18mcec10@nirmauni.ac.in, 18mcec08@nirmauni.ac.in, 18mcec07@nirmauni.ac.in

I. Introduction

Aim of this assignment is to analyze impact of various indexes on any modern database (SQL or NoSQL).

Steps to be followed:

- Select Database
- Analyze performance and cost impact of indexes for various type of query

I. Hash Table inside PostgreSQL

Postgres[1] scans over all the records in one of the tables from the join and saves them in a hash table.

II. Implementation

I. Considerations

Variable	Value	Detail
Database	PostgreSQL	
Analyze scale factor	0.1	Number of tuple inserts, updates, or deletes prior to analyze as a fraction of reltuples.
Analyze Threshold	50	Minimum number of tuple inserts, updates, or deletes prior to analyze.
Block Size	8192	Shows the size of a disk block.
CPU Index Tuple cost	0.005	Sets the planner's estimate of the cost of processing each index entry during an index scan.
CPU operator cost	0.0025	Sets the planner's estimate of the cost of processing each operator or function call.

[1] referred as acronym to PostgreSQL

CPU tuple cost	0.01	Sets the planner's estimate of the cost of processing each tuple (row).
cursor tuple fraction	0.1	Sets the planner's estimate of the fraction of a cursor's rows that will be retrieved.
deadlock timeout	1s	Sets the time to wait on a lock before checking for deadlock.
commit delay	0μs	Sets the delay in microseconds between transaction commit and flushing WAL[2] to disk.
commit sibling	5μs	Sets the minimum concurrent open transactions before performing commit_delay.
Effective cache size	4GB	Sets the planner's assumption about the size of the disk cache.
Effective IO concurrency	1	Number of simultaneous requests that can be handled efficiently by the disk subsystem.
Log Rotation Age	1d	Automatic log file rotation will occur after N minutes.
Log Rotation Size	10MB	Automatic log file rotation will occur after N kilobytes.
Maintenance Work Memory	64MB	Sets the maximum memory to be used for maintenance operations.
Max Connections	100	Sets the maximum number of concurrent connections.
Max files per process	1000	Sets the maximum number of simultaneously open files for each server process.
Max function arguments	100	Shows the maximum number of function arguments.
Max Index Keys	32	Shows the maximum number of index keys.
Max Stack depth	2MB	Sets the maximum stack depth, in kilobytes.
Max WAL Size	1GB	Sets the WAL size that triggers a checkpoint.
Min WAL Size	80MB	Sets the minimum size to shrink the WAL to.
Segment Size	1GB	Shows the number of pages per disk file.

[2]WAL - Write Ahead Logging

Work Memory	4*MB*	Sets the maximum memory to be used for query workspaces.

Table 1: Consideration and variables

II. Schema

```
--
-- PostgreSQL database dump
--

-- Dumped from database version 9.5.14
-- Dumped by pg_dump version 10.3

SET statement_timeout = 0;
SET lock_timeout = 0;
SET idle_in_transaction_session_timeout = 0;
SET client_encoding = 'UTF8';
SET standard_conforming_strings = on;
SELECT pg_catalog.set_config('search_path', '', false);
SET check_function_bodies = false;
SET client_min_messages = warning;
SET row_security = off;

SET default_tablespace = '';

SET default_with_oids = false;

--
-- Name: collection_title; Type: TABLE; Schema: public; Owner: postgres
--

CREATE TABLE public.collection_title (
    id integer NOT NULL,
    title_id character varying(16) NOT NULL,
    ordering integer NOT NULL,
    title character varying(512) NOT NULL,
    region character varying(8),
    language character varying(16),
    types character varying(32),
    attributes character varying(256),
```

```
    description text,
    is_original_title boolean NOT NULL
);


ALTER TABLE public.collection_title OWNER TO postgres;

--
-- Name: collection_title_id_seq; Type: SEQUENCE; Schema: public; Owner:
 ↪  postgres
--

CREATE SEQUENCE public.collection_title_id_seq
    START WITH 1
    INCREMENT BY 1
    NO MINVALUE
    NO MAXVALUE
    CACHE 1;


ALTER TABLE public.collection_title_id_seq OWNER TO postgres;

--
-- Name: collection_title_id_seq; Type: SEQUENCE OWNED BY; Schema: public; Owner:
 ↪  postgres
--

ALTER SEQUENCE public.collection_title_id_seq OWNED BY
 ↪  public.collection_title.id;


--
-- Name: collection_title id; Type: DEFAULT; Schema: public; Owner: postgres
--

ALTER TABLE ONLY public.collection_title ALTER COLUMN id SET DEFAULT
 ↪  nextval('public.collection_title_id_seq'::regclass);


--
-- Name: collection_title collection_title_pkey; Type: CONSTRAINT; Schema:
 ↪  public; Owner: postgres
```

```
--

ALTER TABLE ONLY public.collection_title
    ADD CONSTRAINT collection_title_pkey PRIMARY KEY (id);


--
-- PostgreSQL database dump complete
--
```

III. Status

Total Entries 2096169

III. Query Impact

I. Exact Match/ Point Query

```
EXPLAIN ANALYZE select * from collection_title where title like '%Batman%' and
↪  title_id='tt0060153' and region='US';
```

Table 2: Query Plan

Index Scan using collection_title_pkey on collection_title	(cost=0.43..8.45 rows=1 width=107) (actual time=0.035..0.036 rows=1 loops=1)
Index Cond	$(id = 334596)$
Planning time	0.127ms
Execution time	0.078ms

II. Partial Match

```
explain analyze select * from collection_title where ordering=1;
```

Table 3: Query Plan

Seq Scan on collection_title	(cost=0.00..48832.51 rows=859800 width=107) (actual time=0.020..305.430 rows=910215 loops=1)
Filter	$(ordering = 1)$
Rows Removed by Filter	1185954
Total Rows	2096169
Planning time	0.091ms
Execution time	334.993ms

```
explain analyze select id,title_id, title from collection_title where title like
↪  '%avenger%';
```

Table 4: Query Plan

Seq Scan on collection_title	(cost=0.00..48832.51 rows=207 width=34) (actual time=0.621..490.246 rows=41 loops=1)
Filter	$((title) :: text \ \ '\%avenger\%' :: text)$
Rows Removed by Filter	1185954
Total Rows	2096128
Planning time	0.120 ms
Execution time	490.282 ms

III. Range Query

```
explain analyze select * from collection_title where id>3566 and id <25996;
```

Table 5: Query Plan

Seq Scan on collection_title	(cost=0.00..48832.51 rows=859800 width=107) (actual time=0.020..305.430 rows=910215 loops=1)
Filter	$(ordering = 1)$
Rows Removed by Filter	1185954
Total Rows	2096169
Planning time	0.091ms
Execution time	334.993ms

```
explain analyze select * from collection_title where id>3566 and id <25996;
```

Table 6: Query Plan

Index Scan using collection_title_pkey on collection_title	(cost=0.43..1089.81 rows=25864 width=107) (actual time=0.168..44.540 rows=22429 loops=1)
Index Cond	$((id > 3566)AND(id < 25996))$
Planning time	0.0140 ms
Execution time	45.695 ms