Otimização de Consultas

•••

Pedro Henrique Klein

Consula 1 - Lenta -> C1L

```
SELECT
    bookings.*
FROM
    flights,
    ticket_flights,
   tickets,
    bookings,
    airports
WHERE
    flights.flight id = ticket flights.flight id
    AND
ticket_flights.ticket_no = tickets.ticket_no
    AND
tickets.book_ref = bookings.book_ref
    AND
airports.airport_code = flights.departure_airport
    AND
airports.airport_code = 'BAX'
ORDER BY
    bookings.total_amount DESC
```

Plano de Consulta da C1L

1)
:1)
ws=8391852 loops=1)
pps=1)
loops=27666)
=1 loops=27666)

Consulta 1 - Otimizada -> C10

```
SELECT
     FROM
         BOOKINGS.AIRPORTS
     WHERE
         AIRPORT CODE = 'BAX'
 SELECT
     b.BOOK_REF,
     b.BOOK_DATE,
     b.TOTAL_AMOUNT
 FROM
     BOOKINGS.FLIGHTS AS f
 INNER JOIN BOOKINGS.TICKET FLIGHTS AS tf ON
     f.FLIGHT ID = tf.FLIGHT ID
 INNER JOIN TICKETS AS t ON
     tf.TICKET NO = t.TICKET NO
 INNER JOIN BOOKINGS BOOKINGS AS b ON
     t.BOOK_REF = b.BOOK_REF
 INNER JOIN airports_filter AS af ON
     f.DEPARTURE_AIRPORT = af.AIRPORT_CODE
 ORDER BY
     b. TOTAL_AMOUNT
 DESC;
```

Plano de Consulta da C10

ADC QUERY PLAN
Sort (cost=224126.38224229.88 rows=41399 width=21) (actual time=2619.9112620.951 rows=27666 loops=1)
Sort Key: b.total_amount DESC
Sort Method: quicksort Memory: 2498kB
-> Nested Loop (cost=5323.95220951.64 rows=41399 width=21) (actual time=29.1692601.867 rows=27666 kg
-> Seq Scan on airports_data ml (cost=0.004.30 rows=1 width=4) (actual time=0.0320.033 rows=1 loops
Filter: (airport_code = 'BAX'::bpchar)
Rows Removed by Filter: 103
-> Nested Loop (cost=5323.95220533.35 rows=41399 width=25) (actual time=29.1352597.952 rows=2766
-> Nested Loop (cost=5323.52200927.20 rows=41399 width=11) (actual time=29.1142205.768 rows=2
-> Hash Join (cost=5323.09181201.83 rows=41399 width=18) (actual time=29.0751850.995 rows=
Hash Cond: (tf.flight_id = f.flight_id)
-> Seq Scan on ticket_flights tf (cost=0.00153850.08 rows=8391708 width=18) (actual time=0.0
-> Hash (cost=5309.845309.84 rows=1060 width=8) (actual time=29.01129.012 rows=1017 loo
Buckets: 2048 Batches: 1 Memory Usage: 56kB
-> Seq Scan on flights f (cost=0.005309.84 rows=1060 width=8) (actual time=0.02528.898 i
Filter: (departure_airport = 'BAX'::bpchar)
Rows Removed by Filter: 213850
-> Index Scan using tickets_pkey on tickets t (cost=0.430.48 rows=1 width=21) (actual time=0.012
Index Cond: (ticket_no = tf.ticket_no)
-> Index Scan using bookings_pkey on bookings b (cost=0.430.47 rows=1 width=21) (actual time=0.0
Index Cond: (book_ref = t.book_ref)
Planning Time: 0.589 ms
Execution Time: 2621.764 ms

Consulta 2 - Lenta -> C2L

```
SELECT
      flights.flight_id,
      count(*) total_flight_id
 FROM
 JOIN ticket flights ON
      flights.flight_id = ticket_flights.flight_id
 WHERE
     flights.arrival airport IN (
          airport code
      FROM
         airports
          airport_name =
  'Barnaul Airport')
  flights.departure_airport IN (
         airport_code
      FROM
         airports
     WHERE
         airport name =
  'Barnaul Airport')
  GROUP BY
      flights.flight_id
 ORDER BY
      flights.flight_id
```

Plano de Consulta da C2L

OUERY PLAN Sort (cost=672121.86..672524.73 rows=161150 width=12) (actual time=3220.849..3220.895 rows=1127 loops=1) Sort Key: flights.flight id Sort Method: quicksort Memory: 110kB -> HashAggregate (cost=604647.87..655429.46 rows=161150 width=12) (actual time=3219.855..3220.629 rows=1127 loops=1) Group Key: flights.flight id Planned Partitions: 4 Batches: 1 Memory Usage: 3217kB -> Hash Join (cost=8552.50..250623.25 rows=6293771 width=4) (actual time=68.481..3196.100 rows=55357 loops=1) Hash Cond: (ticket flights.flight id = flights.flight id) -> Seg Scan on ticket flights (cost=0.00..153850.08 rows=8391708 width=4) (actual time=0.110..1230.594 rows=8391852 loops=1) -> Hash (cost=5908.13..5908.13 rows=161150 width=4) (actual time=67.328..67.330 rows=2035 loops=1) Buckets: 262144 Batches: 2 Memory Usage: 2084kB -> Seq Scan on flights (cost=61.13..5908.13 rows=161150 width=4) (actual time=0.862..66.323 rows=2035 loops=1) Filter: ((hashed SubPlan 1) OR (hashed SubPlan 2)) Rows Removed by Filter: 212832 SubPlan 1 -> Seq Scan on airports_data ml (cost=0.00..30.56 rows=1 width=4) (actual time=0.514..0.519 rows=1 loops=1) Filter: ((airport_name ->> lang()) = 'Barnaul Airport'::text) Rows Removed by Filter: 103 SubPlan 2 -> Seq Scan on airports_data ml_1 (cost=0.00..30.56 rows=1 width=4) (actual time=0.216..0.221 rows=1 loops=1) Filter: ((airport_name ->> lang()) = 'Barnaul Airport'::text) Rows Removed by Filter: 103 Planning Time: 0.459 ms Execution Time: 3224,326 ms

Consulta 2 - Otimizada -> C20

```
WITH airports_filter AS (
     SELECT
         AIRPORT_CODE
     FROM
         BOOKINGS.AIRPORTS
     WHERE
         AIRPORT_NAME = 'Barnaul Airport'
 SELECT
     f.FLIGHT_ID,
     COUNT(f.FLIGHT_ID)
 FROM
     BOOKINGS.FLIGHTS AS f
  INNER JOIN BOOKINGS.TICKET_FLIGHTS AS tf ON
     f.FLIGHT_ID = tf.FLIGHT ID
 WHERE
     EXISTS (SELECT * FROM airports_filter WHERE AIRPORT_CODE = f.ARRIVAL_AIRPORT OR AIRPORT_CODE = f.DEPARTURE_AIRPORT)
 GROUP BY
 ORDER BY
 ASC;
```

Plano de Consulta da C20

QUERY PLAN GroupAggregate (cost=211624.40..214434.95 rows=160603 width=12) (actual time=1840.985..1848.568 rows=1127 loops=1) Group Key: f.flight_id -> Sort (cost=211624.40..212025.91 rows=160603 width=4) (actual time=1840.925..1843.457 rows=55357 loops=1) Sort Key: f.flight_id Sort Method: quicksort Memory: 1537kB -> Hash Join (cost=8614.80..195539.75 rows=160603 width=4) (actual time=96.258..1830.206 rows=55357 loops=1) Hash Cond: (tf.flight id = f.flight id) -> Seq Scan on ticket_flights tf (cost=0.00..153850.08 rows=8391708 width=4) (actual time=0.119..957.799 rows=8391852 loops=1) -> Hash (cost=8563.40..8563.40 rows=4112 width=4) (actual time=95.858..95.860 rows=2035 loops=1) Buckets: 8192 Batches: 1 Memory Usage: 136kB -> Nested Loop Semi Join (cost=0.00..8563.40 rows=4112 width=4) (actual time=0.316..95.332 rows=2035 loops=1) Join Filter: ((ml.airport code = f.arrival airport) OR (ml.airport code = f.departure airport)) Rows Removed by Join Filter: 212832 -> Seg Scan on flights f (cost=0.00..4772.67 rows=214867 width=12) (actual time=0.009..18.710 rows=214867 loops=1) -> Materialize (cost=0.00..30.57 rows=1 width=4) (actual time=0.000..0.000 rows=1 loops=214867) -> Seq Scan on airports_data ml (cost=0.00..30.56 rows=1 width=4) (actual time=0.240..0.244 rows=1 loops=1) Filter: ((airport_name ->> lang()) = 'Barnaul Airport'::text) Rows Removed by Filter: 103 Planning Time: 0.422 ms Execution Time: 1848,839 ms

Consulta 3 - Lenta -> C3L

```
⊜ SELECT
     bookings.book_ref,
     count(*)
 FROM
     bookings,
     tickets,
     ticket_flights,
     flights
 WHERE
     tickets.ticket_no = ticket_flights.ticket_no
     AND
 flights.status = 'Scheduled'
     AND
 ticket_flights.flight_id = flights.flight_id
 bookings.book_ref = tickets.book_ref
 GROUP BY
     bookings.book ref
 HAVING
     count(*) > 1
 ORDER BY
     bookings.book_ref;
```

Plano de Consulta da C3L

Studie Graphagregier (cor. 27037603-34119006 rose: 196161 sloth (15) (schuldrine: 221696) 2619 746 rose: 113325 (cops: 1)
Group May bearings book, ref
Risk (pount) > t)
Rose Removed by Steer 19070
-> Garber Werge (post (27037636) 331366 99 rount (9115) width (15) (accounting (2316.95) 2569 356 rount (32995 Scopert))
Warlers Planned 2
Waters Laundred 2
-> Partial Groupinggregate (cost (2001)) 01 271675 60 costs (245577 (42th (15) (actual rine (225) 69) 2364 055 costs (413)2 (copt (1)
Group Kry bookings book, ref
-> Sort (0001/309178.01-20999) 95 cosx (245577 width /7) (actualtities (2259.600.2015/601.cosx (146799.logo) (2)
Som Kiry bookings back, yet
Sort Northad external merge Didd 191380
Worker P. Som Nethod: exernal merge: Didd 3810NB
Worker I: Sort Method: executal merge: Didt 271258
-> Parallel Hush Join (corr 2006/32 Sti 241031 72 roses (245577 witch /1) (semaltrine +1604 947 1776 752 roses (46766 loops /3)
Haids Cond (baddings badd, ref = tidees badd, ref)
-> Parallel Ség Scan on bookings (cost :0 00 22213 29 rows (879639 width :17) (Actual trine :0 009 65 555 rows (705703 (acquii 3)
-> Panallet Hash (cost: 205602 87 205602 87 zous: 245577 width: (7) (actual time: 1815 724: 1815 728 zous: 146798 (acque: 3))
Budent 363144 Sundiec 4 Menory Usage 6464M
 Parallel Hach Join (cost +122661 61 205602 87 nove (245577 width +17) (accusitive +1158 400 1404 643 nove +148768 loops +8)
Hash Cond (rideo) rideo, no i rideo, Rights rideo, no)
 Parallel Seq Scan on riders (post i0:00-61705 (8 rows (2390)6 width (21) (amustrine i0:210-205 975 rows (601266 loops (1))
-> Parallel Hach (com: ITREPELS) 193 118361 93 move 245577 width (14) (amuel draw ITAT 065 TAT 066 move (140T66 loops (1))
Budeto: NG 144 Bandwo B Menory Sough 47045B
-> Parallel Hach Join (com H 514 67 11639) 63 nove-245577 width H 4) (comultine H 1 279 T01 013 nove-146766 loops II)
Hauth Cond: (elder, flights flight; flights flight; id)
 Parallel Seq Scan on tider, flights (cost: 0.00-104696 45 rows: 3496545 width: 10) (actual time: 0.222-305772 rows: 2797284 loops: 3)
> Parallel Hash (cost +4303 90 4203 90 coss (8877 width (d) (Jeanual line +11 071 11 071 raws (5128 loops (3))
Budent (GRM Batchet I Nemony Loop TRAM
>> Paradel Seq Scan on flights (cost ±0 00 ±205 90 rows ±00TT width ±1) (actual time ±0 01 ± 29 (T4 rows ±1516) (acquait 1)
Rec ((status) (test i Schedulet test)
Rows Removed by Street 1994 64
Planning Time: 0 SSH ris.

Consulta 3 - Otimizada -> C30

```
WITH flights_filter AS (
     SELECT
         FLIGHT ID
     FROM
         BOOKINGS.FLIGHTS
     WHERE
         STATUS = 'Scheduled'
 SELECT
     b.BOOK_REF,
     COUNT(b.BOOK_REF)
 FROM
     BOOKINGS BOOKINGS AS b
 INNER JOIN BOOKINGS.TICKETS AS t ON
     b.BOOK_REF = t.BOOK_REF
 INNER JOIN BOOKINGS.TICKET FLIGHTS AS tf ON
     t.TICKET NO = tf.TICKET NO
 INNER JOIN flights_filter AS ff ON
     tf.FLIGHT ID = ff.FLIGHT ID
 GROUP BY
 HAVING
     COUNT(b.BOOK_REF) > 1
 ORDER BY
 ASC;
```

Plano de Consulta da C30

Ruike Grouphgregute (met :270176.03 341190.06 rows:196161 width:15) (actual time :219/.216.2556.259 may:111325 (acpost))	
Group King black pref	
River (passets) black pet() > 1)	
Rases Removed by Renn 19670	
-> Gather Nerge (cost (270176 03: 331366 99 rows (49115) witch (15) (actual time (2191206: 2491 692 rows (13295 (sope (1)	
Workers Planned: 2	
Worlers (aunithed 2	
-> Partial GroupAggregate (cost 1569-178 01 1716/75 60 rows 04/5577 width +15) (acquaintee +2137 109 2241 910 rows +61512 loops +1)	
Group Key h book; ef	
-> Sort (cost (2004)TB 01: 200901 95 roses (215577 width (7) (actual time (2127 091: 221 2395 roses (14970b) loops (1)	
Sort Key biback, ref	
Sort Method external merge Disk 191258	
Worker 0: Som Nethod: external merge-Didd: 272069	
Worker I: Som Nethod: exercial recrypt Didd: 201088	
-> Parallet Hash John (cost 0.09672.56. 244031.72 rows (245577 width (7) (actual time (1526.657 1684.59) rows (146766 book (3)	
Hash Cond (b book) of a book of	
-> Parallet Seq Scan on bookings b. (cod.: 0.00. 2224) 29 rows: 6176279 width: 7) (actual time : 0.141 61 666 rows: (101707) loops: 3)	
Parallel Harb (corr :265602 67 205602 67 core :245577 width (7) (actual time :1355 666 1355 69) (core :146766 (cope :3)	
Sudect 262144 Sundect Memorylange 6464M	
-> Parallel Haidh John (cont i 12266) 64 205602 67 rows (2/5577 width i T) (actual time i 1062 637 1121 429 rows (146766 loops i 1)	
Histor Condition (no intrinse year)	
-> Parallel Seg Scan on tiderest (coat i0 00 G1705 18 rows: 1235018 width (21) (actual time i0 H16 177 952 rows: 961286 loops: il)	
-> Parallel Hach (com: 116391-93 -116391-93 roses 245577 width +14) (actual time +703-367-703-370 roses +146768 boops +3)	
Budenc 262144 Bunther & Mensury Garge 470466	
> Parallel Haich John (cost = 814 6T 11639) 93 roses (245577 width = 14) (actual time = 10 949 66) 052 roses (146766 loops = 1)	
Hash Cond of flight jis : flights flight jis)	
-> Parallel Seq Scan on tides (fightest) (cost: 0.00-101696 45 rows: 3196545 with: 18) (actual time: 0.227-286 616 rows: 2797284 loc	pe:3)
-> Parallel Hadh (post #200 90 +200 90 rows #877 width #) (actual time #10 624 10 624 rows 5128 loops #)	
Budesc (G16) Sunfect Menorylange T1680	
-> ParalletSeqScan on flights (cost 1000 1203 60 rows (66TT width (4) (armalitine (0.012 27 566 rows (1536)) loops: 1)	
Rest ((manu(mos : Scheduledmos)	
Row Removed by Ref. 199464	
Narraing Time: 0 fit it rec	

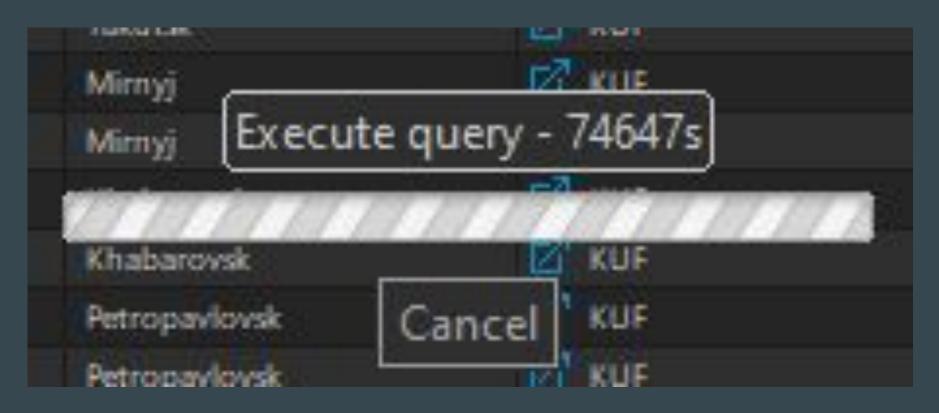
Consulta 4 - Lenta -> C4L

```
■ SELECT

     f2.flight no,
     f2.departure airport,
     dep.airport name AS departure airport name,
     dep.city AS departure city,
     f2.arrival airport,
     arr.airport name AS arrival airport name,
     arr.city AS arrival city,
     f2.aircraft code,
     (SELECT avg(flights.scheduled arrival - flights.scheduled departure)
     FROM bookings.flights WHERE flight no = f2.flight no LIMIT 1) duration,
     (SELECT array agg(DISTINCT to char(flights.scheduled departure, 'ID'::text)::integer)
     FROM bookings.flights WHERE flight no = f2.flight no LIMIT 1) days of week
 FROM
     bookings.flights f2,
     bookings.airports dep,
     bookings.airports arr
 WHERE
     f2.departure airport = dep.airport code AND f2.arrival airport = arr.airport code
     AND dep.city LIKE '%Petro%' OR arr.city LIKE '%Petro%'
```

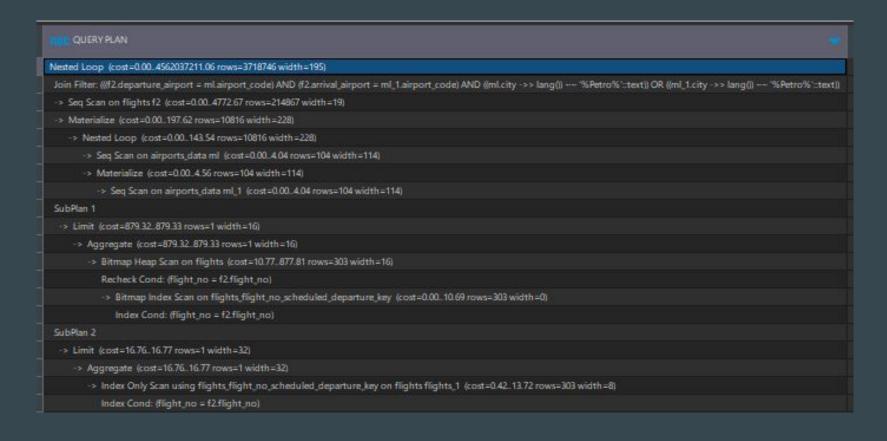
Plano de Consulta da C4L

Acabou não finalizando o Explain Analyse



Plano de Consulta da C4L

Mostrando apenas com o explain



Consulta 4 - Otimizada-> C40

```
SELECT.
         flight no.
         ARRAY AGG(DISTINCT to char(f2.scheduled_departure, 'ID'::text)::integer) AS days_of_week
      FROM
         bookings.flights f2
      GROUP BY
 SELECT
     f.FLIGHT ID,
     f.flight no,
     f.departure airport,
     a.airport name AS departure airport name.
     a.city AS departure city,
     f.arrival airport,
     a.airport name AS arrival airport name,
     a.city AS arrival city,
     f.aircraft code.
     f.scheduled arrival - f.scheduled departure AS duration,
     h.days of week
 FROM
     bookings.flights AS f
 INNER JOIN bookings.airports AS a ON
     f.DEPARTURE AIRPORT = a.AIRPORT CODE
     OR f.ARRIVAL AIRPORT = a.AIRPORT CODE
 INNER JOIN hub AS h ON
     f.flight no = h.flight no
 WHERE
     a.CITY LIKE '%Petro%';
```

Plano de Consulta da C40

QUERY PLAN Hash Join (cost=8078.06.21542.93 rows=4112 width=199) (actual time=131.175.329.552 rows=4979 loops=1) Hash Cond: (f2.flight_no = f.flight_no) GroupAggregate (cost=0.42-9242.14 rows=710 width=39) (actual time=1.276.175.071 rows=710 loops=1) Group Key: f2:flight no -> Index Only Scan using flights_flight_no_scheduled_departure_key on flights f2 (cost=0.42.6547.43 rows=214867 width=15) (actual time=0.041..33.170 rows=214867 loops=1) Heap Fetches: 0 -> Hash (cost=8026.24_8026.24 rows=4112 width=149) (actual time=115.379_115.381 rows=4979 loops=1) Buckets 8192 Batches 1 Memory Usage: 1005kB Nested Loop (cost=0.00.8026.24 rows=4112 width=149) (actual time=1.980_113.583 rows=4979 loops=1) Join Filter: (f.departure_airport = ml.airport_code) OR (f.arrival_airport = ml.airport_code)). Rows Removed by Join Filter: 424755 -> Seg Scan on airports_data_ml_(cost=0.00.30.56 rows=1 width=114) (actual time=0.218.0.477 rows=2 loops=1) Filter: ((city ->> lang()) --- "%Petro%":text) Rows Removed by Filter: 102 Seg Scan on flights f (cost=0.00, 4772.67 rows=214867 width=39) (actual time=0.004, 22.310 rows=214867 loops=2) Planning Time: 1.073 ms Execution Time: 333,358 ms

Ganhos em Tempo de Execução (MS)

Consulta	Tempo Consulta Normal	Tempo Consulta Otimizada	Diferença
1	2945	2620	325
2	3220	1848	1372
3	2649	2556	93
4	-	329	-

Ganhos em Custo de Execução

Consulta	Custo Consulta Normal	Custo Consulta Otimizada	Diferença
1	224.229	224.229	0
2	672.524	214.434	458.090
3	241.190	241.190	0
4	4.562.037.211	21.542	4.562.015.669

Entendendo a pouca diferença entre as consultas 1 e 3

Para poder otimizar ainda mais as consulta 1 e 3, precisaríamos criar índices nas chaves onde ocorrem os JOINS, porém optei por não modificar a "Estrutura" do banco, mantendo assim sua forma original.

Tomei esta decisão pois, as consultas melhoraram significativamente em questão de tempo de execução.

Link para Consultas

• Consultas Normais: <u>LINK</u>

• Consultas Otimizadas: <u>LINK</u>