DISTRIBUTED AND PARALLEL DATABASE SYSTEMS
Project Part 6
Query Optimization

Query 1

The books data of each site is in their respective sites. The scenario for the below is that a user wants to check the availability of fiction books in both USA and France.

For this scenario, the below query will fetch the books information from USA site and books information form France site. The data is then filtered for genre Fiction.

Let's calculate the cost of the below query.

Books table of USA have 100 book records.
Books table of France have 100 book records.
Transferring one record from one site to another will cost 10 units.
Selecting one record within the site will cost 1 unit.

Cost involved in transferring data from books table of USA site to site 1	number of records in USA books table * transferring cost	100 * 10 = 1000
Cost involved in transferring data from books table of France site to site 1	number of records in France books table * transferring cost	100 * 10 = 1000
Both the tables are joined using union		
Cost involved in selecting the fiction books from the new table which is union of USA and France books	Total fiction books in union table * selecting cost	22 * 1 = 22
	Total cost of operation	2022

Relation Algebra Query:

$$\sigma_{genre='fiction'} (\pi_{name, isbn, genre, l\_id} (vishnu.books\_usa) \cup \pi_{name, isbn, genre, l\_id} (sharanya.books\_france))$$

```
select * from (select name,isbn,genre,l_id from vishnu.books_USA
               union
               select name,isbn,genre,l_id from sharanya.books_France )
               temp where temp.genre='Fiction';
```

Query Result | Script Output | DBMS Output | Explain Plan | Autotrace | SQL History

Download | Execution time: 0.005 seconds

	NAME	ISBN	GENRE	L_ID
1	1984	a170	Fiction	11
2	1984	a210	Fiction	21
3	Beloved	a179	Fiction	14
4	Beloved	a219	Fiction	24
5	The Alchemist	a185	Fiction	16
6	The Alchemist	a225	Fiction	26
7	The Catcher in the Ry	a172	Fiction	11
8	The Catcher in the Ry	a212	Fiction	21
9	The Color Purple	a182	Fiction	15
10	The Color Purple	a222	Fiction	25
11	The God of Small Thir	a184	Fiction	15
12	The God of Small Thir	a224	Fiction	25
13	The Great Gatsby	a174	Fiction	12
14	The Great Gatsby	a214	Fiction	22
15	The Handmaids Tale	a181	Fiction	14
16	The Handmaids Tale	a221	Fiction	24
17	The Hitchhikers Guide	a180	Fiction	14
18	The Hitchhikers Guide	a220	Fiction	24
19	The Hunger Games	a176	Fiction	13
20	The Hunger Games	a216	Fiction	23
21	The Road	a189	Fiction	17
22	The Road	a229	Fiction	27

Alternative approach:

Selecting the books with genre fiction from each table and then transferring them to another site.

Cost involved in selecting books with genre fiction in USA site	number of records in USA books table with genre fiction * selecting cost	11 * 1 = 11
Cost involved in selecting books with genre fiction in France site	number of records in France books table with genre fiction * selecting cost	11 * 1 = 11
Cost involved in transferring the selected books to site 1	Total fiction books in both the tables * transferring cost	22 * 10 = 220
Both the tables are joined using union in site 1		
Cost involved in selecting the books from union of tables	Total books in union table * selecting cost	22 * 1 = 22
	Total cost of operation	264

```
8   select name,isbn,genre,l_id from vishnu.books_USA where genre = 'Fiction'
9       union
10      select name,isbn,genre,l_id from sharanya.books_France where genre='Fiction';
```

Query Result | Script Output | DBMS Output | Explain Plan | Autotrace | SQL History

Download ▾ Execution time: 0.003 seconds

	NAME	ISBN	GENRE	L_ID
1	1984	a170	Fiction	11
2	1984	a210	Fiction	21
3	Beloved	a179	Fiction	14
4	Beloved	a219	Fiction	24
5	The Alchemist	a185	Fiction	16
6	The Alchemist	a225	Fiction	26
7	The Catcher in the Ry	a172	Fiction	11
8	The Catcher in the Ry	a212	Fiction	21
9	The Color Purple	a182	Fiction	15
10	The Color Purple	a222	Fiction	25
11	The God of Small Thir	a184	Fiction	15
12	The God of Small Thir	a224	Fiction	25
13	The Great Gatsby	a174	Fiction	12
14	The Great Gatsby	a214	Fiction	22
15	The Handmaids Tale	a181	Fiction	14
16	The Handmaids Tale	a221	Fiction	24
17	The Hitchhikers Guide	a180	Fiction	14
18	The Hitchhikers Guide	a220	Fiction	24
19	The Hunger Games	a176	Fiction	13
20	The Hunger Games	a216	Fiction	23
21	The Road	a189	Fiction	17
22	The Road	a229	Fiction	27

Relation Algebra for the alternative approach:

$$\pi_{\text{name, isbn, genre, l\_id}} \Big(\sigma_{\text{genre='fiction'}}(\text{vishnu.books\_usa}) \cup \sigma_{\text{genre='fiction'}}(\text{sharanya.books\_france}) \Big)$$

From the cost of both the strategies which are 2022 and 264 and from the execution time of both the queries which is 0.005s for the un -optimized query and 0.003s for optimized query, we can conclude that the alternative strategy is the optimized version of the first query.

Query 2

The employee data of each site is in their respective sites. The scenario for the below is that I (admin of site India) wanted to display the name of the employees who are managers from India and USA.

For this scenario, the below query will fetch the employee's information from USA site and employees information from India site. The data is then filtered for role Manager.

Let's calculate the cost of the below query.

Employees table of USA have 20 employee records.
Employees table of India have 20 employee records.
Transferring one record from one site to another will cost 10 units.
Selecting one record within the site will cost 1 unit.

Cost involved in transferring data from employees table of USA site to site 1	number of records in USA employees table * transferring cost	20 * 10 = 200
Cost involved in transferring data from employees table of India site to site 1	number of records in India books table * transferring cost	20 * 10 = 200
Both the tables are joined using union		
Cost involved in selecting the Managers from the new table which is union of USA and India Employees	Total number of managers in union table * selecting cost	2 * 1 = 2
	Total cost of operation	402

Relation Algebra Query:

$$\pi_{\text{name}}\left(\sigma_{\text{role='Manager'}}(\text{sreekar.employees\_India}) \cup (\text{vishnu.employees\_USA})\right)$$

```
1  select name from (select * from employees_India
2                    union
3                    select * from vishnu.employees_USA)
4                    temp where temp.role='Manager';
5
6
```

Query Result | Script Output | DBMS Output | Explain Plan | Autotrace | SQL Hist

Download | Execution time: 0.006 seconds

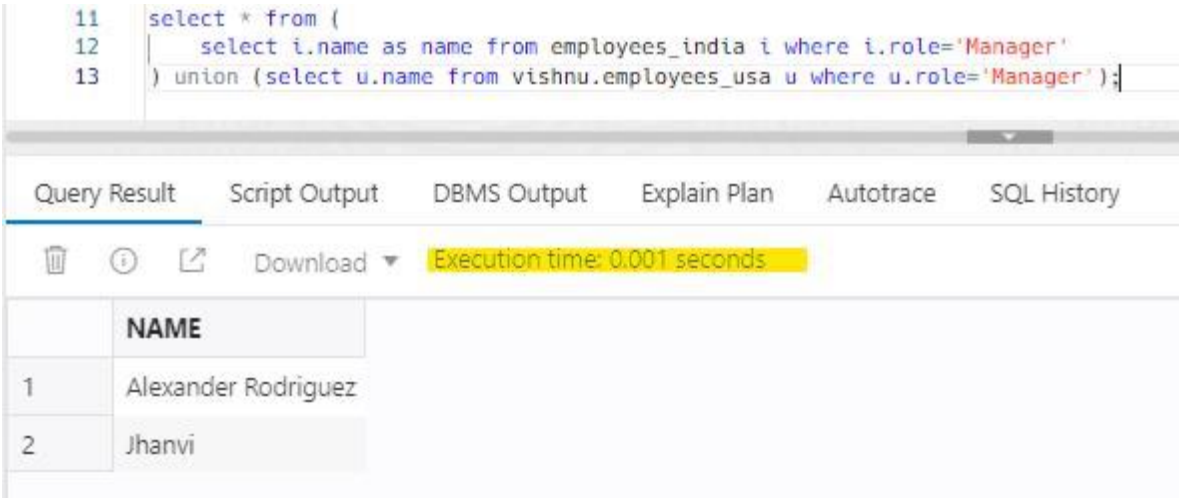
	NAME
1	Jhanvi
2	Alexander Rodriguez

Alternative approach:
Selecting the employees with role manager from each table and then transferring them to another site.

Cost involved in selecting employee with role manager in USA site	number of records in USA employees table with role manager * selecting cost	1 * 10 = 10
Cost involved in selecting employee with role manager in India site	number of records in India employees table with role manager * selecting cost	1 * 10 = 10
Cost involved in transferring the selected employees to site 1	Total managers both the tables * transferring cost	2 * 10 = 20
Both the tables are joined using union in site 1		
Cost involved in selecting the managers from union of tables	Total managers in union table * selecting cost	2 * 1 = 2
	Total cost of operation	42

Relation Algebra Query for alternative approach:

$$\pi_{name}(\sigma_{role='Manager'}(employees\_India)) \cup \pi_{name}(\sigma_{role='Manager'}(vishnu.employees\_USA))$$



From the cost of both the strategies which are 402 and 42 and from the execution time of both the queries which is 0.006s for the un-optimized query and 0.01s for optimized query, we can conclude that the alternative strategy is the optimized version of the first query.

Query 3

The donations data of each site is in their respective sites. The scenario for the below is that a user wants to get the names of the users who donated more than 500 dollars from France and Canada.

For this scenario, the query in the image will fetch the donors name and donated amount information from Canada site and the donors name and donated amount information from France site on to France site. The data is then filtered for the donated amount more than 500.

Let's calculate the cost of the below query.

The donations table of France have 100 donation records.
The donations table of Canada have 100 donation records.
Transferring one record from one site to another will cost 10 units.
Selecting one record within the site will cost 1 unit.

Cost involved in transferring data from donations and users table of Canada site to France site	number of records of donations * transferring cost	100 * 10 = 1000
Cost involved in selecting donations data of France site	number of records of donations * selecting cost	100 * 1 = 100
Both the tables are joined using union		
Cost involved in selecting the donations over 500 from the new table which is union of Canada and France donations	Total donors with donations over 500 in union table * selecting cost	12 * 1 = 12
	Total cost of operation	1112

Relation Algebra Query:

$$\pi_{\text{uname, amount}} (\sigma_{\text{amount}>500} (\pi(\text{donations\_france d} \bowtie_{\text{d.u\_id=f.u\_id}} \text{users\_france f}) \cup \pi(\text{lavnaya.donations\_canada d} \bowtie_{\text{d.u\_id=f.u\_id}} \text{Lavanya.doations\_canada f}))$$

```
11  select u_name,amount from
12  (
13  select * from donations_France d, users_France u where d.u_id=u.u_id
14  union
15  select * from lavanya.donations_Canada d, lavanya.users_Canada u where d.u_id=u.u_id)
16  temp where temp.amount>500;
```

Query Result | Script Output | DBMS Output | Explain Plan | Autotrace | SQL History

Download | Execution time: 0.011 seconds

	U_NAME	AMOUNT
1	Adriano	903
2	Alex	659
3	Amye	747
4	Andrey	956
5	Aymer	580
6	Aymer	717
7	Celestyn	696
8	Christin	823
9	Christin	882
10	Clemmie	750

Alternative approach:

Selecting the donors with donations over 500 from each site and then transferring them to France site.

Cost involved in selecting data from donations and users table of Canada site where donation amount > 500	number of records in USA books table with genre fiction * selecting cost	6 * 1 = 6
Cost involved in transferring selected records to France site	number of records with donations over 500 * transfering cost	6 * 10 = 60
Cost involved in selecting the selected donors with donations over 500 in France site	Total donor with donations over 500 * selecting cost	6 * 1 = 6
Both the tables are joined using union in site 1		
Cost involved in selecting the donors from union of tables	Total donors in union table * selecting cost	12 * 1 = 12
	Total cost of operation	84

Relation Algebra for the alternative approach:

$$\pi_{u\_name,\ amount}(\sigma_{amount>500}(donations\_france\ d \bowtie_{d.u\_id=f.u\_id} users\_france\ f))$$

$$\cup$$

$$\pi_{u\_name,\ amount}(\sigma_{amount>500}(lavnaya.donations\_canada\ d \bowtie_{d.u\_id=f.u\_id} Lavanya.doations\_canada\ f))$$

```
1  select * from (select u.u_name,d.amount
2   from donations_France d, users_France u
3   where d.u_id= u.u_id and d.amount>500)
4  union
5  (select u.u_name,d.amount
6   from lavanya.donations_Canada d, lavanya.users_Canada u
7   where d.u_id= u.u_id and d.amount>500);
```

Query Result | Script Output | DBMS Output | Explain Plan | Autotrace | SQL History

Download ▾ Execution time: 0.004 seconds

	U_NAME	AMOUNT
1	Adriano	903
2	Alex	659
3	Amye	747
4	Andrey	956
5	Aymer	580
6	Aymer	717
7	Celestyn	696
8	Christin	823
9	Christin	882
10	Clemmie	750

From the cost of both the strategies which are 1112 and 84 and from the execution time of both the queries which is 0.011s for the un -optimized query and 0.004s for optimized query, we can conclude that the alternative strategy is the optimized version of the first query.

Query 4

The cabins availability data of each site is in their respective sites. The scenario for the below is that a user wants to know library names that have unavailable cabins in USA and Canada.

For this scenario, the query in the image will fetch the library data and cabin availability information from USA site and the library data and cabin availability information from Canada site on to Canada site. The data is then filtered for the library names with unavailable cabins.

Let's calculate the cost of the below query.

The join of library table and cabin table of USA have 60 records.
The join of library table and cabin table of Canada have 60 records.
Transferring one record from one site to another will cost 10 units.
Selecting one record within the site will cost 1 unit.

Cost involved in transferring data from the join of library and cabins table of USA site to Canada site	number of records of library cabin information * transferring cost	60 * 10 = 600
Cost involved in selecting data from the join of library and cabins table of Canada site	number of records of library cabin information * selecting cost	60 * 1 = 60
Both the tables are joined using union		
Cost involved in selecting the library names and country information with cabins unavailable from the union of tables	Number of records with cabins unavailable in the union table * selecting cost	5 * 1 = 5
	Total cost of operation	665

Relation Algebra Query:

$$\pi_{name,\ country}\ (\sigma_{status='unavailable'}\ (\pi_{l.l\_id,\ l.name,\ l.country,\ c.status}(vishnu.library\_USA\ l \bowtie_{d.u\_id=f.u\_id} vishnu.cabins\_USA\ c)) \cup \pi_{l.l\_id,\ l.name,\ l.country,\ c.status}(library\_Canada\ l \bowtie_{l.l\_id=c.l\_id} cabins\_canada\ c))$$

```
select name,country from (
    select l.*,c.availability from library_Canada l, cabin_Canada c
    where c.l_id=l.l_id
    union
    select l.*,c.availability from vishnu.library_USA l, vishnu.cabin_USA c
    where c.l_id=l.l_id
    ) temp where temp.availability='unavailable';
```

Query Result | Script Output | DBMS Output | Explain Plan | Autotrace | SQL History

Download | Execution time: 0.003 seconds

	NAME	COUNTRY
1	Library Oasis	USA
2	The Bookery	Canada
3	The Book Barracks	Canada
4	Book Boulevard	Canada
5	The Book Brigade	Canada

Alternative approach:

Selecting the library name and country with cabins status unavailable from each site and then transferring them to Canada site.

Cost involved in selecting data from the join of library and cabins table of Canada site where cabin status is unavailable	number of records in Canada cabins with unavailable table * selecting cost	4 * 1 = 4
Cost involved in transferring data from the join of library and cabins table of USA site where cabin status is unavailable	number of records in USA cabins with unavailable table * transferring cost	1 * 10 = 10
Both the tables are joined using union in site 1		
Cost involved in selecting the library name and country of union of the tables	Total records in union table * selecting cost	5 * 1 = 5
	Total cost of operation	19

Relation Algebra for the alternative approach:

$$\pi_{l.name,\ l.country,}\left(\sigma_{status='unavailable'} \left(vishnu.library\_USA\ l \bowtie_{d.u\_id=f.u\_id} vishnu.cabins\_USA\ c \right) \right)$$

$$\cup$$

$$\pi_{l.name,\ l.country,}\left(\sigma_{status='unavailable'} \left(library\_Canada\ l \bowtie_{l.l\_id=c.l\_id} cabins\_canada\ c \right) \right)$$

```
1  select * from (
2  select name,country from library_Canada l, cabin_Canada c
3   where c.l_id=l.l_id and availability='unavailable'
4  union
5  select name,country from vishnu.library_USA l, vishnu.cabin_USA c
6   where c.l_id=l.l_id and availability='unavailable');
```

Query Result | Script Output | DBMS Output | Explain Plan | Autotrace | SQL H

Download ▾ Execution time: 0.001 seconds

	NAME	COUNTRY
1	Book Boulevard	Canada
2	Library Oasis	USA
3	The Book Barracks	Canada
4	The Book Brigade	Canada
5	The Bookery	Canada

From the cost of both the strategies which are 665 and 19 and from the execution time of both the queries which is 0.003s for the un -optimized query and 0.001s for optimized query, we can conclude that the alternative strategy is the optimized version of the first query.

Query 5

The books data of each site are in their respective sites. The scenario for the below is that a user wants to know names of un-returned books from USA and England.

For this scenario, the query in the image will fetch the book transactions of USA site and the book transactions of England site on to England site. The data is then filtered for the book names with return date null.

Let's calculate the cost of the below query.

The join of books table and book transactions table of USA have 100 records.
The join of books table and book transactions table of England have 100 records.
Transferring one record from one site to another will cost 10 units.
Selecting one record within the site will cost 1 unit.

Cost involved in transferring data from the join of books and book transactions table of USA site to England site	number of records of book transactions * transferring cost	100 * 10 = 1000
Cost involved in selecting data from the join of books and book transactions table of England site	number of records of book transactions * selecting cost	100 * 1 = 100
Both the tables are joined using union		
Cost involved in selecting the book names with returned date null, from the union of tables	Number of records return date null from the union of tables * selecting cost	6 * 1 = 6
	Total cost of operation	1106

Relation Algebra Query:

$$\pi_{b.name} (\sigma_{bt.returned\_date='null'} (\pi_{b.*,\ bt.returned\_date,} (vishnu.books\_USA\ b \bowtie_{b.b\_id=bt.u\_id} vishnu.book\_transactions\_USA\ bt)) \cup \pi_{b.*,\ bt.returned\_date,} (books\_england\ b \bowtie_{b.b\_id=bt.u\_id} book\_transactions\_england\ bt)))$$

```
11  select name from (
12      SELECT b.*,t.returned_date
13      FROM books_England b, book_transactions_England t
14      WHERE t.b_id = b.b_id
15      union
16      SELECT b.*,t.returned_date
17      FROM vishnu.books_USA b, vishnu.book_transactions_USA t
18      WHERE t.b_id = b.b_id
19  )temp where temp.returned_date is null;
20
```

Query Result | Script Output | DBMS Output | Explain Plan | Autotrace

Download ▾ Execution time: 0.008 seconds

	NAME
1	1984
2	The Hunger Games
3	The Diary of Anne Fra
4	Beloved
5	The Hitchhikers Guide
6	The Handmaids Tale

Alternative approach:

Selecting the book name with return date null from each site and then transferring them to England site.

Cost involved in transferring books with return data null in USA site	number of records in USA with books return date null * transferring cost	3 * 10 = 30
Cost involved in selecting books with return data null in England site	number of records in USA with books return date null * selecting cost	3 * 1 = 3
Both the tables are joined using union in site 1		
Cost involved in selecting the book name from union of the tables	Total records in union table * selecting cost	6 * 1 = 6
	Total cost of operation	39

Relation Algebra for the alternative approach:

$$\pi_{b.name}(\sigma_{bt.returned\_date='null'}($$

$$vishnu.books\_USA\ b \bowtie_{b.b\_id=bt.u\_id} vishnu.book\_transactions\_USA\ bt$$

$$))$$

$$\cup$$

$$\pi_{b.name}(\sigma_{bt.returned\_date='null'}($$

$$books\_england\ b \bowtie_{b.b\_id=bt.u\_id} book\_transactions\_england\ bt$$

$$))$$

```
1  select * from(
2      SELECT b.name
3      FROM books_England b, book_transactions_England t
4      WHERE t.b_id = b.b_id and t.returned_date IS NULL
5      union
6      SELECT b.name
7      FROM vishnu.books_USA b, vishnu.book_transactions_USA t
8      WHERE t.b_id = b.b_id and t.returned_date IS NULL);
9
```

Query Result | Script Output | DBMS Output | Explain Plan | Autotrace

Download ▾ Execution time: 0.002 seconds

	NAME
1	1984
2	The Hunger Games
3	The Diary of Anne Frank
4	Beloved
5	The Hitchhikers Guide to the Galaxy
6	The Handmaids Tale

From the cost of both the strategies which are 1106 and 39 and from the execution time of both the queries which is 0.008s for the un -optimized query and 0.002s for optimized query, we can conclude that the alternative strategy is the optimized version of the first query.

Individual Contribution:

- I have understood the concept of optimization and have created a scenario (**QUERY 2**) of displaying the employees who are managers from site India and Site USA.
- The query can be written in many ways, I have used a general query in which the employees table from site India and site USA are merged using union and then that merged table is filtered to get the managers. In this process, the time taking is more and cost is also high since the tables are being merged and a new table is creating. Hence, this is not an optimized query.
- So, the optimized solution can be as follows, filtered both employees table who are managers in site India and site USA which gives few tuples and then this is merged. Hence only a few couple of rows will be displayed. In this process the cost will also be very little and the time for execution will also be less. Hence, this can be an optimized version.