

## Task 3.4

1.

Original:

```
SELECT *
```

```
FROM film
```

**Explained:** "Seq Scan on film (cost=0.00..64.00 rows=1000 width=388)"

Revised:

```
SELECT film_id, title
```

```
FROM film
```

**Explained:** "Seq Scan on film (cost=0.00..64.00 rows=1000 width=19)"

They both had the same cost when ran but the revised version is more effective because we don't extract the information that we don't need also in the real world it took less time to run the revised query than the original

2.








Query		Query History	
1	SELECT	title, release_year, rental_rate	
2	FROM	film	
3	ORDER BY	title, release_year, rental_rate	ASC
Data output		Messages	Notifications
<div><div></div><div></div><div></div><div></div><div></div><div></div><div></div></div>			
	title	release_year	rental_rate
	character varying (255)	integer	numeric (4,2)
1	Academy Dinosaur	2006	0.99
2	Ace Goldfinger	2006	4.99
3	Adaptation Holes	2006	2.99
4	Affair Prejudice	2006	2.99
5	African Egg	2006	2.99
6	Agent Truman	2006	2.99
7	Airplane Sierra	2006	4.99
8	Airport Pollock	2006	4.99
9	Alabama Devil	2006	2.99
10	Aladdin Calendar	2006	4.99
11	Alamo Videotape	2006	0.99
12	Alaska Phantom	2006	0.99
13	Ali Forever	2006	4.99
14	Alice Fantasia	2006	0.99
15	Alien Center	2006	2.99
16	Alley Evolution	2006	2.99
17	Alone Trip	2006	0.99

[Film A to Z](#)

3.

What is the average rental rate for each rating category?








Query		Query History
1	<b>SELECT</b> rating, <b>AVG</b> (rental_rate)	
2	<b>FROM</b> film	
3	<b>GROUP BY</b> rating	

Data output		Messages	Notifications
<div></div>			
	rating mpaa_rating	avg numeric	
1	R	2.938717948	
2	NC-17	2.970952380	
3	G	2.888876404	
4	PG	3.051855670	
5	PG-13	3.034843049	

[Film AVG](#)

What are the minimum and maximum rental durations for each rating category?

Query		Query History
1	<b>SELECT</b> rating, <b>MIN</b> (rental_duration), <b>MAX</b> (rental_duration)	
2	<b>FROM</b> film	
3	<b>GROUP BY</b> rating	

Data output		Messages	Notifications
<div></div>			
	rating mpaa_rating	min smallint	max smallint
1	R	3	7
2	NC-17	3	7
3	G	3	7
4	PG	3	7
5	PG-13	3	7

[Film MIN-MAX](#)

4.

**Can you outline the procedure for migrating the data and who will be responsible for it?**

The process of migrating data is assigned to data engineers and is a procedure known as ETL, which stands for Extract, Transform and Load. This process is necessary to make data compatible with the format and the rest of the data that is being used in the analysis.

**What problems do you foresee if you start analyzing the data before it's been loaded into the data warehouse?**

The biggest problem that comes from analyzing the data before it's being loaded into the data warehouse is that the format of the data that haven't been migrated might be incompatible with the data that is already available in the warehouse which can create misleading analysis and inaccurate insights