

Group Assignment 1: Data Migration

Nazar Tkhir, Severyn Shykula, Petro Prokopets, Mykola Vysotskyi

Team Roles

Nazar, Petro: silver

Severyn: question answering, monitoring

Mykola: bronze

Solution was architected together

GitHub repo link: <https://github.com/four-dragons-casino/databricks-group-assignment>

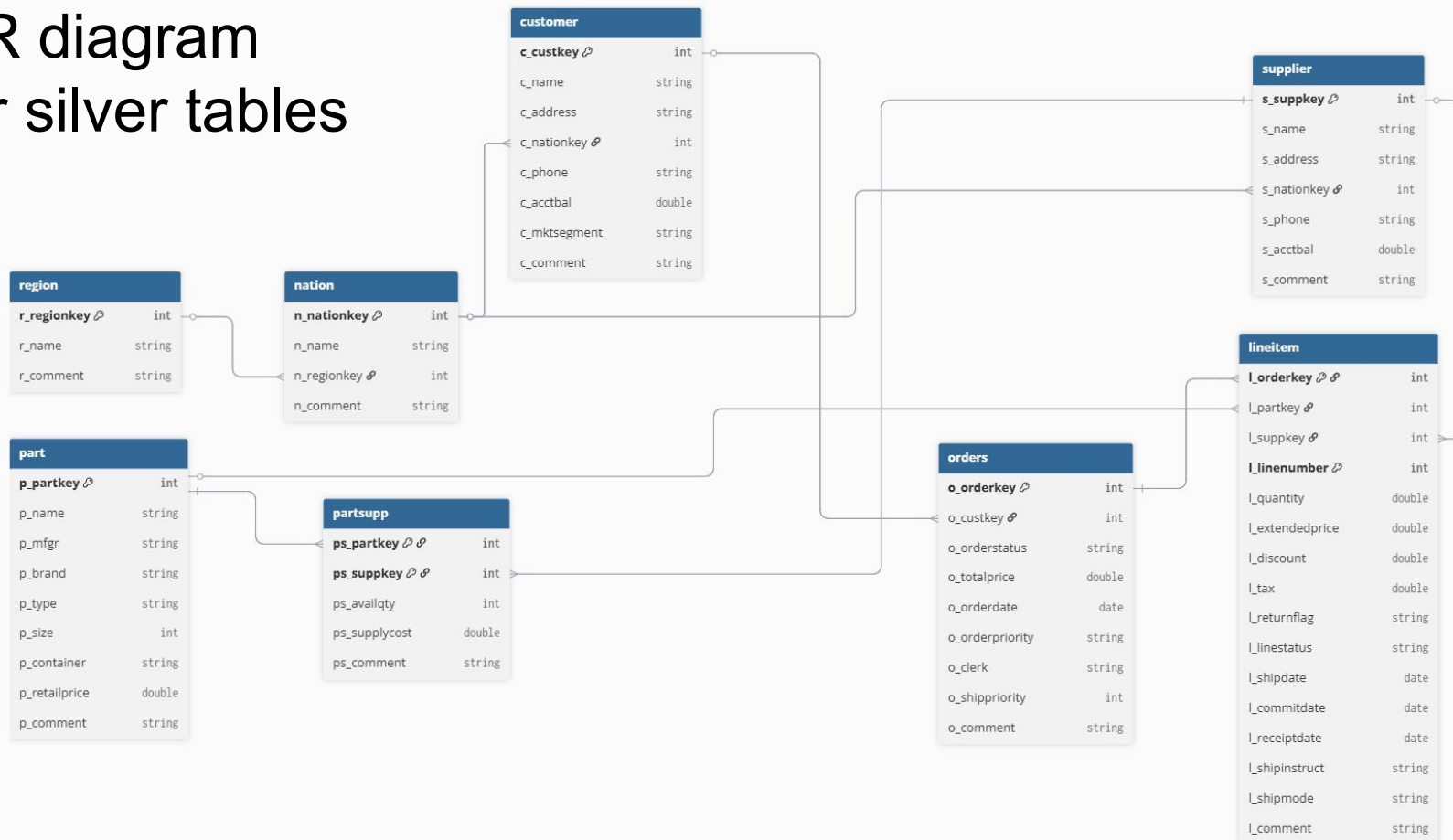
Challenges

Координація роботи в дата
брікс, узгодження назв і т.д.

Ніхто не вміє робити презентації

Нові тулзи відповідно
нові челенджі з ними,
швидке вивчення

ER diagram for silver tables



Demo time

Answers to questions

How many items in total customer 446044 bought?

▶ ✓ 11:13 AM (11s)

```
%sql
SELECT SUM(l_quantity)
FROM retail_migration.silver.lineitem l
JOIN retail_migration.silver.orders o ON l.l_orderkey = o.o_orderkey
WHERE o.o_custkey = 446044;
```

> [See performance \(1\)](#)

Table ▾ +

	1.2 SUM(l_quantity)
1	3055

First purchase date of 564787

▶ ✓ 11:13 AM (2s)

```
%sql
SELECT MIN(o_orderdate)
FROM retail_migration.silver.orders
WHERE o_custkey = 564787;
```

> [See performance \(1\)](#)

▶ [_sqldf](#): pyspark.sql.connect.dataframe.DataFrame = [MIN(o_orderdate): date]

Table ▾ +

	📅 MIN(o_orderdate)
1	1992-03-04

Answers to questions

Orders count for 85909

▶ ▼ ✓ 11:13 AM (1s)

```
%sql
SELECT COUNT(*)
FROM retail_migration.silver.orders
WHERE o_custkey = 85909;
```

> [See performance \(1\)](#)

Table ▼		+
	1 ² ₃ COUNT(*)	
1	39	

Nationality of customer who made most orders

▶ ▼ ✓ 11:13 AM (3s)

```
%sql
SELECT n.n_name, COUNT(o.o_orderkey) as total_orders
FROM retail_migration.silver.orders o
JOIN retail_migration.silver.customer c ON o.o_custkey = c.c_custkey
JOIN retail_migration.silver.nation n ON c.c_nationkey = n.n_nationkey
GROUP BY n.n_name
ORDER BY total_orders DESC
LIMIT 1;
```

> [See performance \(1\)](#)

▶ [_sqlidf](#): pyspark.sql.connect.dataframe.DataFrame = [n_name: string, total_orders: long]

Table ▼			+
	A ^B _C n_name	1 ² ₃ total_orders	
1	INDONESIA	304068	

Thank you!