# Aggregation with Grouping

Dr Paolo Guagliardo


Fall 2020 (v20.1.1)

This page is intentionally left blank

# General idea

**Account**

| Num | Branch | CID | Balance |
|---|---|---|---|
| 111 | London | 1 | 1330.00 |
| 222 | London | 2 | 1756.00 |
| 333 | Edinburgh | 1 | 450.00 |

How much money does **each** customer have in total **across all of his accounts**?

Idea

1. Partition Account into **groups** (one per customer) of rows
2. Sum balances in each group separately
3. Take the union of the results for each group

**Account**

| Num | Branch | CID | Balance |
|---|---|---|---|
| 111 | London | 1 | 1330.00 |
| 222 | London | 2 | 1756.00 |
| 333 | Edinburgh | 1 | 450.00 |

| Num | Branch | CID | Balance |
|---|---|---|---|
| 111 | London | 1 | 1330.00 |
| 333 | Edinburgh | 1 | 450.00 |

| Num | Branch | CID | Balance |
|---|---|---|---|
| 222 | London | 2 | 1756.00 |

| | CID | SUM |
|---|---|---|
| | 1 | 1780.00 |

| | CID | SUM |
|---|---|---|
| | 2 | 1756.00 |

| CID | SUM |
|---|---|
| 1 | 1780.00 |
| 2 | 1756.00 |

# Grouping in SQL

**Account**

| Number | Branch | CustID | Balance |
|---|---|---|---|
| 111 | London | 1 | 1330.00 |
| 222 | London | 2 | 1756.00 |
| 333 | Edinburgh | 1 | 450.00 |

How much money does **each** customer have in total **across all of his accounts**?

```
SELECT    A.custid, SUM(A.balance)
FROM      Account A
GROUP BY  A.custid ;
```

Answer:

| CustID | SUM |
|---|---|
| 1 | 1780.00 |
| 2 | 1756.00 |

**Account**

| Number | Branch | CustID | Balance |
|---|---|---|---|
| 111 | London | 1 | 1330.00 |
| 222 | London | 2 | 1756.00 |
| 333 | Edinburgh | 1 | 450.00 |

| CustID | Balance |
|---|---|
| 1 | 1330.00 |
| 1 | 450.00 |

| CustID | Balance |
|---|---|
| 2 | 1756.00 |

| CustID | SUM |
|---|---|
| 1 | 1780.00 |

| CustID | SUM |
|---|---|
| 2 | 1756.00 |

| CustID | SUM |
|---|---|
| 1 | 1780.00 |
| 2 | 1756.00 |

# Grouping in SQL: Another example

**Account**

| Number | Branch | CustID | Balance |
|---|---|---|---|
| 111 | London | 1 | 1330.00 |
| 222 | London | 2 | 1756.00 |
| 333 | Edinburgh | 1 | 450.00 |

How much money is there in **total in each branch**?

```
SELECT    A.branch, SUM(A.balance)
FROM      Account A
GROUP BY  A.branch ;
```

Answer:

| Branch | SUM |
|---|---|
| London | 3086.00 |
| Edinburgh | 450.00 |

**Account**

| Number | Branch | CustID | Balance |
|---|---|---|---|
| 111 | London | 1 | 1330.00 |
| 222 | London | 2 | 1756.00 |
| 333 | Edinburgh | 1 | 450.00 |

| Branch | Balance |
|---|---|
| London | 1330.00 |
| London | 1756.00 |

| Branch | Balance |
|---|---|
| Edinburgh | 450.00 |

| Branch | SUM |
|---|---|
| London | 3086.00 |

| Branch | SUM |
|---|---|
| Edinburgh | 450.00 |

| Branch | SUM |
|---|---|
| London | 3086.00 |
| Edinburgh | 450.00 |

# YAGE: Yet Another Grouping Example

Account

| Number | Branch | CustID | Balance |
|---|---|---|---|
| 111 | London | 1 | 1330.00 |
| 222 | London | 2 | 1756.00 |
| 333 | Edinburgh | 1 | 450.00 |
| 444 | Edinburgh | 2 | 613.00 |
| 555 | London | 1 | 300.00 |
| 666 | Edinburgh | 2 | 1217.00 |

How much money does **each customer** have in **each branch**?

```
SELECT    A.custid, A.branch, SUM(A.balance)
FROM      Account A
GROUP BY  A.custid, A.branch ;
```

Answer:

| CustID | Branch | SUM |
|---|---|---|
| 1 | London | 1630.00 |
| 1 | Edinburgh | 450.00 |
| 2 | London | 1756.00 |
| 2 | Edinburgh | 1830.00 |

Account

| Number | Branch | CustID | Balance |
|---|---|---|---|
| 111 | London | 1 | 1330.00 |
| 222 | London | 2 | 1756.00 |
| 333 | Edinburgh | 1 | 450.00 |
| 444 | Edinburgh | 2 | 613.00 |
| 555 | London | 1 | 300.00 |
| 666 | Edinburgh | 2 | 1217.00 |

| Branch | CustID | Balance |
|---|---|---|
| London | 1 | 1330.00 |
| London | 1 | 300.00 |

| Branch | CustID | Balance |
|---|---|---|
| London | 2 | 1756.00 |

| Branch | CustID | Balance |
|---|---|---|
| Edinburgh | 1 | 450.00 |

| Branch | CustID | Balance |
|---|---|---|
| Edinburgh | 2 | 613.00 |
| Edinburgh | 2 | 1217.00 |

| Branch | CustID | SUM |
|---|---|---|
| London | 1 | 1630.00 |

| Branch | CustID | SUM |
|---|---|---|
| London | 2 | 1756.00 |

| Branch | CustID | SUM |
|---|---|---|
| Edinburgh | 1 | 450.00 |

| Branch | CustID | SUM |
|---|---|---|
| Edinburgh | 2 | 1830.00 |

| CustID | Branch | SUM |
|---|---|---|
| 1 | London | 1630.00 |
| 1 | Edinburgh | 450.00 |
| 2 | London | 1756.00 |
| 2 | Edinburgh | 1830.00 |

# Beware

In queries with **GROUP BY**, attributes in **SELECT must**

- appear in the **GROUP BY**

  **or**

- be used in an aggregate function

The following query gives an error:

```
SELECT    A.custid, A.branch, SUM(A.balance)
FROM      Account A
GROUP BY  A.branch ;
```

# Filtering based on aggregation

**Account**

| Number | Branch | CustID | Balance |
|---|---|---|---|
| 111 | London | 1 | 1330.00 |
| 222 | London | 2 | 1756.00 |
| 333 | Edinburgh | 1 | 450.00 |

Branches with a total balance (across accounts) of at least 500?

```
SELECT    A.branch, SUM(A.balance)
FROM      Account A
GROUP BY  A.branch
HAVING    SUM(A.balance) >= 500 ;
```

Answer:

| Branch | SUM |
|---|---|
| London | 3086.00 |

# Order of evaluation

1. Take rows from the (joined) tables listed in **`FROM`**
2. Discard rows not satisfying the **`WHERE`** condition
3. Partition rows according to attributes in **`GROUP BY`**
4. **Compute aggregates**
5. Discard rows not satisfying the **`HAVING`** condition
6. Output the values of expressions listed in **`SELECT`**

This page is intentionally left blank

# Aggregation and arithmetic (1)

**Account**

| Number | Branch | CustID | Balance | Spend |
|---|---|---|---|---|
| 111 | London | 1 | 1330.00 | 250.00 |
| 222 | London | 2 | 1756.00 | 356.00 |
| 333 | Edinburgh | 1 | 450.00 | 0.00 |

Money available in total to each customer across his accounts

```
SELECT    A.custid, SUM(A.balance - A.spend)
FROM      Account A
GROUP BY  A.custid ;
```

Answer:

| CustID | SUM |
|---|---|
| 1 | 1530.00 |
| 2 | 1400.00 |

**Account**

| Number | Branch | CustID | Balance | Spend |
|---|---|---|---|---|
| 111 | London | 1 | 1330.00 | 250.00 |
| 222 | London | 2 | 1756.00 | 356.00 |
| 333 | Edinburgh | 1 | 450.00 | 0.00 |

| CustID | Balance | Spend |
|---|---|---|
| 1 | 1330.00 | 250.00 |
| 1 | 450.00 | 0.00 |

| CustID | Balance | Spend |
|---|---|---|
| 2 | 1756.00 | 356.00 |

| CustID | Balance − Spend |
|---|---|
| 1 | 1080.00 |
| 1 | 450.00 |

| CustID | Balance − Spend |
|---|---|
| 2 | 1400.00 |

| CustID | SUM |
|---|---|
| 1 | 1530.00 |

| CustID | SUM |
|---|---|
| 2 | 1400.00 |

| CustID | SUM |
|---|---|
| 1 | 1530.00 |
| 2 | 1400.00 |

# Aggregation and arithmetic (2)

**Account**

| Number | Branch | CustID | Balance | Spend |
|---|---|---|---|---|
| 111 | London | 1 | 1330.00 | 250.00 |
| 222 | London | 2 | 1756.00 | 356.00 |
| 333 | Edinburgh | 1 | 450.00 | 0.00 |

Money available in total to each customer across his accounts

```
SELECT    A.custid, SUM(A.balance) - SUM(A.spend)
FROM      Account A
GROUP BY  A.custid ;
```

Answer:

| CustID | ?column? |
|---|---|
| 1 | 1530.00 |
| 2 | 1400.00 |

**Account**

| Number | Branch | CustID | Balance | Spend |
|---|---|---|---|---|
| 111 | London | 1 | 1330.00 | 250.00 |
| 222 | London | 2 | 1756.00 | 356.00 |
| 333 | Edinburgh | 1 | 450.00 | 0.00 |

| CustID | Balance | Spend |
|---|---|---|
| 1 | 1330.00 | 250.00 |
| 1 | 450.00 | 0.00 |

| CustID | Balance | Spend |
|---|---|---|
| 2 | 1756.00 | 356.00 |

| CustID | SUM | SUM |
|---|---|---|
| 1 | 1780.00 | 250.00 |

| CustID | SUM | SUM |
|---|---|---|
| 2 | 1756.00 | 356.00 |

| CustID | ?column? |
|---|---|
| 1 | 1530.00 |

| CustID | ?column? |
|---|---|
| 2 | 1400.00 |

| CustID | ?column? |
|---|---|
| 1 | 1530.00 |
| 2 | 1400.00 |