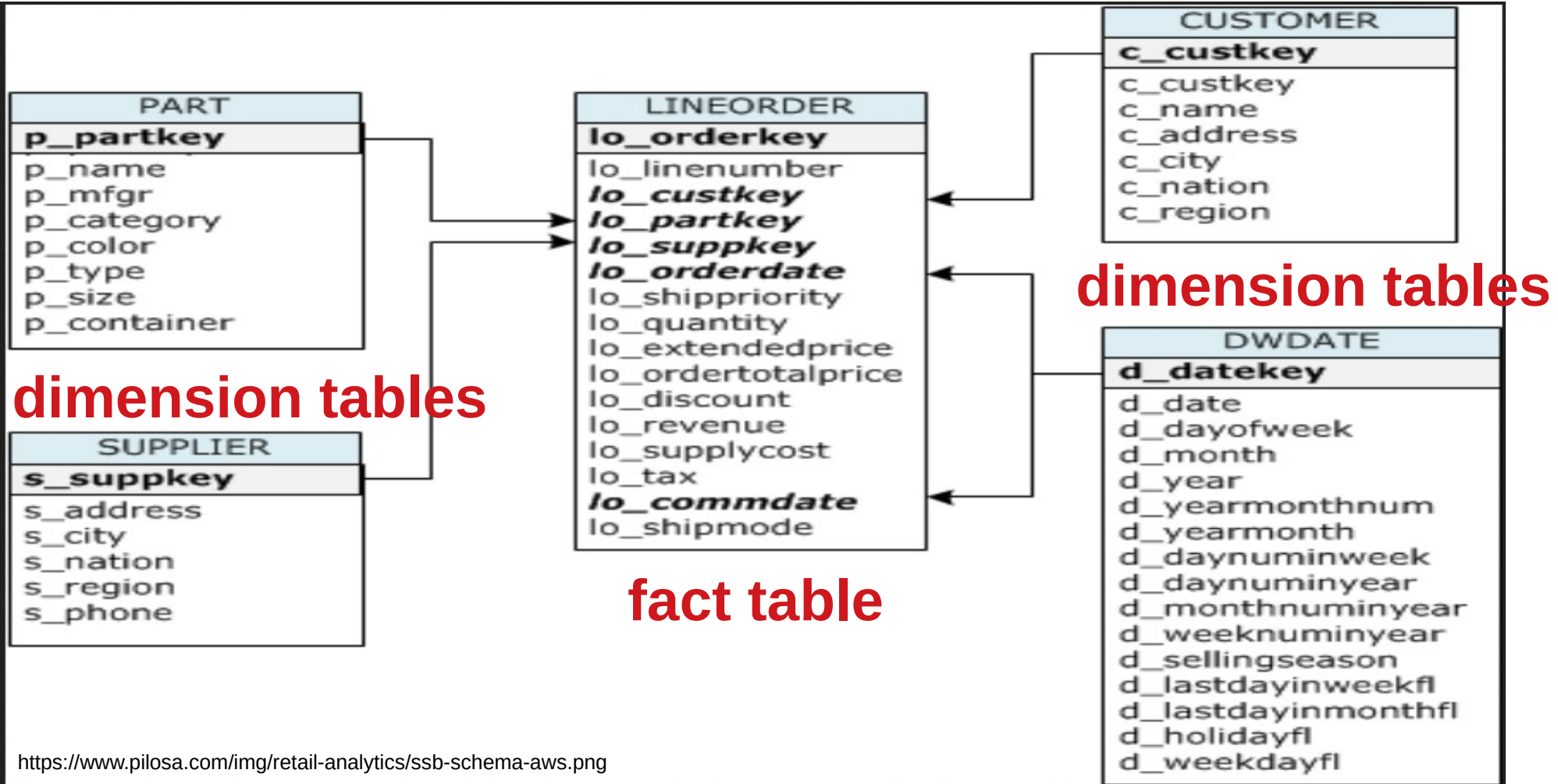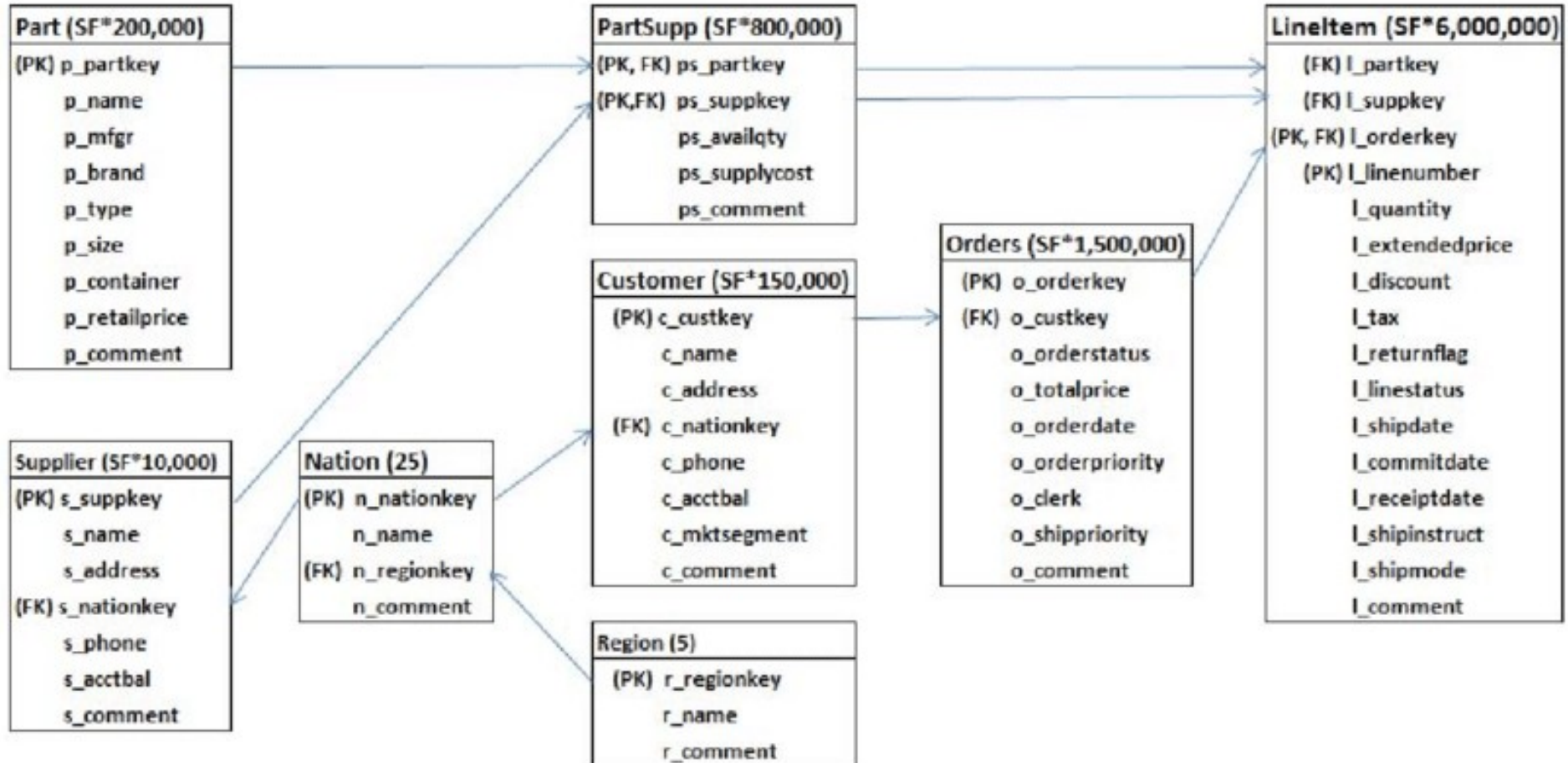# On-Line Analytic Processing (OLAP) & Data Cubes

OLAP vs. OLTP

- On-Line Analytic Processing (OLAP)
  - Decision-support over data warehouses
  - Highly complex queries with one or more aggregations
  - Examine large amounts of data even when the result is small
    - Queries in Lab 6 over TPC-H
- On-Line Transaction Processing (OLTP)
  - Modification operations (transactions)
  - Touch a tiny portion (one or a few tuples) of the database
    - Record a new order in TPC-H

Star Schema

PART

p_partkey
p_name
p_mfgr
p_category
p_color
p_type
p_size
p_container

SUPPLIER

s_suppkey
s_address
s_city
s_nation
s_region
s_phone

dimension tables

LINEORDER

lo_orderkey
lo_linenumber
lo_custkey
lo_partkey
lo_suppkey
lo_orderdate
lo_shippriority
lo_quantity
lo_extendedprice
lo_ordertotalprice
lo_discount
lo_revenue
lo_supplycost
lo_tax
lo_commdate
lo_shipmode

fact table

CUSTOMER

c_custkey
c_custkey
c_name
c_address
c_city
c_nation
c_region

dimension tables

DWDATE

d_datekey
d_date
d_dayofweek
d_month
d_year
d_yearmonthnum
d_yearmonth
d_daynuminweek
d_daynuminyear
d_monthnuminyear
d_weeknuminyear
d_sellingseason
d_lastdayinweekfl
d_lastdayinmonthfl
d_holidayfl
d_weekdayfl

https://www.pilosa.com/img/retail-analytics/ssb-schema-aws.png

TPC-H Schema (Snowflake)

Part (SF*200,000)
- (PK) p_partkey
- p_name
- p_mfgr
- p_brand
- p_type
- p_size
- p_container
- p_retailprice
- p_comment

PartSupp (SF*800,000)
- (PK, FK) ps_partkey
- (PK,FK) ps_suppkey
- ps_availqty
- ps_supplycost
- ps_comment

LineItem (SF*6,000,000)
- (FK) l_partkey
- (FK) l_suppkey
- (PK, FK) l_orderkey
- (PK) l_linenumber
- l_quantity
- l_extendedprice
- l_discount
- l_tax
- l_returnflag
- l_linestatus
- l_shipdate
- l_commitdate
- l_receiptdate
- l_shipinstruct
- l_shipmode
- l_comment

Customer (SF*150,000)
- (PK) c_custkey
- c_name
- c_address
- (FK) c_nationkey
- c_phone
- c_acctbal
- c_mktsegment
- c_comment

Orders (SF*1,500,000)
- (PK) o_orderkey
- (FK) o_custkey
- o_orderstatus
- o_totalprice
- o_orderdate
- o_orderpriority
- o_clerk
- o_shippriority
- o_comment

Supplier (SF*10,000)
- (PK) s_suppkey
- s_name
- s_address
- (FK) s_nationkey
- s_phone
- s_acctbal
- s_comment

Nation (25)
- (PK) n_nationkey
- n_name
- (FK) n_regionkey
- n_comment

Region (5)
- (PK) r_regionkey
- r_name
- r_comment

OLAP Query Example

select n_name, sum(o_totalprice) as tot_orders

from orders, customer, nation, region

where o_custkey = c_custkey

    and c_nationkey = n_nationkey

    and n_regionkey = r_regionkey

    and o_orderdate >= '1996-01-01'

    and o_orderdate < '1997-01-01'

    and r_name = 'AMERICA'

group by n_name

order by tot_orders desc

CANADA|18482207.74

BRAZIL|15273545.8

UNITED STATES|11750866.68

ARGENTINA|11502493.16

PERU|9312955.18

Slicing & Dicing OLAP Queries

SELECT <dicing attributes & aggregations>

FROM <fact table joined with dimension tables>

WHERE <**slicing attributes**>

GROUP BY <**dicing attributes**>

Data Exploration with Drill-down and Roll-up

select n_name, sum(o_totalprice) as tot_orders

from orders, customer, nation, region

where o_custkey = c_custkey

    and c_nationkey = n_nationkey

    and n_regionkey = r_regionkey

    and **o_orderdate >= '1996-01-01'**

    and o_orderdate < '1997-01-01'

    and r_name = 'AMERICA'

group by **n_name**

order by tot_orders desc

CANADA|18482207.74

BRAZIL|15273545.8

UNITED STATES|11750866.68

ARGENTINA|11502493.16

PERU|9312955.18

Drill-down on Market Segment in US

select c_mktsegment, sum(o_totalprice) as tot_orders

from orders, customer, nation

where o_custkey = c_custkey

    and c_nationkey = n_nationkey

    and o_orderdate >= '1996-01-01'

    and o_orderdate < '1997-01-01'

    and **n_name = 'UNITED STATES'**

group by **c_mktsegment**

AUTOMOBILE|1764146.3

BUILDING|3949798.52

FURNITURE|2463719.39

HOUSEHOLD|2807178.64

MACHINERY|766023.83

Drill-down on Month for BUILDING

select substr(o_orderdate, 6, 2) as month, sum(o_totalprice) as tot_orders

from orders, customer, nation

where o_custkey = c_custkey

    and c_nationkey = n_nationkey

    and o_orderdate >= '1996-01-01'

    and o_orderdate < '1997-01-01'

    and n_name = 'UNITED STATES'

    and **c_mktsegment = 'BUILDING'**

group by **month**

01|307934.57

04|449200.42

05|504249.66

06|535603.54

07|197825.34

08|133971.12

09|719143.56

10|446284.43

11|401000.0

12|254585.88

Roll-up on Month

select substr(o_orderdate, 6, 2) as month, sum(o_totalprice) as tot_orders

from orders, customer, nation

where o_custkey = c_custkey

   and c_nationkey = n_nationkey

   and o_orderdate >= '1996-01-01'

   and o_orderdate < '1997-01-01'

   and n_name = 'UNITED STATES'

group by **month**

01|763247.39

02|589382.5

03|41703.87

04|1424955.06

05|1107433.27

06|1239444.19

07|992346.92

08|848086.1

09|1957533.34

10|609445.63

11|1620398.94

12|556889.47

Data Cube

Measures
Sales amount
Units sold
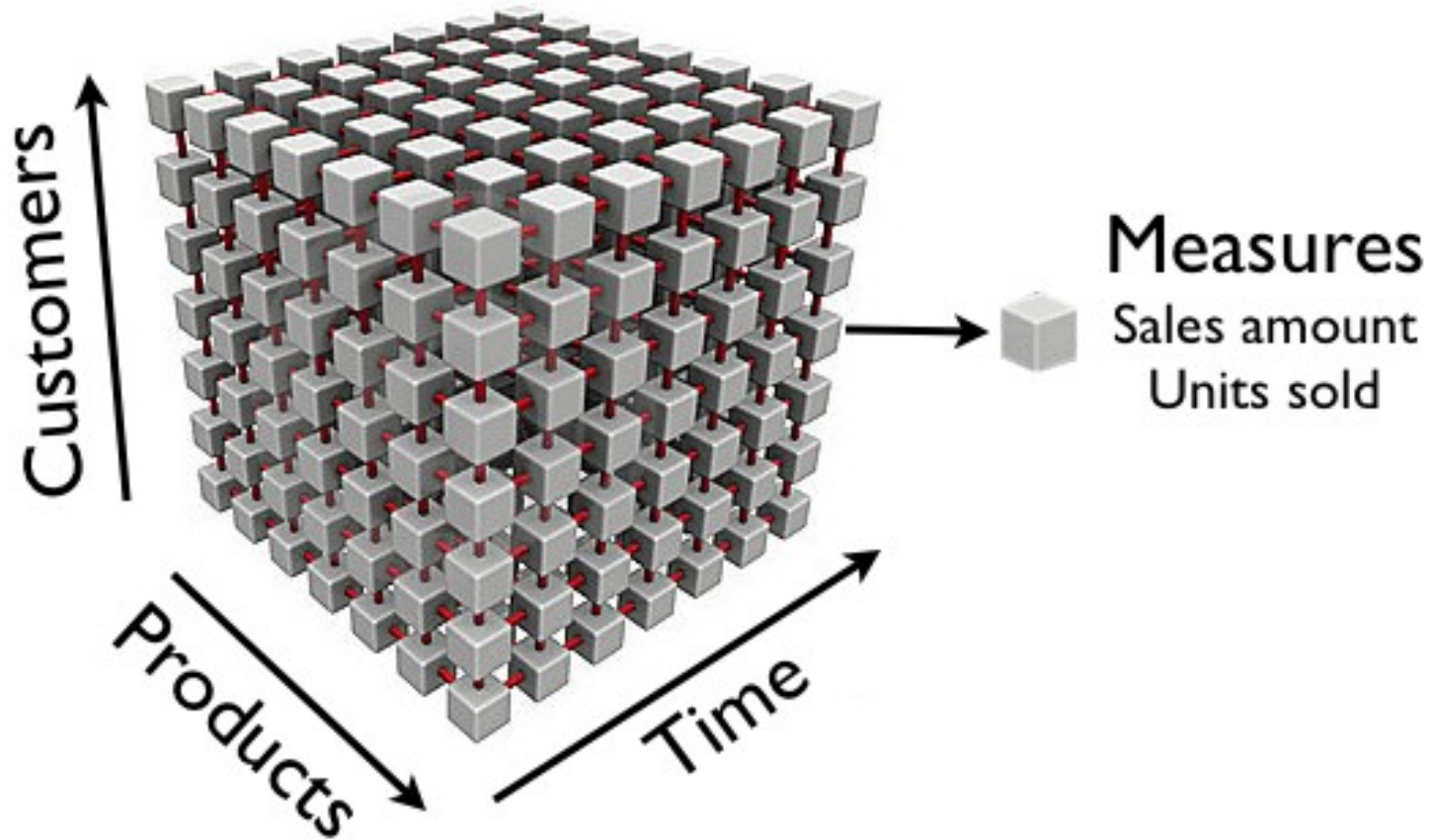
Build Data Cube (1)

select n_name as country, c_mktsegment as segment,

    substr(o_orderdate, 6, 2) as month,

    sum(o_totalprice) as tot_orders

from orders, customer, nation, region

where o_custkey = c_custkey

    and c_nationkey = n_nationkey

    and n_regionkey = r_regionkey

    and o_orderdate >= '1996-01-01'

    and o_orderdate < '1997-01-01'

    and r_name = 'AMERICA'

group by n_name, c_mktsegment, month

UNITED STATES|AUTOMOBILE|02|320234.13

UNITED STATES|AUTOMOBILE|03|41703.87

UNITED STATES|AUTOMOBILE|04|519989.12

UNITED STATES|AUTOMOBILE|05|366946.75

UNITED STATES|AUTOMOBILE|07|115110.42

UNITED STATES|AUTOMOBILE|08|133691.66

UNITED STATES|AUTOMOBILE|09|73907.33

UNITED STATES|AUTOMOBILE|11|88538.53

UNITED STATES|BUILDING|01|307934.57

UNITED STATES|BUILDING|04|449200.42

UNITED STATES|BUILDING|05|504249.66

UNITED STATES|BUILDING|06|535603.54

UNITED STATES|BUILDING|07|197825.34

Build Data Cube (2)

```
select '*' as country, segment, month, tot_orders
from
    (select
        c_mktsegment as segment,
        substr(o_orderdate, 6, 2) as month,
        sum(o_totalprice) as tot_orders
    from orders, customer, nation, region
    where o_custkey = c_custkey
        and c_nationkey = n_nationkey
        and n_regionkey = r_regionkey
        and o_orderdate >= '1996-01-01'
        and o_orderdate < '1997-01-01'
        and r_name = 'AMERICA'
    group by c_mktsegment, month)
```

```
*|FURNITURE|07|1536410.0
*|FURNITURE|08|1275646.64
*|FURNITURE|09|850020.17
*|FURNITURE|10|627374.67
*|FURNITURE|11|1674244.08
*|FURNITURE|12|596075.08
*|HOUSEHOLD|01|857604.53
*|HOUSEHOLD|02|1268791.05
*|HOUSEHOLD|03|734929.51
*|HOUSEHOLD|04|977667.03
```

Build Data Cube (3)

```
select '*' as country, '*' as segment, month, tot_orders
from
    (select
        substr(o_orderdate, 6, 2) as month,
        sum(o_totalprice) as tot_orders
    from orders, customer, nation, region
    where o_custkey = c_custkey
        and c_nationkey = n_nationkey
        and n_regionkey = r_regionkey
        and o_orderdate >= '1996-01-01'
        and o_orderdate < '1997-01-01'
        and r_name = 'AMERICA'
    group by month)
```

||01|3951668.23

||02|6019772.02

||03|5905427.19

||04|5324806.52

||05|5117392.48

||06|5055926.24

||07|5821844.72

||08|7009655.79

||09|5697575.73

||10|5306528.08

||11|6371999.76

||12|4739471.8

Build Data Cube (4)

select '*' as country, '*' as segment, '*' as month, tot_orders

from

  (select

    sum(o_totalprice) as tot_orders

  from orders, customer, nation, region

  where o_custkey = c_custkey

    and c_nationkey = n_nationkey

    and n_regionkey = r_regionkey

    and o_orderdate >= '1996-01-01'

    and o_orderdate < '1997-01-01'

    and r_name = 'AMERICA')

||*|66322068.56

SQL Data Cube Operator

create materialized view DataCube as

select n_name as country, c_mktsegment as segment,

    substr(o_orderdate, 6, 2) as month,

    sum(o_totalprice) as tot_orders

from orders, customer, nation, region

where o_custkey = c_custkey

    and c_nationkey = n_nationkey

    and n_regionkey = r_regionkey

    and o_orderdate >= '1996-01-01'

    and o_orderdate < '1997-01-01'

    and r_name = 'AMERICA'

group by n_name, c_mktsegment, month **WITH CUBE**

Data Cube in SQLite

- create table DataCube (

    country char(50), segment char(50), month char(10), tot_orders decimal(20,4),

    primary key (country, segment, month))

- insert into DataCube

    select n_name as country, c_mktsegment as segment, substr(o_orderdate, 6, 2) as month, sum(o_totalprice) as tot_orders

    group by n_name, c_mktsegment, month

    - UNION

    select '*' as country, segment, month, tot_orders

    group by c_mktsegment, month

    select country, '*' as segment, month, tot_orders

    select country, segment, '*' as month, tot_orders

    - UNION

    select '*' as country, '*' as segment, month, tot_orders

    group by month

    select '*' as country, segment, '*' as month, tot_orders

    select country, '*' as segment, '*' as month, tot_orders

    - UNION

    select '*' as country, '*' as segment, '*' as month, tot_orders

Data Exploration with Data Cube

- Data exploration with drill-down and roll-up
  - select country, tot_orders from DataCube where segment = '*' and month = '*'
- Drill-down on market segment in US
  - select segment, tot_orders from DataCube
    where country = 'UNITED STATES' and month = '*'
- Drill-down on month for BUILDING
  - select month, tot_orders from DataCube
    where country = 'UNITED STATES' and segment = 'BUILDING'
- Roll-up on month
  - select month, tot_orders from DataCube
    where country = 'UNITED STATES' and segment = '*'