MTN.BI.03

SQL FOR ANALYSIS

Dataset Transformations: Pivoting and Unpivoting

Author: **Aliaksei Belablotski**
Senior Software Engineer
Aliaksei_Belablotski@epam.com

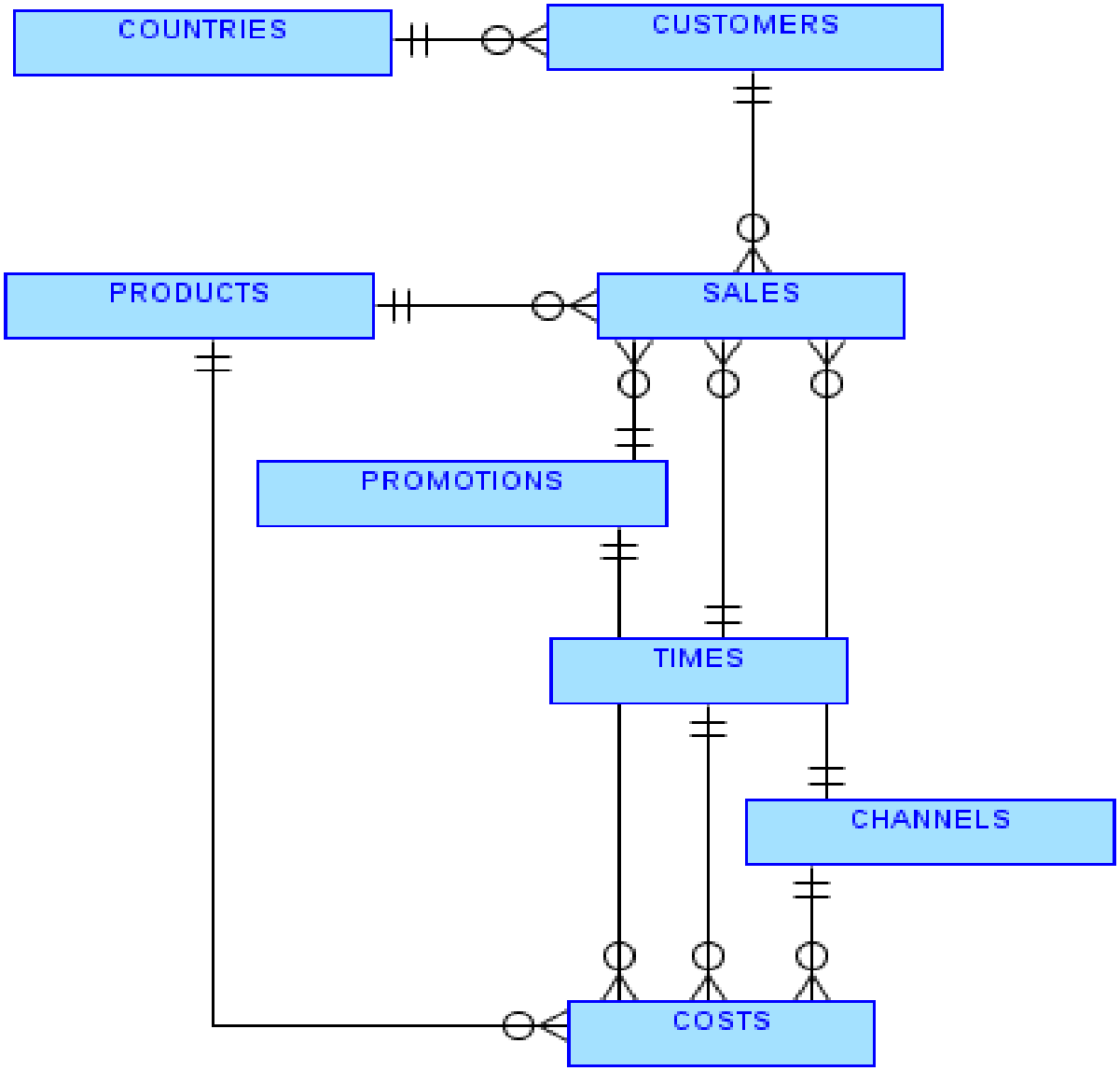
2012 © EPAM Systems, RD Dep.

Objectives

1. Star queries
2. Minimarts
3. Pre-calculated totals
4. Pivoting
5. Unpivoting
6. Dataset reusing

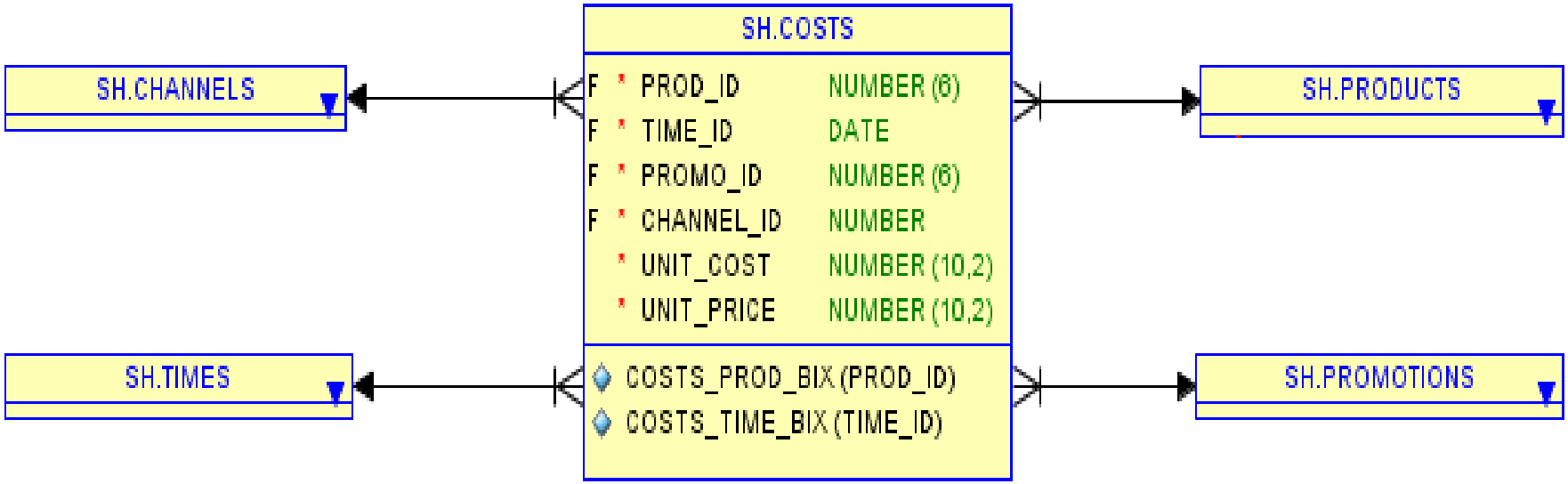
2011 © EPAM Systems, RD Dep.

Sales History Sample Database



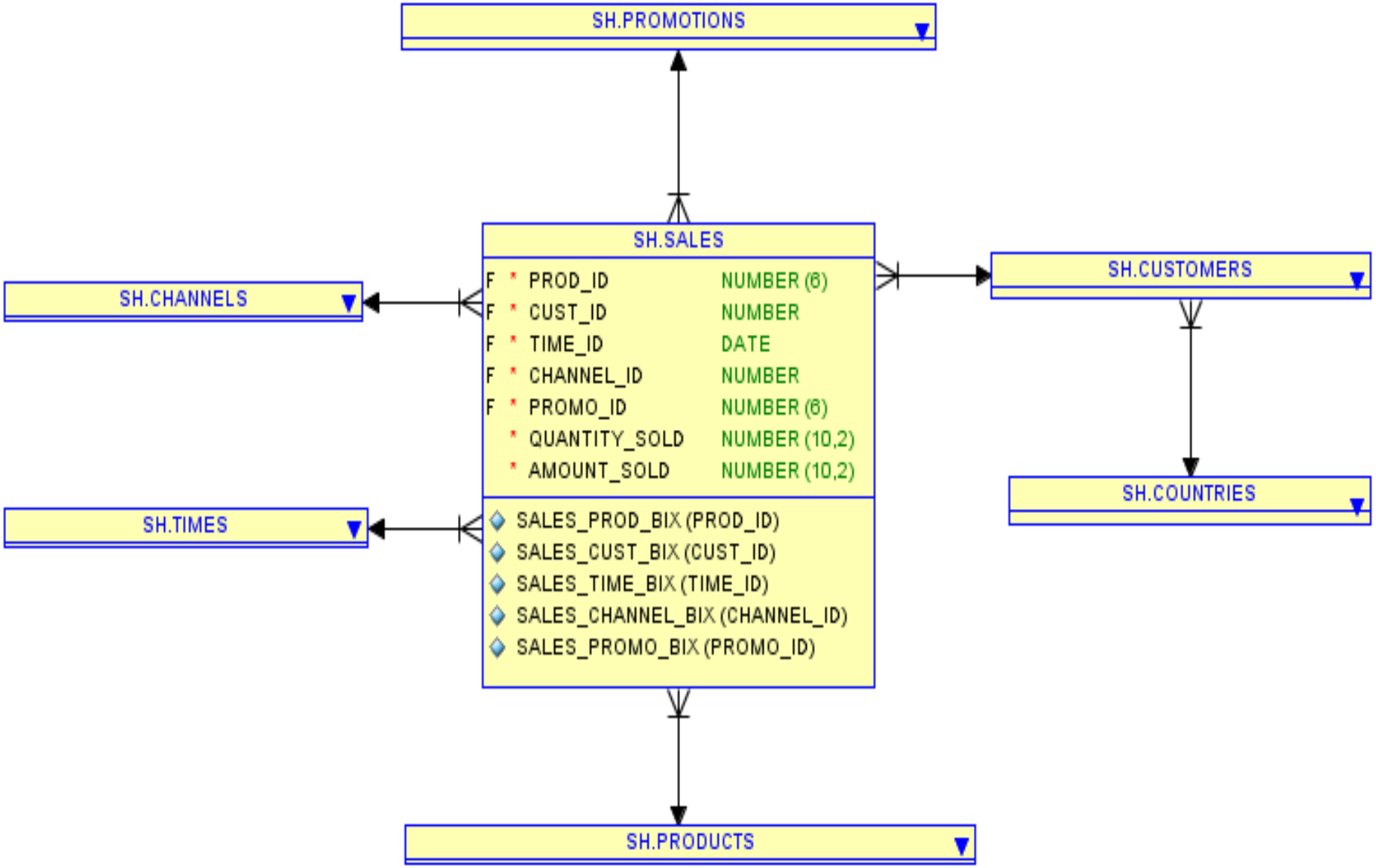
2011 © EPAM Systems, RD Dep.

Costs Star



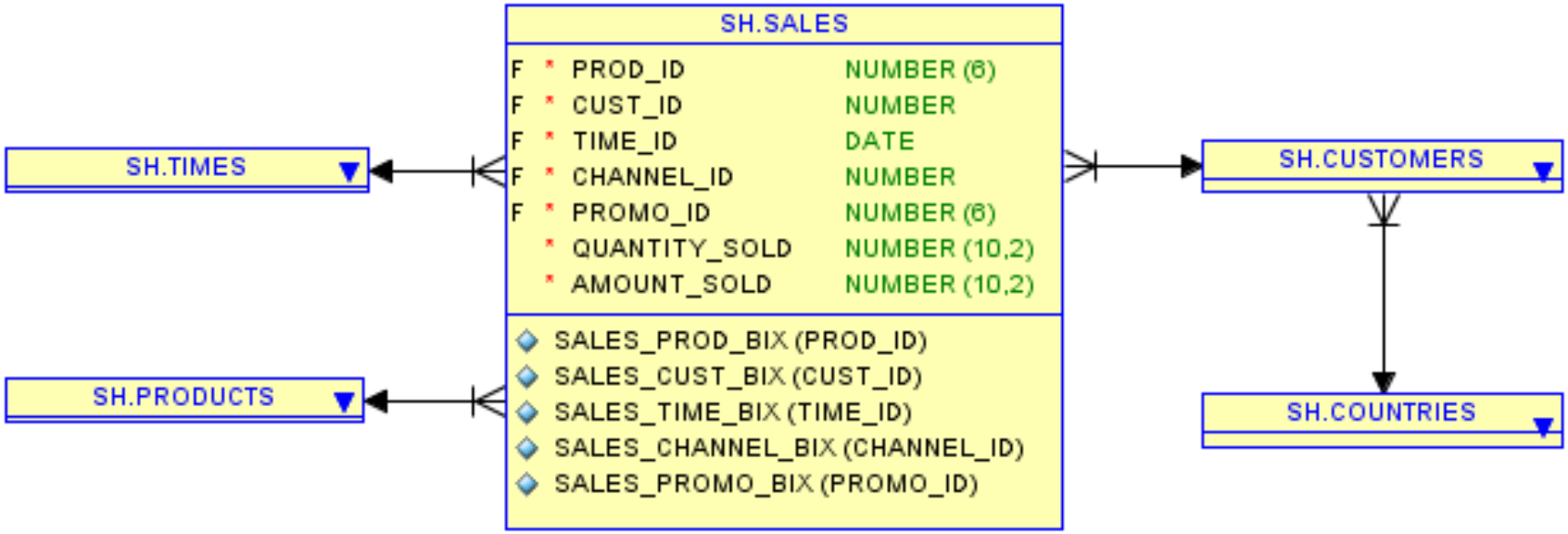
2011 © EPAM Systems, RD Dep.

Sales Star



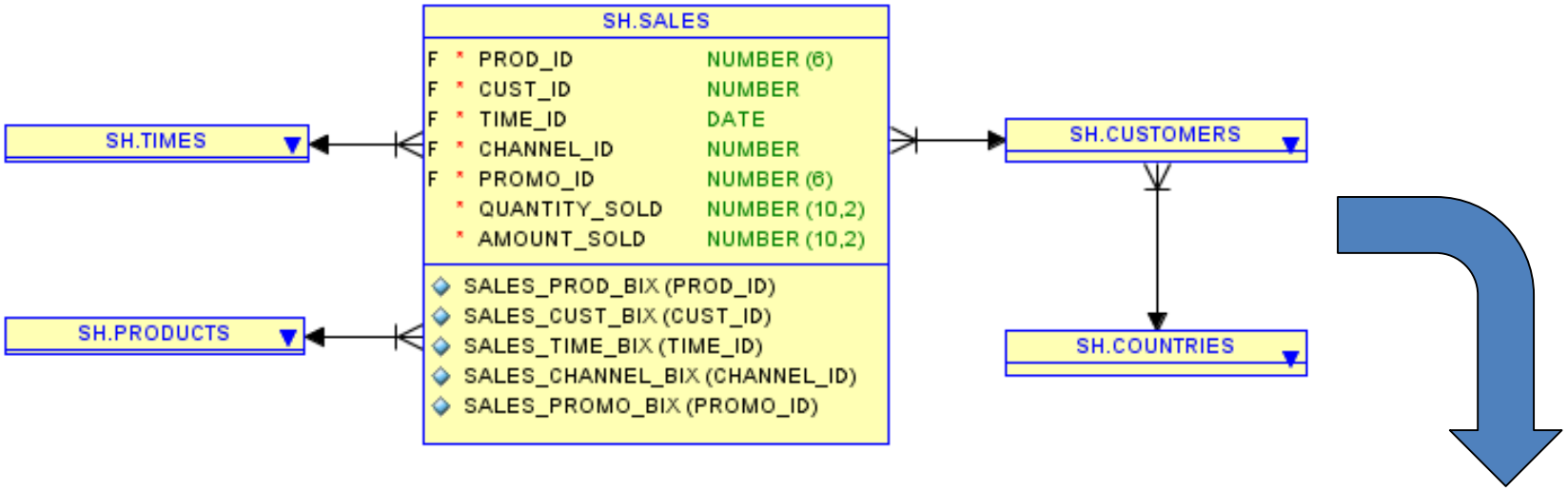
2011 © EPAM Systems, RD Dep.

Star (time, products, customers)



2011 © EPAM Systems, RD Dep.

Star schema transformed to flat minimart



PRODUCT	COUNTRY	CHANNEL	QUARTER	AMOUNT_SOLD	QUANTITY_SOLD
5MP Telephoto Digi...	Spain	3	01	27668.59	24
5MP Telephoto Digi...	United Ki...	3	01	65908.81	57
17" LCD w/built-in...	United St...	3	01	600901.7	487
Mini DV Camcorder ...	Australia	4	01	19538.12	15
Mini DV Camcorder ...	United St...	2	01	311831.67	241
Mini DV Camcorder ...	France	2	01	10439	8
Mini DV Camcorder ...	United Ki...	4	01	5323.22	4
17" LCD w/built-in...	Canada	3	01	31790.14	26
17" LCD w/built-in...	Japan	4	01	14538.52	12
Envoy 256MB - 40GB	United Ki...	2	01	17961.21	18
Envoy 256MB - 40GB	Japan	3	01	69178.06	69
Envoy 256MB - 40GB	Australia	4	01	8697.59	9

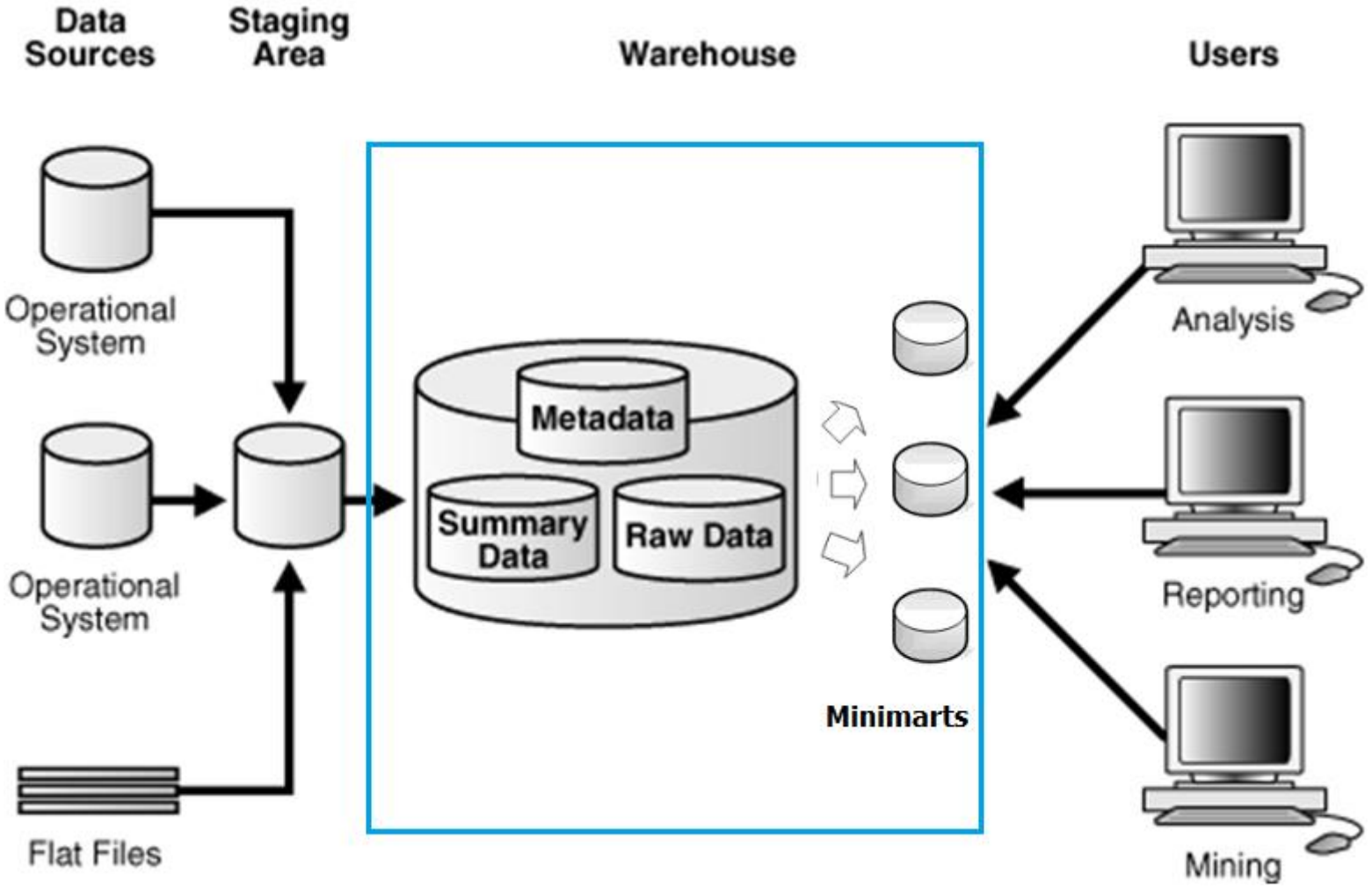
2011 © EPAM Systems, RD Dep.

Star query

```
select PROD_NAME PRODUCT,
  COUNTRY_NAME COUNTRY,
  CHANNEL_ID CHANNEL,
  SUBSTR(CALENDAR_QUARTER_DESC, 6,2) QUARTER,
  SUM(AMOUNT_SOLD) AMOUNT_SOLD,
  SUM(QUANTITY_SOLD) QUANTITY_SOLD
from SALES,
  TIMES,
  CUSTOMERS,
  COUNTRIES,
  PRODUCTS
where SALES.TIME_ID     = TIMES.TIME_ID
and SALES.PROD_ID       = PRODUCTS.PROD_ID
and SALES.CUST_ID       = CUSTOMERS.CUST_ID
and CUSTOMERS.COUNTRY_ID = COUNTRIES.COUNTRY_ID
group by PROD_NAME,
  COUNTRY_NAME,
  CHANNEL_ID,
  SUBSTR(CALENDAR_QUARTER_DESC, 6, 2) ;
```

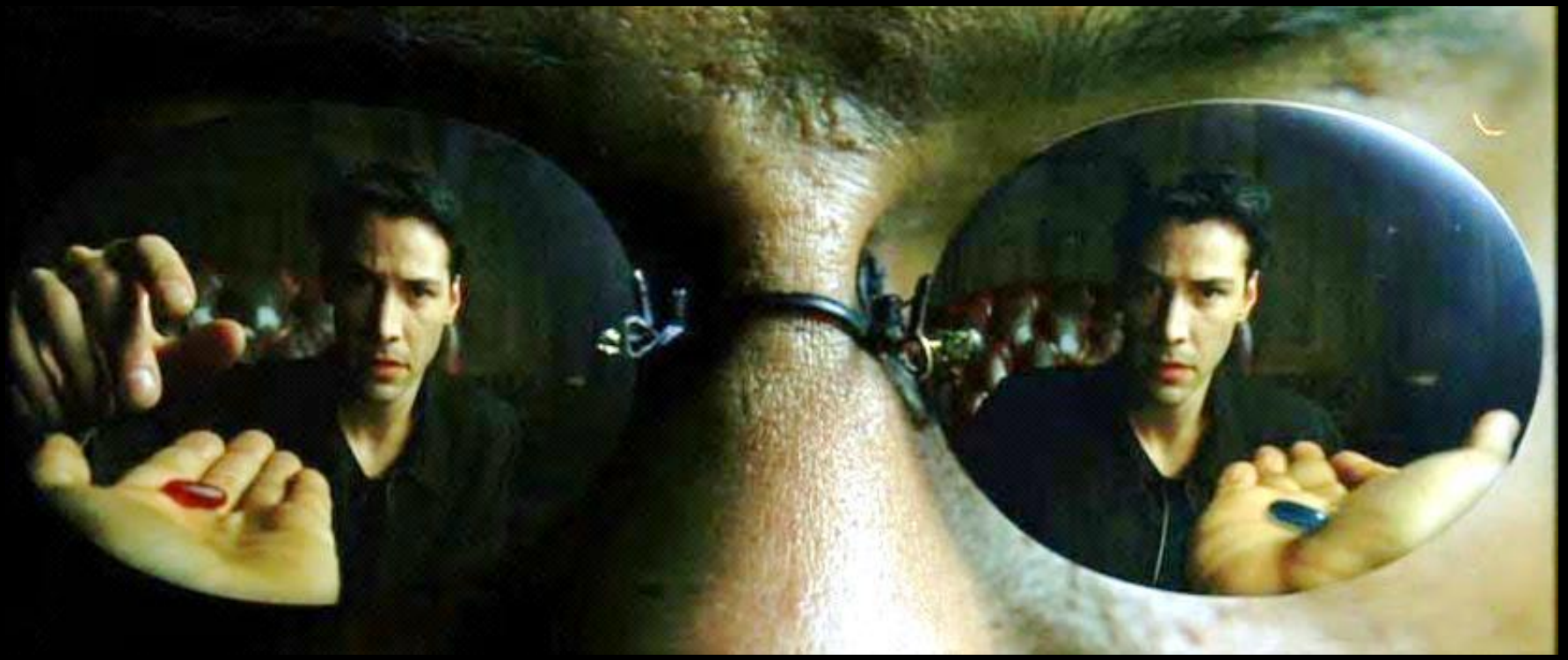

2011 © EPAM Systems, RD Dep.

Data Warehouse with Minimarts



2011 © EPAM Systems, RD Dep.

Star schema queries or minimarts?



2011 © EPAM Systems, RD Dep.

Minimarts organization

- View
- Materialized View
- Table

2011 © EPAM Systems, RD Dep.

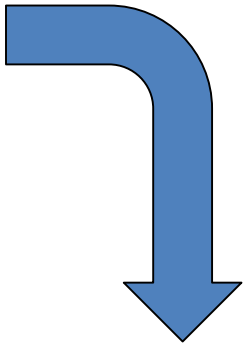
Example: Minimart organized as View

```
create view SALES_VIEW as
select PROD_NAME PRODUCT,
  COUNTRY_NAME COUNTRY,
  CHANNEL_ID CHANNEL,
  SUBSTR(CALENDAR_QUARTER_DESC, 6,2) QUARTER,
  SUM(AMOUNT_SOLD) AMOUNT_SOLD,
  SUM(QUANTITY_SOLD) QUANTITY_SOLD
from SALES, TIMES,
  CUSTOMERS,
  COUNTRIES,
  PRODUCTS
where SALES.TIME_ID     = TIMES.TIME_ID
and SALES.PROD_ID       = PRODUCTS.PROD_ID
and SALES.CUST_ID       = CUSTOMERS.CUST_ID
and CUSTOMERS.COUNTRY_ID = COUNTRIES.COUNTRY_ID
group by PROD_NAME,
  COUNTRY_NAME,
  CHANNEL_ID,
  SUBSTR(CALENDAR_QUARTER_DESC, 6, 2) ;
```


2011 © EPAM Systems, RD Dep.

Select data from SALES_VIEW

```
select *
from SALES_VIEW;
```



Views in Sales History:

- CAL_MONTH_SALES_MV
- FWEEK_PSCAT_SALES_MV

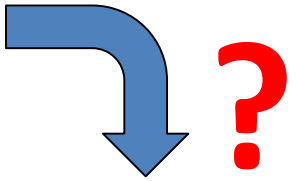
PRODUCT	COUNTRY	CHANNEL	QUARTER	AMOUNT_SOLD	QUANTITY_SOLD
5MP Telephoto Digi...	Spain	3	01	27668.59	24
5MP Telephoto Digi...	United Ki...	3	01	65908.81	57
17" LCD w/built-in...	United St...	3	01	600901.7	487
Mini DV Camcorder ...	Australia	4	01	19538.12	15
Mini DV Camcorder ...	United St...	2	01	311831.67	241
Mini DV Camcorder ...	France	2	01	10439	8
Mini DV Camcorder ...	United Ki...	4	01	5323.22	4
17" LCD w/built-in...	Canada	3	01	31790.14	26
17" LCD w/built-in...	Japan	4	01	14538.52	12
Envoy 256MB - 40GB	United Ki...	2	01	17961.21	18
Envoy 256MB - 40GB	Japan	3	01	69178.06	69
Envoy 256MB - 40GB	Australia	4	01	8697.59	9

2011 © EPAM Systems, RD Dep.

Cross tabular report

PRODUCT	CHANNEL	SUM(AMOUNT_SOLD)
1.44MB External 3.5" Diskette	2	60120.52
1.44MB External 3.5" Diskette	3	137330.01
1.44MB External 3.5" Diskette	4	22167.94
128MB Memory Card	2	168783.39
128MB Memory Card	3	313505.83
128MB Memory Card	4	89044.53
17" LCD w/built-in HDTV Tuner	2	1690316.63
17" LCD w/built-in HDTV Tuner	3	4442061.35
17" LCD w/built-in HDTV Tuner	4	1056793.79
18" Flat Panel Graphics Monitor	2	1127568.55
18" Flat Panel Graphics Monitor	3	3017888.81
18" Flat Panel Graphics Monitor	4	1148972.72
18" Flat Panel Graphics Monitor	9	204297.73
256MB Memory Card	2	127390.3
256MB Memory Card	3	393111.15
256MB Memory Card	4	71203.21

```
select PRODUCT,
    CHANNEL,
    SUM(AMOUNT_SOLD)
from SALES_VIEW
group by
    PRODUCT,
    CHANNEL
order by
    PRODUCT,
    CHANNEL;
```



PRODUCT	DIRECT_SALES	INTERNET_SALES	PARTNERS	TELESALES
1.44MB External 3.5" Diskette	137330.01	22167.94	60120.52	(null)
128MB Memory Card	313505.83	89044.53	168783.39	(null)
17" LCD w/built-in HDTV Tuner	4442061.35	1056793.79	1690316.63	(null)
18" Flat Panel Graphics Monitor	3017888.81	1148972.72	1127568.55	204297.73
256MB Memory Card	393111.15	71203.21	127390.3	(null)

Traditional pivot puzzle (step 1 – diagonals)

```
select PRODUCT, CHANNEL,
  SUM(DECODE(CHANNEL, 2, AMOUNT_SOLD, 0)) as  PARTNERS,
  SUM(DECODE(CHANNEL, 3, AMOUNT_SOLD, 0)) as  DIRECT_SALES,
  SUM(DECODE(CHANNEL, 4, AMOUNT_SOLD, 0)) as INTERNET_SALES,
  SUM(DECODE(CHANNEL, 9, AMOUNT_SOLD, 0)) as TELESALES
from SALES_VIEW
group by PRODUCT, CHANNEL order by PRODUCT, CHANNEL;
```

PRODUCT	CHANNEL	PARTNERS	DIRECT_SALES	INTERNET_SALES	TELESALES
1.44MB External 3.5" Diskette	2	60120.52	0	0	0
1.44MB External 3.5" Diskette	3	0	137330.01	0	0
1.44MB External 3.5" Diskette	4	0	0	22167.94	0
128MB Memory Card	2	168783.39	0	0	0
128MB Memory Card	3	0	313505.83	0	0
128MB Memory Card	4	0	0	89044.53	0
17" LCD w/built-in HDTV Tuner	2	1690316.63	0	0	0
17" LCD w/built-in HDTV Tuner	3	0	4442061.35	0	0
17" LCD w/built-in HDTV Tuner	4	0	0	1056793.79	0
18" Flat Panel Graphics Monitor	2	1127568.55	0	0	0
18" Flat Panel Graphics Monitor	3	0	3017888.81	0	0
18" Flat Panel Graphics Monitor	4	0	0	1148972.72	0
18" Flat Panel Graphics Monitor	9	0	0	0	204297.73
256MB Memory Card	2	127390.3	0	0	0
256MB Memory Card	3	0	393111.15	0	0
256MB Memory Card	4	0	0	71203.21	0

2011 © EPAM Systems, RD Dep.

Traditional pivot puzzle (step 2 – aggregation)

```
select DIA.PRODUCT as PRODUCT,
  SUM(DIA.PARTNERS) as PARTNERS,
  SUM(DIA.DIRECT_SALES) as DIRECT_SALES,
  SUM(DIA.INTERNET_SALES) as INTERNET_SALES,
  SUM(DIA.TELESALES) as TELESALES
from (
  select PRODUCT, CHANNEL,
    SUM(DECODE(CHANNEL, 2, AMOUNT_SOLD, 0)) as PARTNERS,
    SUM(DECODE(CHANNEL, 3, AMOUNT_SOLD, 0)) as DIRECT_SALES,
    SUM(DECODE(CHANNEL, 4, AMOUNT_SOLD, 0)) as INTERNET_SALES,
    SUM(DECODE(CHANNEL, 9, AMOUNT_SOLD, 0)) as TELESALES
  from SALES_VIEW
  group by PRODUCT, CHANNEL
) DIA
group by PRODUCT order by PRODUCT;
```

PRODUCT	PARTNERS	DIRECT_SALES	INTERNET_SALES	TELESALES
1.44MB External 3.5" Diskette	60120.52	137330.01	22167.94	0
128MB Memory Card	168783.39	313505.83	89044.53	0
17" LCD w/built-in HDTV Tuner	1690316.63	4442061.35	1056793.79	0
18" Flat Panel Graphics Monitor	1127568.55	3017888.81	1148972.72	204297.73
256MB Memory Card	127390.3	393111.15	71203.21	0

How to create pivot table more effectively?

PIVOT Syntax

```
SELECT …
FROM   …
PIVOT [XML]
  (
     pivot_clause
     pivot_for_clause
     pivot_in_clause
   )
WHERE  …
```

***pivot_clause*:** defines the columns to be aggregated (pivot is an aggregate operation);

***pivot_for_clause*:** defines the columns to be grouped and pivoted;

***pivot_in_clause*:** defines the filter for the column(s) in the pivot_for_clause (i.e. the range of values to limit the results to). The aggregations for each value in the pivot_in_clause will be transposed into a separate column (where appropriate).

PIVOT Operator

```
select *
from
  (select PRODUCT, CHANNEL, AMOUNT_SOLD from SALES_VIEW) S
  PIVOT (SUM(AMOUNT_SOLD)
            for CHANNEL
            in (2 as PARTNERS, 3 as DIRECT_SALES,
                4 as INTERNET_SALES, 9 as TELESALES))
order by PRODUCT;
```

PRODUCT	PARTNERS	DIRECT_SALES	INTERNET_SALES	TELESALES
1.44MB External 3.5" Diskette	60120.52	137330.01	22167.94	(null)
128MB Memory Card	168783.39	313505.83	89044.53	(null)
17" LCD w/built-in HDTV Tuner	1690316.63	4442061.35	1056793.79	(null)
18" Flat Panel Graphics Monitor	1127568.55	3017888.81	1148972.72	204297.73
256MB Memory Card	127390.3	393111.15	71203.21	(null)

Pivoting on Multiple Columns

```
select *
from
  (select PRODUCT, CHANNEL, QUARTER, QUANTITY_SOLD from
SALES_VIEW)
  PIVOT (SUM(QUANTITY_SOLD)
    for (CHANNEL, QUARTER) in ((5, '02') as CATALOG_Q2,
      (4, '01') as INTERNET_Q1, (4, '04') as INTERNET_Q4,
      (2, '02') as PARTNERS_Q2, (9, '03') as TELE_Q3
    )
)
order by PRODUCT;
```

PRODUCT	CATALOG_Q2	INTERNET_Q1	INTERNET_Q4	PARTNERS_Q2	TELE_Q3
1.44MB External 3.5" Diskette	(null)	557	853	1302	(null)
128MB Memory Card	(null)	332	512	742	(null)
17" LCD w/built-in HDTV Tuner	(null)	181	247	345	(null)
18" Flat Panel Graphics Monitor	(null)	236	286	259	(null)
256MB Memory Card	(null)	78	194	321	(null)

Pivoting: Multiple Aggregates

```
SELECT *
from
  (select PRODUCT, CHANNEL, AMOUNT_SOLD, QUANTITY_SOLD
   from SALES_VIEW)
  PIVOT (SUM(AMOUNT_SOLD) as SUMS, SUM(QUANTITY_SOLD) as SUMQ
    FOR channel IN (2, 3, 4, 9)
)
ORDER BY product;
```

PRODUCT	2_SUMS	2_SUMQ	3_SUMS	3_SUMQ	4_SUMS	4_SUMQ	9_SUMS	9_SUMQ
1.44MB External 3.5" Diskette	60120.52	6455	137330.01	14189	22167.94	2464	(null)	(null)
128MB Memory Card	168783.39	3078	313505.83	5446	89044.53	1701	(null)	(null)
17" LCD w/built-in HDTV Tuner	1690316.63	1461	4442061.35	3625	1056793.79	924	(null)	(null)
18" Flat Panel Graphics Monit	1127568.55	1076	3017888.81	2775	1148972.72	1127	204297.73	227
256MB Memory Card	127390.3	1120	393111.15	3878	71203.21	543	(null)	(null)

Pivoting: Multiple Aggregates

```
SELECT *
from
  (select PRODUCT, CHANNEL, AMOUNT_SOLD, QUANTITY_SOLD
   from SALES_VIEW)
  PIVOT (SUM(AMOUNT_SOLD) AS SUMS, SUM(QUANTITY_SOLD) AS SUMQ
    FOR channel IN (select 2 from dual)
)
ORDER BY product;
```

ORA-00936: missing expression
00936. 00000 - "missing expression"
*Cause:
*Action:
Error at Line: 6 Column: 21

The pivot syntax doesn't support a dynamic list of values in the pivot_in_clause.

2011 © EPAM Systems, RD Dep.

Unpivot

- An unpivot does not reverse a PIVOT operation.
- Unpivot rotates data from columns into rows.
- If you are working with pivoted data, an UNPIVOT operation cannot reverse any aggregations that have been made by PIVOT.

2011 © EPAM Systems, RD Dep.

UNPIVOT Syntax

```
SELECT …
FROM  …
UNPIVOT
[INCLUDE|EXCLUDE NULLS]
  ( unpivot_clause
    unpivot_for_clause
    unpivot_in_clause )
WHERE  …
```

unpivot_clause: this clause specifies a name for a column to represent the unpivoted measure values.

unpivot_for_clause: the unpivot_for_clause specifies the name for the column that will result from our unpivot query. The data in this column describes the measure values in the unpivot_clause column;

unpivot_in_clause: this contains the list of pivoted columns (not values) to be unpivoted.

Pivoted mart

```
create view SALES_PROD_QUARTER as
select *
from
  (select PRODUCT, QUARTER, QUANTITY_SOLD from SALES_VIEW)
  PIVOT (SUM(QUANTITY_SOLD)
    for QUARTER in ('01' as Q1, '02' as Q2, '03' as Q3, '04' as Q4)
  )

select *
from SALES_PROD_QUARTER
order by 1
```

PRODUCT	Q1	Q2	Q3	Q4
1.44MB External 3.5" Diskette	6098	5112	6050	5848
128MB Memory Card	1963	2361	3069	2832
17" LCD w/built-in HDTV Tuner	1492	1387	1591	1540
18" Flat Panel Graphics Monitor	1386	1149	1330	1340
256MB Memory Card	1179	1533	1455	1374

Unpivoting dataset

```
select *
from SALES_PROD_QUARTER
  UNPIVOT INCLUDE NULLS (
    QUANTITY_SOLD
    for QUARTER in (Q1, Q2, Q3, Q4)
  )
order by PRODUCT, QUARTER;
```

PRODUCT	QUARTER	QUANTITY_SOLD
1.44MB External 3.5" Diskette	Q1	6098
1.44MB External 3.5" Diskette	Q2	5112
1.44MB External 3.5" Diskette	Q3	6050
1.44MB External 3.5" Diskette	Q4	5848
128MB Memory Card	Q1	1963
128MB Memory Card	Q2	2361
128MB Memory Card	Q3	3069
128MB Memory Card	Q4	2832
17" LCD w/built-in HDTV Tuner	Q1	1492
17" LCD w/built-in HDTV Tuner	Q2	1387
17" LCD w/built-in HDTV Tuner	Q3	1591
17" LCD w/built-in HDTV Tuner	Q4	1540
18" Flat Panel Graphics Monitor	Q1	1386
18" Flat Panel Graphics Monitor	Q2	1149
18" Flat Panel Graphics Monitor	Q3	1330
18" Flat Panel Graphics Monitor	Q4	1340
256MB Memory Card	Q1	1179
256MB Memory Card	Q2	1533
256MB Memory Card	Q3	1455
256MB Memory Card	Q4	1374

Unpivot aliasing options

```
select *
from SALES_PROD_QUARTER
  UNPIVOT INCLUDE NULLS (
    QUANTITY_SOLD
    for QUARTER in (Q1 as 'Quarter 1',
                    Q2 as 'Quarter 2',
                    Q3 as 'Quarter 3',
                    Q4 as 'Quarter 4')
  )
order by PRODUCT, QUARTER;
```

PRODUCT	QUARTER	QUANTITY_SOLD
1.44MB External 3.5" Diskette	Quarter 1	6098
1.44MB External 3.5" Diskette	Quarter 2	5112
1.44MB External 3.5" Diskette	Quarter 3	6050
1.44MB External 3.5" Diskette	Quarter 4	5848
128MB Memory Card	Quarter 1	1963
128MB Memory Card	Quarter 2	2361
128MB Memory Card	Quarter 3	3069
128MB Memory Card	Quarter 4	2832
17" LCD w/built-in HDTV Tuner	Quarter 1	1492
17" LCD w/built-in HDTV Tuner	Quarter 2	1387
17" LCD w/built-in HDTV Tuner	Quarter 3	1591
17" LCD w/built-in HDTV Tuner	Quarter 4	1540
18" Flat Panel Graphics Monitor	Quarter 1	1386
18" Flat Panel Graphics Monitor	Quarter 2	1149
18" Flat Panel Graphics Monitor	Quarter 3	1330
18" Flat Panel Graphics Monitor	Quarter 4	1340
256MB Memory Card	Quarter 1	1179
256MB Memory Card	Quarter 2	1533
256MB Memory Card	Quarter 3	1455
256MB Memory Card	Quarter 4	1374
3 1/2" Bulk diskettes, Box of 100	Quarter 1	3316

How to Unpivot without analytical SQL?

2011 © EPAM Systems, RD Dep.

Traditional Unpivoting

```
select PRODUCT, 'Q1' as QUARTER, Q1 as QUANTITY_SOLD
from SALES_PROD_QUARTER
union
select PRODUCT, 'Q2' as QUARTER, Q2 as QUANTITY_SOLD
from SALES_PROD_QUARTER
union
select PRODUCT, 'Q3' as QUARTER, Q3 as QUANTITY_SOLD
from SALES_PROD_QUARTER
union
select PRODUCT, 'Q4' as QUARTER, Q4 as QUANTITY_SOLD
from SALES_PROD_QUARTER
order by 1, 2
```

PRODUCT	QUARTER	QUANTITY_SOLD
1.44MB External 3.5" Diskette	Q1	6098
1.44MB External 3.5" Diskette	Q2	5112
1.44MB External 3.5" Diskette	Q3	6050
1.44MB External 3.5" Diskette	Q4	5848
128MB Memory Card	Q1	1963
128MB Memory Card	Q2	2361
128MB Memory Card	Q3	3069
128MB Memory Card	Q4	2832

Totals for Pivoted Dataset

```
select *
  from
    (select PRODUCT, CHANNEL, AMOUNT_SOLD from SALES_VIEW) S
    PIVOT (SUM(AMOUNT_SOLD)
      for CHANNEL in (2 as PARTNERS, 3 as DIRECT_SALES,
                      4 as INTERNET_SALES, 9 as TELESALES)) P
order by 1;
```

PRODUCT	PARTNERS	DIRECT_SALES	INTERNET_SALES	TELESALES
1.44MB External 3.5" Diskette	60120.52	137330.01	22167.94	(null)
128MB Memory Card	168783.39	313505.83	89044.53	(null)
17" LCD w/built-in HDTV Tuner	1690316.63	4442061.35	1056793.79	(null)
18" Flat Panel Graphics Monitor	1127568.55	3017888.81	1148972.72	204297.73
256MB Memory Card	127390.3	393111.15	71203.21	(null)
3 1/2" Bulk diskettes, Box of 100	88777.23	249909.31	31518.02	(null)
...	...	...	...	
Xtend Memory	109293	217011.38	40553.93	(null)
Y Box	618499.9	1081050.96	382767.45	11.99
-> (null)	26346342.32	57875260.6	13706802.03	277426.26

Totals for Pivoted Dataset

```
with PIVOTED_DATASET as (
  select *
  from
    (select PRODUCT, CHANNEL, AMOUNT_SOLD from SALES_VIEW) S
    PIVOT (SUM(AMOUNT_SOLD)
      for CHANNEL in (2 as PARTNERS, 3 as DIRECT_SALES,
                      4 as INTERNET_SALES, 9 as TELESALES)) P
)
select 1, P1.PRODUCT, P1.PARTNERS, P1.DIRECT_SALES,
          P1.INTERNET_SALES, P1.TELESALES
from PIVOTED_DATASET P1
union all
select  2, NULL, SUM(P2.PARTNERS), SUM(P2.DIRECT_SALES),
         SUM(P2.INTERNET_SALES), SUM(P2.TELESALES)
from PIVOTED_DATASET P2
order by 1, 2;
```


2011 © EPAM Systems, RD Dep.

Totals for Pivoted Dataset

```
select 1, P1.PRODUCT, P1.PARTNERS, P1.DIRECT_SALES,
          P1.INTERNET_SALES, P1.TELESALES
from PIVOTED_DATASET P1
union all
select  2, NULL, SUM(P2.PARTNERS), SUM(P2.DIRECT_SALES),
          SUM(P2.INTERNET_SALES), SUM(P2.TELESALES)
from PIVOTED_DATASET P2
order by 1, 2;
```

1	PRODUCT	PARTNERS	DIRECT_SALES	INTERNET_SALES	TELESALES
1	1.44MB External 3.5" Diskette	60120.52	137330.01	22167.94	(null)
1	128MB Memory Card	168783.39	313505.83	89044.53	(null)
1	17" LCD w/built-in HDTV Tuner	1690316.63	4442061.35	1056793.79	(null)
1	18" Flat Panel Graphics Monitor	1127568.55	3017888.81	1148972.72	204297.73
1	256MB Memory Card	127390.3	393111.15	71203.21	(null)
1	3 1/2" Bulk diskettes, Box of 100	88777.23	249909.31	31518.02	(null)
...	...	...	...		
1	Xtend Memory	109293	217011.38	40553.93	(null)
1	Y Box	618499.9	1081050.96	382767.45	11.99
2	(null)	26346342.32	57875260.6	13706802.03	277426.26

MTN.BI.03

SQL FOR ANALYSIS

Questions and Answers

Dataset Transformations: Pivoting and Unpivoting

Author: **Aliaksei Belablotski**
Senior Software Engineer
Aliaksei_Belablotski@epam.com

2012 © EPAM Systems, RD Dep.