Data Warehousing and Decision Support, part 2

CS634
Class 23, Apr 27, 2016

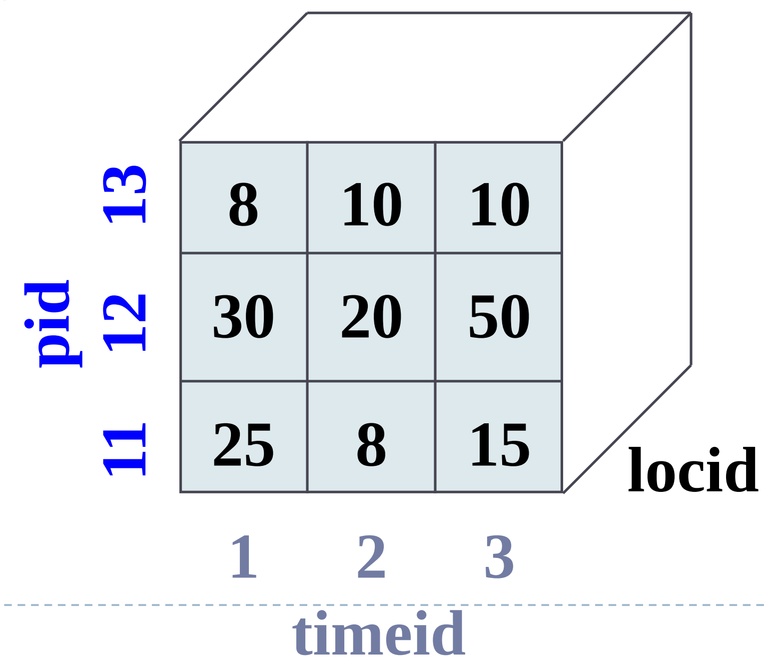
Slides based on "Database Management Systems" 3rd ed, Ramakrishnan and Gehrke, Chapter 25

Multidimensional Data Model

SalesCube(pid, timeid, locid, sales)

- Collection of numeric measures, which depend on a set of dimensions.
  - E.g., measure **sales**, dimensions **Product** (key: pid), **Location** (locid), and **Time** (timeid).
  - Full table, pg. 851

Slice locid=1 is shown:

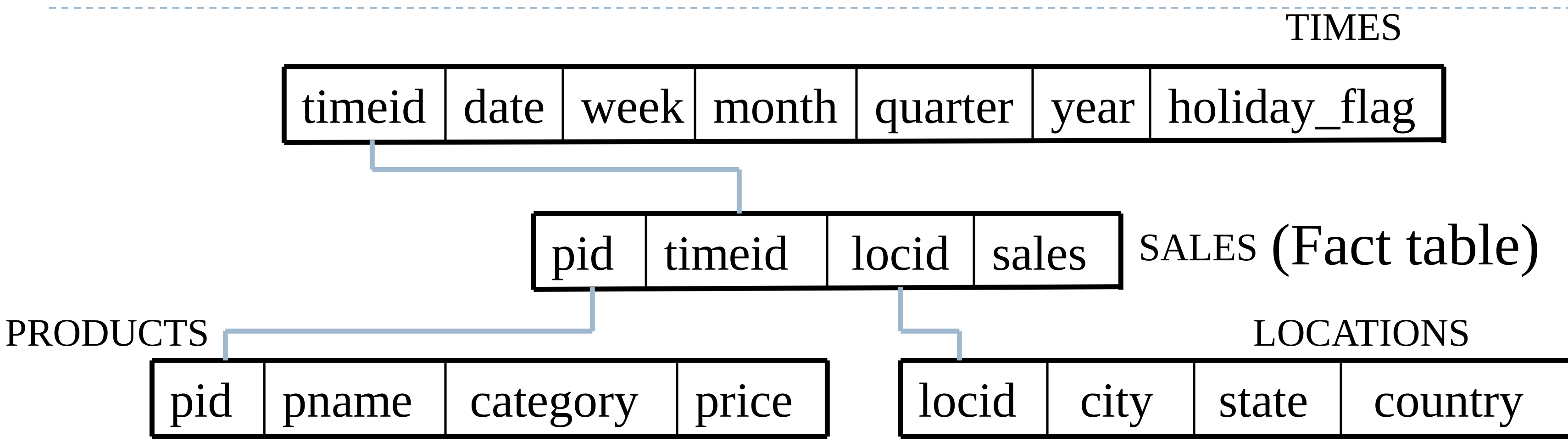


Fact Table

pid	timeid	locid	sales
11	1	1	25
11	2	1	8
11	3	1	15
12	1	1	30
12	2	1	20
12	3	1	50
13	1	1	8
13	2	1	10
13	3	1	10
11	1	2	35

• • •

Star Schema underlying OLAP, used in RDB DW



- Fact/cube table in BCNF; dimension tables not normalized.
  - Dimension tables are small; updates/inserts/deletes are rare. So, anomalies less important than good query performance.
- This kind of schema is very common in DW and OLAP, and is called a star schema; computing the join of all these relations is called a star join.
- Note: in OLAP, this is not what the user sees, it's hidden underneath
- In DW, this is the basic setup, but usually with more dimensions
- Here only one measure, sales, but can have several

Star queries on star schemas

- Oracle definition: a query that joins a large (fact) table to a number of small (dimension) tables, with provided WHERE predicates on the dimension tables to reduce the result set to a very small percentage of the fact table
- The select list has sum(sales), etc., as desired (measures)
- This calculates certain cells of a pivot table, and is used for other analysis too.

```
SELECT store.sales_district, time.fiscal_period,
  SUM(sales.dollar_sales) FROM sales, store, time
WHERE sales.store_key = store.store_key AND
  sales.time_key = time.time_key AND
  store.sales_district IN ('San Francisco', 'Los
  Angeles') AND time.fiscal_period IN ('3Q95', '4Q95',
  '1Q96')
GROUP BY store.sales_district,time.fiscal_period;
```

Star queries

- Oracle: A better way to write the query would be:

(i.e., give the QP a hint on how to do it)

```
SELECT ... FROM sales
 WHERE store_key IN
    ( SELECT store_key FROM store
     WHERE sales_district IN ('WEST', 'SOUTHWEST'))
 AND time_key IN
    ( SELECT time_key FROM time
     WHERE quarter IN ('3Q96', '4Q96', '1Q97'))
 AND product_key IN
    ( SELECT product_key FROM product
     WHERE department = 'GROCERY')
 GROUP BY …;
```

- Oracle will rewrite the query this way if you add the STAR_TRANSFORMATION hint to your SQL, or the DBA has set STAR_TRANSFORMATION_ENABLED

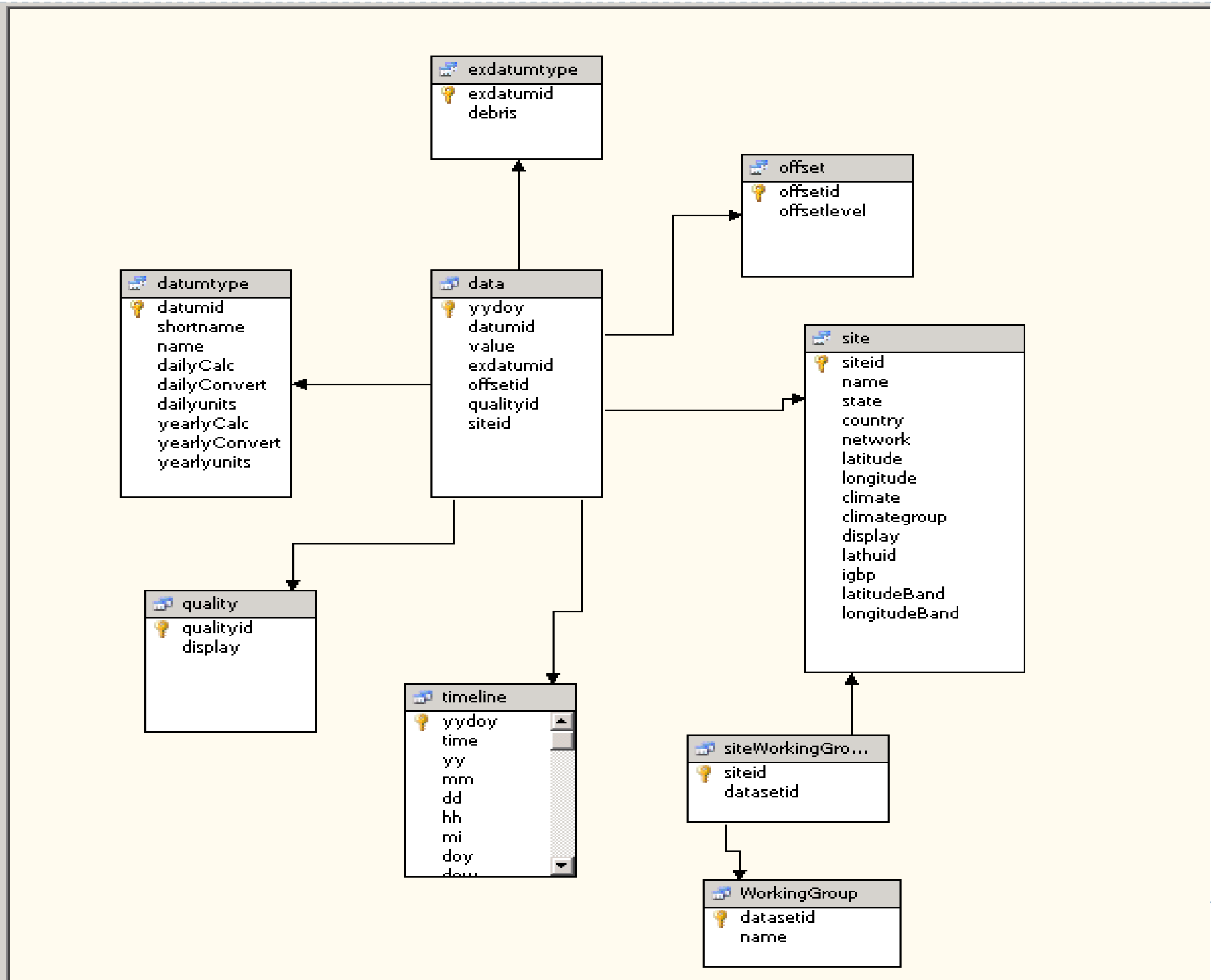
Excel can do Star queries

- Recall GROUP BY queries for individual crosstab entries
- A Star query is of this form, plus WHERE clause predicates on dimension tables such as
  - store.sales_district IN ('WEST', 'SOUTHWEST')
  - time.quarter IN ('3Q96', '4Q96', '1Q97')
- Excel allows “filters” on data that correspond to these predicates of the WHERE clause

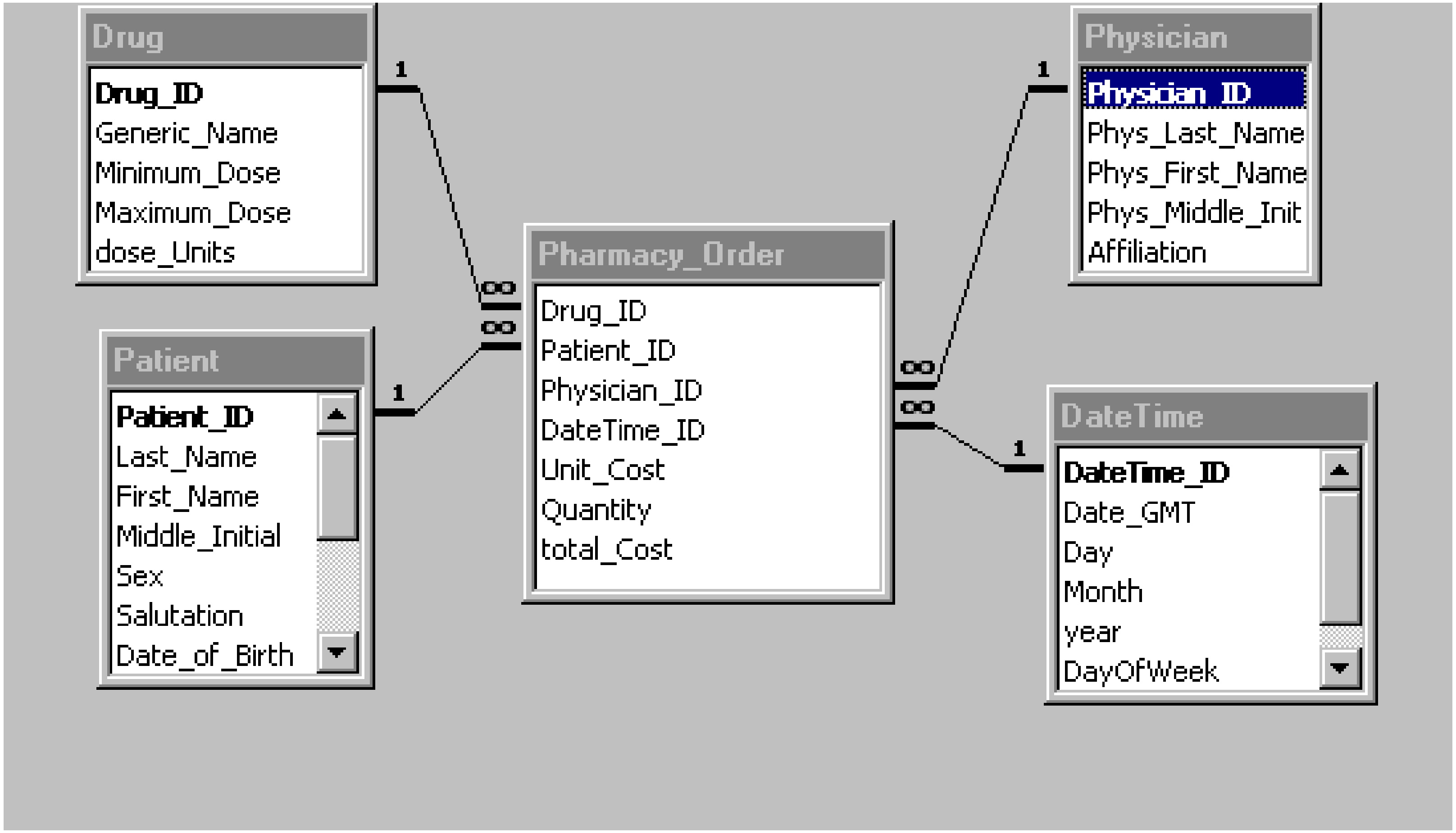
Star schemas arise in many fields

- The dimensions: the facts of the matter
  - What: product
  - Where: store
  - When: time
  - How/why: promotion
- This can be generalized to other subjects: ecology
  - What: temperature
  - Where: location and height
  - When: time
  - How/why: quality of data
  - Which: working group

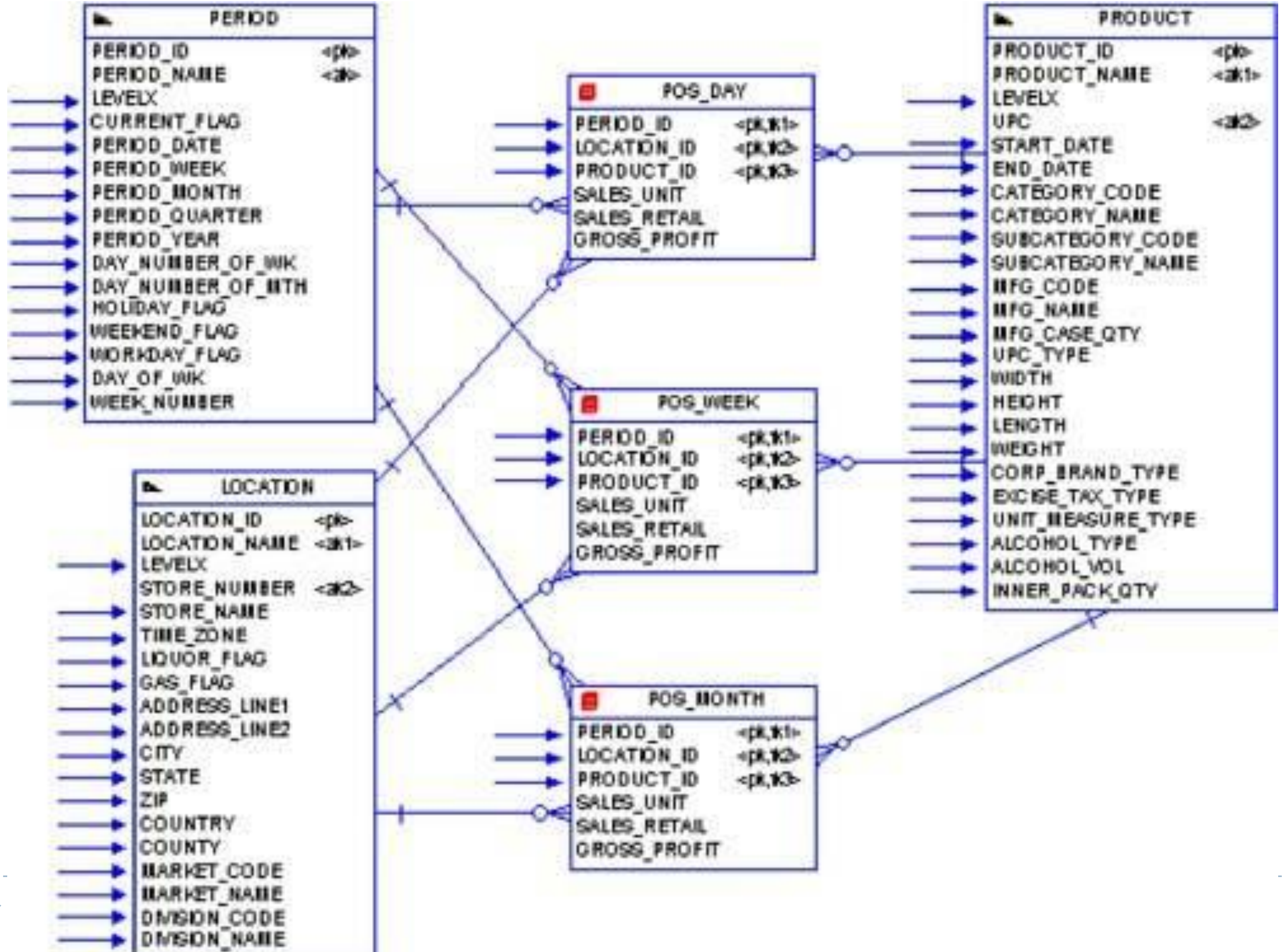
Star schema from ecology



Star Schema from Medicine



What's this?



PERIOD
PERIOD_ID <pk>
PERIOD_NAME <ak>
LEVELX
CURRENT_FLAG
PERIOD_DATE
PERIOD_WEEK
PERIOD_MONTH
PERIOD_QUARTER
PERIOD_YEAR
DAY_NUMBER_OF_WK
DAY_NUMBER_OF_MTH
HOLIDAY_FLAG
WEEKEND_FLAG
WORKDAY_FLAG
DAY_OF_WK
WEEK_NUMBER
POS_DAY
PERIOD_ID <pk,fk1>
LOCATION_ID <pk,fk2>
PRODUCT_ID <pk,fk3>
SALES_UNIT
SALES_RETAIL
GROSS_PROFIT
PRODUCT
PRODUCT_ID <pk>
PRODUCT_NAME <ak1>
LEVELX
UPC <ak2>
START_DATE
END_DATE
CATEGORY_CODE
CATEGORY_NAME
SUBCATEGORY_CODE
SUBCATEGORY_NAME
MFG_CODE
MFG_NAME
MFG_CASE_QTY
UPC_TYPE
WIDTH
HEIGHT
LENGTH
WEIGHT
CORP_BRAND_TYPE
EXCISE_TAX_TYPE
UNIT_MEASURE_TYPE
ALCOHOL_TYPE
ALCOHOL_VOL
INNER_PACK_QTY
POS_WEEK
PERIOD_ID <pk,fk1>
LOCATION_ID <pk,fk2>
PRODUCT_ID <pk,fk3>
SALES_UNIT
SALES_RETAIL
GROSS_PROFIT
LOCATION
LOCATION_ID <pk>
LOCATION_NAME <ak1>
LEVELX
STORE_NUMBER <ak2>
STORE_NAME
TIME_ZONE
LIQUOR_FLAG
GAS_FLAG
ADDRESS_LINE1
ADDRESS_LINE2
CITY
STATE
ZIP
COUNTRY
COUNTY
MARKET_CODE
MARKET_NAME
DIVISION_CODE
DIVISION_NAME
POS_MONTH
PERIOD_ID <pk,fk1>
LOCATION_ID <pk,fk2>
PRODUCT_ID <pk,fk3>
SALES_UNIT
SALES_RETAIL
GROSS_PROFIT

Indexes related to data warehousing

- Example Bitmap index: if interested, see Bitmap Indexes

Bit-vector:
1 bit for each possible value.
Many queries can be answered using bit-vector ops!

F M

sex	custid	name	sex	rating	rating
10	112	Joe	M	3	00100
10	115	Ram	M	5	00001
01	119	Sue	F	5	00001
10	112	Woo	M	4	00010

Indexing for DW, cont: Join Indexes

- Consider the join of Sales, Products, Times, and Locations,
  - A join index can be constructed to speed up such joins. The index contains [s,p,t,l] if there are tuples (with sid) s in Sales, p in Products, t in Times and l in Locations that satisfy the join conditions.
  - Can do one dimension column at a time, put <f_rid, c1> in c1's join index, where f_rid is the fact table RID and c1 the dimension-table value we're interested in.
  - It's as if c1 is an additional column of the fact table, with a normal index <c1, f_rid> to allow finding rows with certain c1.
  - Related topic: materialized views, cover later.
  - Bitmap indexes are a good match here…

Organizing huge fact tables

- If the query retrieves 1000 or even 10,000 rows from the fact table, it's still pretty fast without special organization (10,000 random i/os = 100 seconds, faster on RAID)
- The problem is that retrieving 100,000 random rows in a huge fact table (itself billions of rows) means 100,000 page i/os (1000 seconds) unless we do something about the fact table organization
- Traditional solution for scattered i/o problem: clustered table.
- But what to cluster on—time? Product? Store?
- Practical simple answer: time, so can insert smoothly and extend the table, delete old stuff in a range
- But we can do better…

Well, how does Teradata do it?

By multi-dimensional partitioning (toy example):

```
CREATE TABLE Sales (storeid INTEGER NOT NULL,
productid INTEGER NOT NULL, salesdate DATE NOT NULL,
sales DECIMAL(13,2), totalsold INTEGER, note
VARCHAR(256), PRIMARY KEY (storeid, productid,
salesdate)) PARTITION BY
( RANGE_N(salesdate BETWEEN DATE '2012-01-01' AND
     DATE '2016-12-31' EACH INTERVAL '1' YEAR),
  RANGE_N(storeid BETWEEN 1 AND 300 EACH 100),
  RANGE_N(productid BETWEEN 1 AND 400 EACH 100));
```

- This table is first partitioned by year based on salesdate.
- Next, within each year the data will be partitioned by storeid in groups of 100.
- Finally, within each year/storeid group, the data will be partitioned by productid in groups of 100.
- One cell: sales in 2015 for one group of 100 stores and one group of 100 productids (365x100x100 rows)

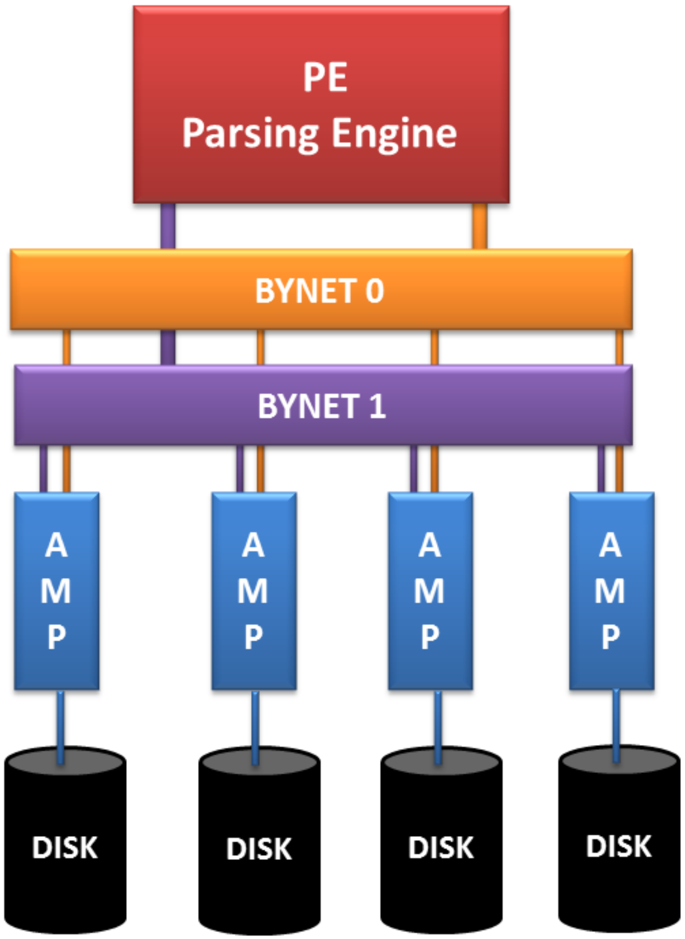
How multi-dimensional partitioning speeds up queries

- Suppose 500 stores, 500 products
- So the partitioning sets up 5 productid ranges, 5 storeid ranges, as well as 5 time ranges
  - Query on 2016 reads only 1/5 of data
  - Query on store 25, all years, reads 1/5 of data
  - Query on store 25 for 2016 reads only 1/5*1/5 = 1/25 of data
  - Query on store 25, product 44 for 2016 reads only 1/5*1/5*1/5 of data
  - This assumes the query processor is smart about partitions…
- Also helps with huge delete needed when a year gets archived: just drop the year's partition

Teradata System

Partitioning puts a set of cube cells on each node

Star query pulls data from a subset of cells scattered across nodes



AMP : Access Module Processor

Partitioning: physical organization

- Not covered by SQL standard
- So we have to look at each product for details
- But similar basic capabilities
- Oracle says start thinking about partitioning if your table is over 2GB in size.
- Another way of saying it: start thinking about partitioning if your table and indexes can't fit in the database buffer pool.
  - Don't forget to size up the buffer pool to, say, ½ memory when you install the database!
- Partitioning can also be used with Hadoop

Partitioning Example

- Consider a warehouse with 10TB of data, made up of 2 TB per year of sales data, for 5 years.
- End of year: has grown to 12 TB, need to clean out oldest 2TB, or put it in archive area.
- Or do this every month.
- Either way, massive delete. Could delete rows on many pages, lowering #rows/page, thus query performance. Will take a long time for a big table.
- With partitioning, we can just drop a partition, create a new one for the new year/month. All the surviving extents still have the same rows.
- So most warehouses are partitioned by year or month.

Partitioning

- The following works in Oracle and mysql:

```
create table sales (year int,  yearday int,
            product varchar(10),sales decimal(10,2))
     partition by range (year)
    (partition p1 values less than (2010),
     partition p2 values less than (2011),
     partition p3 values less than (2012),
     partition p4 values less than (2013);
```

- Here the sales table is created with 4 partitions. Partition p1 will contain rows of year 2009 and earlier. Partition p2 will contain rows of year 2010, and so on..

Partitioning by time

- Considering example table partitioned by year
- So if we're interested in data from a certain year, the disks do one seek, then read, read, read…
  - Much more efficient than if all the years are mixed up on disk. Partitioning is doing a kind of clustering.
  - We could partition by month instead of by year and get finer-grained clustering
- To add a partition to sales table give the following command.

```
alter table sales
    add partition p6 values less than (2014);
```

- Similarly can drop a partition of old data

Oracle Partitioning

- In Oracle, each partition has its own extents, like an ordinary table or index does. So each extent will have data all from one year.
- We read-mostly data, we should make sure the extents are at least 1MB, so say 16MB in size. In Oracle we could create the one tablespace with a default storage clause early in our setup
- Could be across two RAID sets, each with 1MB stripes

```
CREATE TABLESPACE dw_tspace
 DATAFILE 'fname1' SIZE 3000G,'fname2' SIZE 3000G
 DEFAULT STORAGE (INITIAL 16M NEXT 16M);
```

Types of Partitioning

- In Oracle and mysql you can partition a table by
  - Range Partitioning (example earlier)
  - Hash Partitioning
  - List Partitioning (specify list of key values for each partition)
  - Composite Partitioning (uses subpartitions of range or list partitions)
- Much more to this than we can cover quickly, but plenty of documentation online
- Idea from earlier: put cells of cube/fact table together in various different places. Need last item in above list.
- But Oracle docs/tools shy away from 3-level cases (they do work, because I've done it)

Cube-related partitioning in Oracle

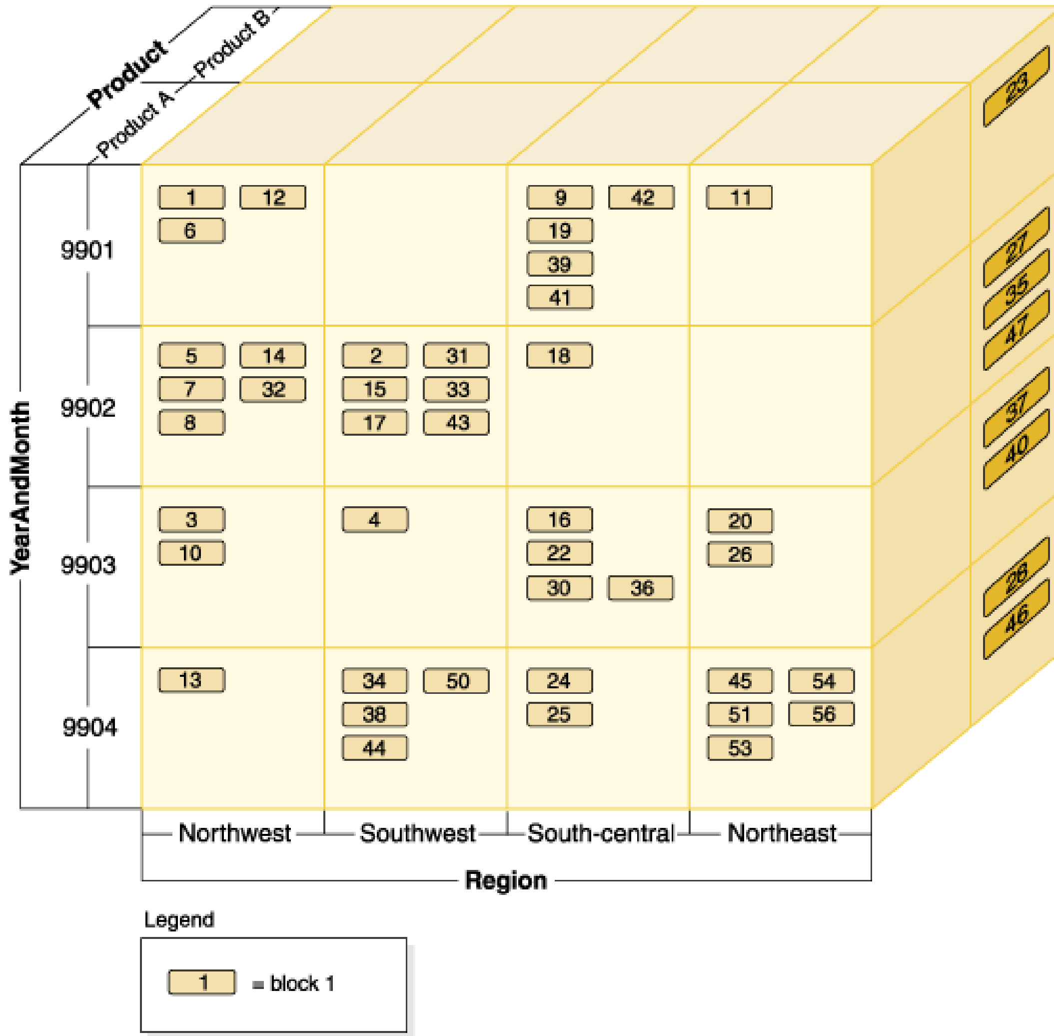
```
create table sales (year int, dayofyear int,  product varchar(10),
                    sales decimal(10,2))
PARTITION BY RANGE (year)
SUBPARTITION BY  HASH(product) SUBPARTITIONS 8
   (partition p1  values less than (2008),
    partition p2 values less than (2009),
    partition p3 values less than (2010),
    partition p4 values less than (2011),
    partition p5 values less than (2012);
));
```

- Here have 40 partitions
- Subpartitions are also made of extents (in Oracle), so now in one extent we have a certain subset of products in a certain year.
- With partitions and subpartitions, we are getting a kind of multi-dimensional clustering, by two dimensions.

DB2's Multi-dimensional Clustering (MDC)

Example 3-dim clustering, following cube dimensions.

Note this is not partitioning, but can be used with partitioning



Characteristics of a mainstream DB2 data warehouse fact table, from DB2 docs

- A typical warehouse fact table, might use the following design: Create data partitions on the Month column.
- Define a data partition for each period you roll-out, for example, 1 month, 3 months.
- Create MDC dimensions on Day and on 1 to 4 additional dimensions. Typical dimensions are: product line and region.

Example DB2 partition/MDC table

```
CREATE TABLE orders (YearAndMonth INT,
     Province CHAR(2), sales DECIMAL(12,2))
 PARTITION BY RANGE (YearAndMonth)
     (STARTING 9901 ENDING 9904 EVERY 2)
 ORGANIZE BY (YearAndMonth, Province);
```

- Partition by time for easy roll-out
- Use MDC for fast cube-like queries
  - All data for yearandmonth = '9901' and province='ON' (Ontario) in one disk area
  - Note this example has no dimension tables
  - Could use prodid/1000, etc. as MDC computed column—but does the QP optimize queries properly for this?

Partition Pruning

- The QP needs to be smart about partitions/MDC cells
- From Oracle docs, the idea: "Do not scan partitions where there can be no matching values".
- Example: partitions of table t1 based on region_code:

  PARTITION BY RANGE(region_code)
  (PARTITION p0 VALUES LESS THAN (64),
  PARTITION p1 VALUES LESS THAN (128),
  PARTITION p2 VALUES LESS THAN (192),
  PARTITION p3 VALUES LESS THAN MAXVALUE);

  Query:

  ```
  SELECT fname, lname, region_code, dob FROM t1
     WHERE region_code > 125 AND region_code < 130;
  ```
- QP should prune partitions p0 (region_code too low) and p3 (too high).
- But the capability is somewhat fragile in practice.

Partition Pruning is fragile

- From dba.stackexchange.com:
- The problem with this approach is that partition_year must be explicitly referenced in queries or partition pruning (highly desirable because the table is large) doesn't take effect. (Can't ask users to add predicates to queries with dates in them)
- Answer:
- … Your view has to apply some form of function to start and end dates to figure out if they're the same year or not, so I believe you're out of luck with this approach.
- Our solution to a similar problem was to create materialized views over the base table, specifying different partition keys on the materialized views.
- So need to master materialized views to be an expert in DW.

Parallelism is essential to huge DWs

Table 1: Parallelism approaches taken by different data warehouse DBMS vendors, from "How to Build a High-Performance Data Warehouse" by David J. DeWitt, Ph.D.; Samuel Madden, Ph.D.; and Michael Stonebraker, Ph.D.
(I've added bold for the biggest players, green for added entries)

Shared Memory (least scalable)	Shared Disk (medium scalable)	Shared Nothing (most scalable)
Microsoft SQL Server PostgreSQL MySQL	**Oracle RAC** Sybase IQ	**Teradata** **IBM DB2** Netezza EnterpriseDB (Postgres) Greenplum Vertica MySQL Cluster SAP HANA

Shared-nothing vs. Shared-disk