**44-560 Advanced Topics in Database Systems**

**Introduction to Data Warehousing**

1. Consider the dimensional model given below for a grocery store chain.



Assume that

* + there are 500 stores;
  + there are 200,000 products;
  + data will be collected for 5 years (5 \* 365 days);
  + on any given day at any given store, approximately 5,000 different items are sold.

Answer the questions below. Assume the average field size is 10 bytes.

* 1. How many rows will **DateDim** contain?
  2. How many rows will **ProductDim** contain?
  3. How many rows will **StoreDim** contain?
  4. How many total bytes of storage will the dimension tables require?
  5. How many rows will the fact table contain?
  6. How many bytes of storage will the fact table contain?

1. Suppose we use a coarser grain and maintain data for stores and dates but not individual products. For each store, we lump all product sales together each day. The new model looks like this:



Answer the questions below. Assume the average field size is 10 bytes.

* 1. How many rows will **DateDim** contain?
  2. How many rows will **StoreDim** contain?
  3. How many total bytes of storage will the dimension tables require?
  4. How many rows will the fact table contain?
  5. How many bytes of storage will the fact table contain?

1. Suppose we maintain data for dates and products, but not for individual stores. Then our model will look like this:



Answer the questions below, assuming

* + the average field size is 10 bytes, and
  + on a given day, across all stores in the chain, approximately 10,000 different products are sold.
  1. How many rows will **DateDim** contain?
  2. How many rows will **ProductDim** contain?
  3. How many total bytes of storage will the dimension tables require?
  4. How many rows will the fact table contain?
  5. How many bytes of storage will the fact table contain?

1. Construct a star schema for a national bookstore chain. The relevant dimensions and attributes are as follows:

* **BookDim**. Attributes: ISBN, title, publisher, year of publication.
* **StoreDim**. Attributes: store number, store name, street address, city, state, zip, name of manager, phone number.
* **DateDim**. Attributes: day, month, year, quarter, holiday flag, and weekend flag.

Facts to be recorded are number of copies sold, dollar amount, and dollar cost.

The chain carries approximately 25,000 titles and there are 150 stores nationwide. Each store carries an average of 15,000 titles. On any given day, at any given store, approximately 200 different titles are sold. Data will be stored for 5 years (5 \* 365 days).

For the problems below, assume the average field size is 5 bytes.

* 1. Place your star schema in the space below. Remember to use surrogate keys for the primary keys of the dimension tables.
  2. How many rows will the **BookDim** table contain?
  3. Find the size (in bytes) of **BookDim**.
  4. How many rows will the **StoreDim** table contain?
  5. Find the size (in bytes) of **StoreDim**.
  6. How many rows will the **DateDim** table contain?
  7. Find the size (in bytes) of **DateDim**.
  8. How many rows will the fact table contain when all five years of data has been stored?
  9. Find the size (in bytes) of the fact table when all five years of data has been stored.

1. Construct a star schema that enables a university with multiple campuses to track enrollments on the various campuses. The relevant dimensions and attributes are:

* **CampusDim**. Attributes: CampusID, campus name, city, state, zip.
* **StudentDim**. Attributes: StudentID, name of student, street address, city, state, zip, majorID (assume a student has only one major), major name.
* **DateDim**. Attributes: semester (fall, spring, or summer), year.

Facts to be recorded are number of credit hours the student is enrolled in on each campus and the dollar amount of revenue generated for a campus by the student’s enrollment.

The university has 5 campuses. Total student enrollment for the five campuses is approximately 100,000 each semester. Each semester (fall, spring, and summer), about 30,000 students transfer or graduate and are replaced by new students. In any given semester, about 20,000 students are enrolled at two different campuses. Data will be stored for 10 years.

Note that when a student transfers or graduates, that student is *not* removed from the database, because we are storing historical data.

For the problems below, assume the average field size is 10 bytes.

* 1. Place your star schema in the space below. Remember to use surrogate keys for the primary keys of your dimension tables.
  2. How many rows will the **CampusDim** table contain?
  3. How many rows will the **StudentDim** table contain at the end of the ten year period?
  4. How many rows will the **DateDim** table contain?
  5. Find the size (in bytes) of **StudentDim**.
  6. How many rows will the fact table contain when all ten years of data has been stored?
  7. Find the size (in bytes) of the fact table when all ten years of data has been stored.