EXTENDED BAKERY Database

The EXTENDED BAKERY database/dataset is a collection of data documenting the operations of a chain of bakery stores.

The Data

A *nameless bakery chain* has a number of stores in the West Coast states of California, Oregon and Washington, as well as in nearby Nevada and Arizona. The stores make and sell a variety of bakery items (goods), and a number of drinks. Each store has a number of employees working at it. Customers come to the stores and make purchases. Each purchase can include multiple items purchased (e.g., a cup of tea and a cookie) and multiple "copies" of the same item (e.g., two almond twists). Each purchase is associated with an employee who rung it.

It is assumed that

- All stores have exactly the same item availability (i.e., all drinks and pastries are equally available at all locations).
- All stores have the same prices on all items.
- Prices are stable and do not change in the time horizon for which the data is available.
- The database does not track sales taxes.

The Database

The database contains the following data:

- **Goods.** The `Goods` table stores information about the bakery offerings. Each item on the menu has a unique id, a name that consists of two parts: the Flavor ("almond", "chocolate", "napoleon") and Food type ("twist", "cake", "cookie"). We also store information on whether the item is a pastry or a drink, and the price of the item.

- **Locations.** The `Location` table stores information about individual bakery stores. Each store has a unique store number, as well as a full address composed of street address, city, zip code and state.

- **Employees.** The `Employee` table stores information about employees working in the store. Employees are uniquely identified by their employee id. The table stores their full name (first name, last name), their dates of employment (hire date and fire date), name of their position (barista, cashier, manager), and whether the employee is *full-time* or *part-time*.

- **Receipts.** The `Reciepts` table combined with the `Items` table (see below) documents the actual purchases made at the stores. The `reciepts` table stores the receipt number (unique id), sale date, of of the store where the purchase was made and employee Id of the employee who rang the purchase. There is also a flag stating whether the purchase was made on a weekend[1] and a flag indicating whether the purchase was made with cash or credit card[2]

- **Items.** The itemized list of purchases for each repeipt is stored in the `Items` table. The table stores the receipt number, id of the purchased good and the quantity purchased. There may be multiple goods purchased on a single receipt, but each receipt has only one entry for any single item (i.e., a purchase of three almond twists is recorded as a single tuple in the `Items` table with the `Quantity` attribute set to 3).

The Files

The `EXTENDED BAKERY` dataset consists of the following files:

Filename	Explanation
`EB-setup.sql`	database setup script (`CREATE TABLE` statements)
`EB-cleanup.sql`	`DROP TABLE` statements
`EB-insert.sql`	runs all other `...-insert.sql` scripts
`EB-build-employee.sql`	inserts data into the `Employees` table
`EB-build-goods.sql`	inserts data into the `Goods` table
`EB-build-items.sql`	inserts data into the `Items` table
`EB-build-location.sql`	inserts data into the `Location` table
`EB-build-receipts.sql`	inserts data into the `Receipts` table
`EB-verify.sql`	reports tuple counts in all database tables

These files can be downloaded one-by-one, or as a single zip file from the following URL:

http://wiki.csc.calpoly.edu/datasets/wiki/ExtendedBakery1000

Please make sure you read the README file for the EXTENDED BAKERY dataset (available from http://wiki.csc.calpoly.edu/datasets/wiki/ExtendedBakery) carefully before you start your work. To set up your database, run the following commands:

[1]This can be deduced from the sales date, but the weekend flag is there for simplicity.
[2]The bakery accepts no other forms of payment.

```
> $ORACLE_HOME/bin/sqlplus  <yourAccount>@ora10g

SQL> start EB-setup.sql

...

SQL> start EB-insert.sql
```

Note, the last command, may take some time. You can run `start EB-verify.sql` command to verify that all data is there. File `results-of-verify.txt` shows the tuple counts.

Variants

There are four actual sets of data in the EXTENDED BAKERY dataset. They are different by content and size. The table below outlines the four sets, and shows the urls from which they can be downloaded.

Set of data	Size	URL
EXTENDED BAKERY-1000	1,000 receipts	http://wiki.csc.calpoly.edu/datasets/wiki/ExtendedBakery1000
EXTENDED BAKERY-5000	5,000 receipts	http://wiki.csc.calpoly.edu/datasets/wiki/ExtendedBakery5000
EXTENDED BAKERY-20K	20,000 receipts	http://wiki.csc.calpoly.edu/datasets/wiki/ExtendedBakery20K
EXTENDED BAKERY-75K	75,000 receipts	http://wiki.csc.calpoly.edu/datasets/wiki/ExtendedBakery75K