```
create table dim_subscription_status
( subscription_status_key    int not null
, subscription_status_code   char(2)
, description                varchar(50)
, source_system_code         tinyint
, create_timestamp           datetime
, update_timestamp           datetime
, constraint pk_dim_subscription_status
  primary key clustered (subscription_status_key)
)

create table dim_channel
( channel_key         int not null
, name                varchar(20)
, description         varchar(50)
, start_date          smalldatetime
, end_date            smalldatetime
, status              varchar(10)
, source_system_code  tinyint    not null
, create_timestamp    datetime   not null
, update_timestamp    datetime   not null
, constraint pk_dim_channel
  primary key clustered (channel_key)
)
```

While we are on the permission management topic, it is relevant to talk about *opting out*, or revoking the permission. It is mandatory that the customer is able to unsubscribe or revoke their permission at any time from any campaign. If it is an e-mail campaign, we need to have a link on every campaign pointing to an unsubscribe page. This page will update the front-end database that will then be propagated to the data warehouse. As a result, the subscription will be deactivated, in other words, by updating the subscription end date and status. It is important that the propagation of unsubscription information to the CRM data warehouse is frequent, ideally intraday or at least once a day, so that we can make sure customers who have revoked their permissions do not get subsequent campaigns. If it is a general or specific permission, the ETL will update the corresponding permission column on the customer dimension from Y to N.

Delivery and Response Data

Campaign delivery data is about whether the campaign that we sent is delivered to the intended recipient. For example, if the campaign is delivered using an e-mail message, it may not get into the target mailbox because the mailbox is full. Campaign response data is about the reaction from the customer after receiving our campaign. For example, they may be interested in a particular offer in the e-mail and click it to purchase the product.

To understand campaign delivery and response data, let's go through the process of campaign delivery. Take the Amadeus Entertainment Music Weekly Newsletter, which has 100,000 recipients, as an example. After the campaign is created, it is executed and sent to the target

recipients. Let's say that when the campaign was executed, there were 100,000 target recipients. Let's say 1,000 of these target recipients were filtered out because of the banned list table (refer to the earlier discussion of negative permissions), and 2,000 were filtered out because the recipient e-mail addresses have bounced four times. So, 97,000 actually went out: 90,000 by e-mail; 4,000 by post; 2,000 by RSS; and 1,000 by text message. Out of the 90,000 e-mail messages, say 86,000 reached the intended recipients, and 4,000 were undelivered (because the e-mail domain did not exist, because of an invalid mailbox name, because the mailbox was full, and so on). Out of the 86,000 e-mail messages that were delivered, 25,000 were opened/read, and 5,000 were clicked through; that is, the customers went to the web pages specified in the e-mail campaign. Out of these 5,000 customers, 500 made an order (purchased some products). Figure 14-3 illustrates the figures in this process.
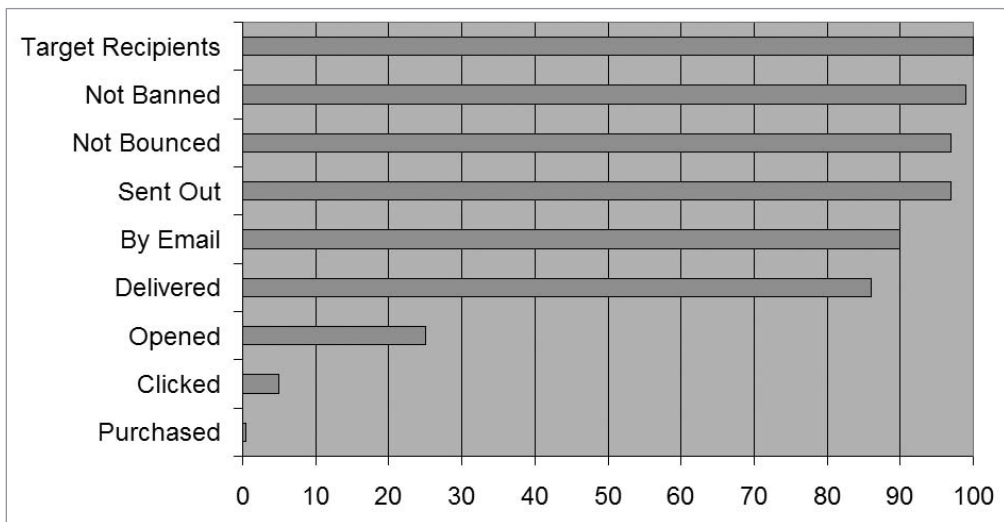

Figure 14-3. *Number of customers involved in each step of campaign delivery process*

The number of e-mail messages sent out, the number of messages delivered, the number of messages not delivered (bounced), and the number of e-mail messages not sent out (either because of banned list or bounce list) are called *campaign delivery data*, and the number of e-mail messages opened/read, the number of messages clicked by the customers, and the number of customers purchased some products are called *campaign response data*. Both campaign delivery data and campaign response data are important because they indicate the success or failure of the campaign; for example, if the open rate and click-through rate (CTR) are high, the campaign is successful.

The open and click-through rates vary from industry to industry. Amongst other things, they depend on whether we implement double opt-in (cleaner list), they depend on who the target audience is (business to business or business to consumer), and they depend on whether the segmentation is specific or general. If you still want a rough indicator, for business to consumer with double opt-in and specific segmentation, you can expect an open rate from 5 percent to 15 percent.

It would be useful to understand how campaign delivery and response data can be collected by the CRM system. In other words, how do we know that an e-mail has been opened

or read? How do we know that the customer has clicked certain links in the e-mail? How do we detect which customer purchased which products and identify which campaign initiated the purchase? The techniques and methods of obtaining delivery and response data within the CRM systems vary greatly from application to application. If it is an e-mail campaign, the delivery status can be obtained from SMTP response codes, which is the number sent back by the target e-mail server, such as 250 for delivered, 530 for access denied, 552 for exceeding the mailbox limit, 521 for a domain that does not except e-mail, 421 for a domain that's not valid/available, and so on. Sometimes the interpretation of the response codes is not so straightforward; for example, code 250 can mean several things, because it actually means that the requested action is done OK, so the meaning depends on what the requested action was. For a text message, we need to use the appropriate protocol instead of SMTP, such as an SMS delivery report that indicates whether the text message has reached the target cell phone or cannot be delivered.

Now that you have an understanding of campaign delivery and response data, I'll cover the pros and cons of implementing them in a data warehouse. The benefit of using a data warehouse (as opposed to an operational system such as an ERP, CDI, or ODS) is that we can use the delivery and response data for campaign segmentation. This is because the delivery and response data will be located on the same platform as the other data used for a selection query (for example, customer data, shopping/purchase/activity data, and permission data) during the campaign selection and segmentation process. For example, we might select customers who have received a campaign e-mail about certain music product types but did not open the e-mail *or* opened the e-mail but did not click any offer in the e-mail. We can then target these groups with specific follow-up campaigns encouraging them to go to our web site to look at the products.

The second benefit of implementing campaign delivery and response data in the data warehouse is that we have better performance when loading data and performing analysis. The amount of campaign delivery and response data can be billions of rows within a few years. For example, we could have 35 campaigns that send e-mails to 20 million target subscribers every week, which equates to 1 billion a year. The grain of campaign delivery and response data is the number of target recipients. That means that for every message created by the campaign, we will have a row, whether the message is sent or not. Because of the volume, we can take advantage of a data warehouse platform, which handles a large volume of data better than operational systems. This is especially true when querying the campaign response data for customer analysis and purchase analysis, such as when analyzing customers who have clicked but not purchased or when performing demographic analysis against customers who have received certain types of campaigns in the last six months.

So, those are two benefits of implementing campaign delivery and response data in a dimensional data warehouse. But sometimes we are not able to do that if our CRM application is a packaged application. For example, the CRM application may not work with a dimensional data warehouse; it may work only with a highly normalized operational database. In this case, we could load the campaign delivery and response data from the normalized CRM database into a dimensional data warehouse for analysis.

Figure 14-4 shows the design of an implementation of CRM campaign delivery and response data in a data warehouse. It consists of a campaign delivery and response fact table with six dimensions.
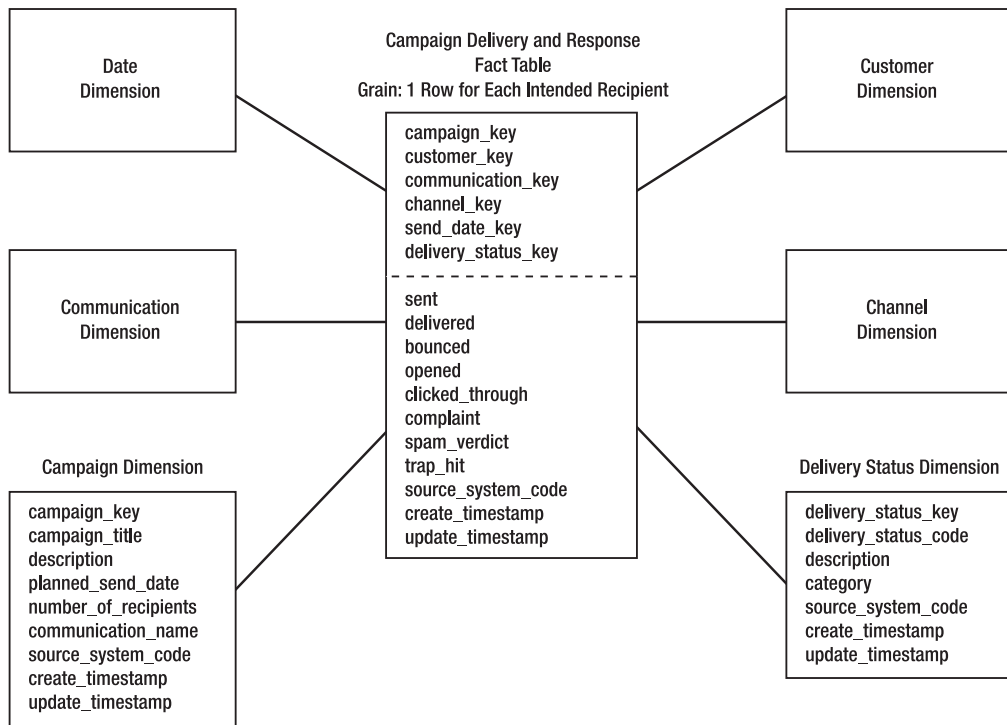
Figure 14-4. *Schema design for implementing campaign delivery and response data in a data warehouse*

Let's go through the schema in Figure 14-4. It has one fact table and six dimension tables: date, communication, campaign, customer, channel, and delivery status. The grain of the fact table is one row for each intended recipient in the campaign. For example, say we have 10 campaigns that were sent last week, each containing 20,000 customers. In this case, we have 200,000 rows in this fact table.

I discussed this schema in Chapter 5. But to refresh, let's go through each key column on the fact table. The campaign key identifies which campaign was sent, for example: Weekly France Music Newsletter 01/14/2008. The customer key identifies to which customer the campaign was sent. The communication key identifies the communication to which the campaign belongs. For example, the Weekly France Music Newsletter 01/14/2008 campaign is an instance of the Weekly France Music Newsletter communication. The channel key indicates how the campaign was sent, such as via e-mail. The send date indicates when the campaign was sent. The delivery status indicates whether the campaign was delivered successfully to the intended recipient.

Listing 14-3 shows the DDL of Figure 14-4.

Listing 14-3. *DDL for Implementing Campaign Delivery and Response Data in a Data Warehouse*

```
create table fact_campaign_result
( campaign_key          int not null
, customer_key          int not null
, communication_key     int not null
, channel_key           int not null
, send_date_key         int not null
, delivery_status_key   int not null
, sent                  int
, delivered             int
, bounced               int
, opened                int
, clicked_through       int
, complaint             int
, spam_verdict          int
, trap_hit              int
, source_system_code    tinyint
, create_timestamp      datetime
, update_timestamp      datetime
, constraint pk_fact_campaign_result
  primary key clustered
  ( campaign_key, send_date_key, customer_key )
)

create table dim_campaign
( campaign_key        int not null
, campaign_title      varchar(50) not null
, description         varchar(100)
, planned_send_date   smalldatetime
, number_of_recipients int
, communication_name  varchar(50)
, source_system_code  tinyint    not null
, create_timestamp    datetime   not null
, update_timestamp    datetime   not null
, constraint pk_dim_campaign
  primary key clustered (campaign_key)
)

create table dim_delivery_status
( delivery_status_key   int not null
, delivery_status_code  int not null
, description           varchar(50)
, category              varchar(20)
, source_system_code    tinyint    not null
, create_timestamp      datetime   not null
, update_timestamp      datetime   not null
, constraint pk_dim_delivery_status
  primary key clustered (delivery_status_key)
)
```

After we load campaign delivery and response data into the data warehouse, we can use it for various analyses, including calculating the campaign result statistics such as open rate, click-through rate, and purchase rate by any demographic attributes, by campaign type, and by product hierarchy. For example, we can compare the open, click-through, and purchase rates for different regions across all product categories. Or we can analyze them by occupation group, age group, and interest group. We can analyze how many customers clicked each link on the campaign we sent. We can compare the campaign response rates with the statistics from previous campaign sends, with the company-wide average and with the industry standards. Figure 14-5 shows the open rates across different regions by product category for a certain period.
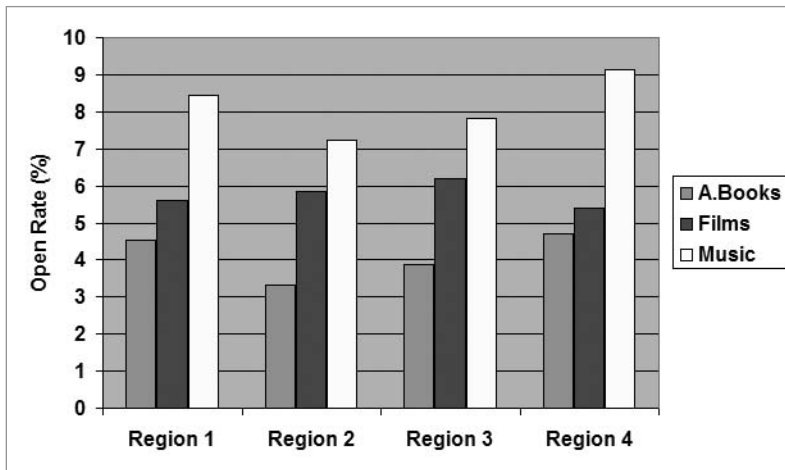

Figure 14-5. *Open rate across different regions by product category*

We can also use the campaign response data to analyze the effectiveness of certain product promotion campaigns by comparing the open rate and click-through rate across product types, as shown in Figure 14-6. The bubble size indicates the campaign size (the number of target recipients). The bigger the bubble, the bigger the campaign.
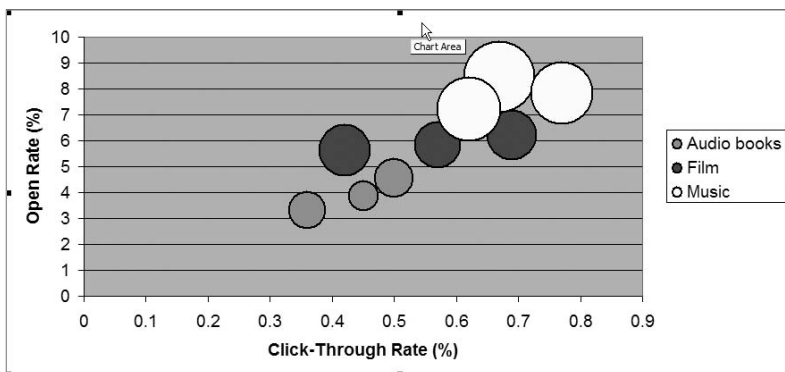

Figure 14-6. *Open rate, click-through rate, and campaign size across different product categories*

Campaign delivery data enables us to analyze the bounce rate across different e-mail domains and understand the reasons for the rate. The bounce rate is the number of e-mail messages not delivered divided by the number of e-mail messages sent. E-mail domains are the domain name of the customer e-mail address. For example, Figure 14-7 shows Amadeus Entertainment's average bounce rate for the top six e-mail domains in a particular week.
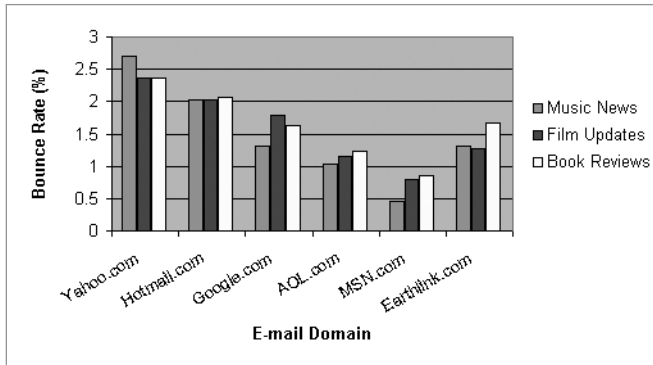

Figure 14-7. *Bounce rate across different e-mail domains*

Customer Analysis

In addition to the campaign response analysis discussed earlier, we can use CRM data in the data warehouse for analyzing customer activities and behavior. This analysis of customer activities and behavior is what we call *customer analysis*. There are two main categories of customer analysis: descriptive analysis and predictive analysis. In descriptive analysis, we try to understand or describe customer behavior and activities. In predictive analysis, we try to predict or forecast customer behavior and activities. The following are examples of descriptive analysis:

E-commerce industry: Purchase pattern by demographic, where we try to understand the purchase and browsing behavior of our customer group by demographic attributes, such as age, occupation, income, location, family, and interest.

Telecommunication industry: Usage analysis against products, such as analyzing a customer's cell phone usage to see whether the calling plan they are on now is the most appropriate one. We could then use the CRM campaign to encourage switching to the correct calling plan to increase customer service satisfaction.

Insurance industry: Risk pattern identification, where we calculate customer risk factors and insurance premiums according to customer activities. For example, for car insurance, the risk groups could be determined according to the base location, driving area, type of cars, vehicle usage, and traveling time.

Utility industry: Customer profitability analysis, where we find out which customers we are making money on and which ones we are losing money from and by how much.