- **Premium:** This is typically the written premium associated with each record. If the line of business has multiple coverages, this information is recorded by coverage (represented either as a separate record or via a coverage indicator field). For example, personal auto insurance databases track premium separately for bodily injury, property damage, comprehensive, collision, etc. Earned and in-force premium can be calculated from the information on the record.
- **Exposure:** This is typically the written exposure associated with each record. If the line of business has multiple coverages, this information is recorded by coverage.
- **Characteristics**: Characteristics include rating variables, underwriting variables, and any other available information regarding the risk represented by the record. Some characteristics describe the policy as a whole (e.g., the year the policy originated with the company) and as such are the same for every record associated with a particular policy and period of exposure. Other characteristics describe individual risks (e.g., make/model of automobile) and consequently vary between different records on the same policy.

As separate records are generated for midterm adjustments, the characteristics corresponding to each record are those that were in effect during the relevant period of exposure (e.g., if records are split to reflect a deductible change, the first record shows the initial deductible and the subsequent record(s) shows the new deductible).

Frequently, risk characteristic information is captured in multiple databases across the company and, as such, may be difficult to obtain and merge. For some rating characteristics, it is advantageous to capture a stable element from which the rating characteristic can be derived. For example, age of driver is a typical rating variable for personal automobile insurance; however, it is better to capture the date of birth of the driver on the data record because the driver's date of birth will not change from one policy period to the next but the driver's age will.

The following example homeowners policies can help clarify the construction of the policy database:

- Policy A is written on January 1, 2010, with an annual premium of $1,100. The home is located in Territory 1 and the insured has a $250 deductible. The policy remains unchanged for the full term of the policy.
- Policy B is written on April 1, 2010, with an annual premium of $600. The home is located in Territory 2 and the insured has a deductible of $250. The policy is canceled on December 31, 2010.
- Policy C is written on July 1, 2010, with an annual premium of $1,000. The home is located in Territory 3 and has a deductible of $500. On January 1, 2011, the insured decreases the deductible to $250. The full annual term premium after the deductible change is $1,200.

Policy A expired at its original expiration date and had no changes, thus the entire policy can be represented with one record.

Policy B was canceled before the policy expired. This is represented by two records. The first record for Policy B contains the information known at the inception of the policy (e.g., one exposure and $600 in written premium). The second record represents an adjustment for the cancelation such that when aggregated, the two records show a result net of cancellation. As the policy was canceled 75% of the way through the policy period, the second record should show -0.25 exposure and -$150 (=25% x -$600) of written premium.

Policy C expired at the original expiration date, but has a mid-term adjustment; this is represented by three records. The first record includes all the information at policy inception. The second record negates the portion of the original policy that is unearned at the time of the amendment (i.e., -0.50 exposure and -$500 premium and deductible equal to $500). The third record represents the information applicable to the portion of the policy written with the new deductible (i.e., +0.50 exposure and +$600 premium and deductible equal to $250).

Table 3.1 is an example policy database for the three policies described above.

3.1 Policy Database

Policy	Original Effective Date	Original Termination Date	Transaction Effective Date	Ded	Terr	Other Chars	Written Exposure	Written Premium
A	01/01/10	12/31/10	01/01/10	$250	1	…	1.00	$1,100
B	04/01/10	03/31/11	04/01/10	$250	2	…	1.00	$600
B	04/01/10	03/31/11	12/31/10	$250	2	…	-0.25	-$150
C	07/01/10	06/30/11	07/01/10	$500	3	…	1.00	$1,000
C	07/01/10	06/30/11	01/01/11	$500	3		-0.50	-$500
C	07/01/10	06/30/11	01/01/11	$250	3	…	0.50	$600

*For illustrative purposes this is ordered by policy rather than transaction effective date.

In a more sophisticated data mart, information for Policy B would be aggregated to one record that shows a "net" exposure of 0.75 and "net" written premium of $450. Similarly, information for Policy C would be aggregated to two records representing before and after the deductible change. The first record would reflect the period of time with the $500 deductible and would have a "net" exposure of 0.50 and "net" written premium of $500. The second record reflecting the period of time with the $250 deductible would be identical to the third record in the original example. The exposure is 0.50 and written premium is $600. This type of transaction aggregation is required for statistical ratemaking analysis such as generalized linear models (discussed in more detail in Chapter 10).

Claims Database

Most companies maintain a separate database to capture all available information about the claims on a specific policy. In a claims database, each record generally represents a transaction tied to a specific claim (e.g., a payment or a change in reserve). The fields contain dates or other explanatory information with respect to that claim. Similar to the policy database, claims involving multiple coverages or causes of loss may be represented as separate records or via indicator fields.

The following are the fields typically present for each record on the claims database:

- **Policy identifier**
- **Risk identifier(s):** If relevant, the claim database contains a way to identify the risk that had the claim. This will be necessary to match the claim to the corresponding record in the policy database.
- **Claim identifier:** The claim database contains a unique identifier for each specific claim. This same identifier is used if the claim has multiple claim transaction records.
- **Claimant identifier:** The claim database contains a unique identifier for each specific claimant on a particular claim.