Understanding Data Structure

In order to use reporting tools such as Actuate and DataExpress, you need to become familiar with your data and understand how it is organized. The following topics are covered on this page. Click on a link to go to a topic or begin at the top to read all the topics.

- Data Dictionary
- Data Elements
- Data Sets
- Databases
- Obtaining Data Dictionary Information
- Working with Database Information

Basic Data Structure

Data is organized on the HP3000 in a hierarchical structure that consists of databases, data sets, and data elements.

- A **database** is a collection of data organized for rapid processing, searching, and retrieval. A database consists of related data sets. An example of a database is SM (Student Management).
- A **data set** consists of related data elements (fields) that define the contents of a record, for example, all the data elements that identify a student's address. An example of a data set is STU-D (Student Detail) in the SM database.
- A **data element** is a group of related bytes that identifies a unit of information, such as street address, city name, state name, and postal code. A data element defines the attributes of a field. Examples of data elements are STU-STREET, STU-CITY, STU-ST, and STU-ZIP, which are a few of the data elements in the STU-D data set in the SM database.
- A **byte** is a one-character piece of information. It is the smallest unit of information in a database. A byte can be extracted using a character string computation.

Data Dictionary

The Data Dictionary describes the data elements, data sets, and databases that are used in the application system programs and processes. It does not contain the actual data; rather it describes the type, location, usage, and relationships of the Data Dictionary entities.

The Data Dictionary describes the following data entities: data elements, data sets, databases.

Data Elements

A data element is the smallest entity in a database that can be accessed without using a computation. It defines a single piece of information, for example, a student's street address (STU-STREET) or a student's ZIP code (STU-ZIP). Data elements typically define a field name on a screen or report. The Data Dictionary defines various attributes of data elements.

You might encounter some of these terms in DataExpress when you are working with data elements:

- The **data element short name** consists of 15 or less characters These are often very cryptic abbreviations, for example, CLVL-GPA. The data element short names are used to create DataExpress procedures.
- The **data element long name** is a more complete or descriptive name for a data element; for example, College Level Grade Point Average is the long name for CLVL-GPA.
- The **data element type** indicates how the data for a data element is stored in the computer. Some of the data element types you might encounter are:

    X   **Alphanumeric**. Data elements that are classified as type "X" are text-only, character strings. They are non-numeric and cannot be used in arithmetic expressions.

    Z   **Zoned decimal numeric**

    P   **Packed decimal numeric**

    I   **Integer numeric**

    J   **Integer numeric**

A plus sign (+) added to a numeric data element type indicates that the element cannot have negative values.

- The **size** indicates the maximum number of positions (spaces) there are for entering the data. For example, the size for the data element STU-ST (Student State) is two positions.
- The **length** indicates how the data is stored in the computer. The HP computer stores all data in increments of two; therefore, a data element that is five positions long requires six positions of storage space. For example, the size for the high school code (HI-SCHL) is three positions, but its storage length is four positions.
- **Dec** refers to the number of decimal positions for the data element.
- The **edit mask** provides information about how the data appears. Edit masks indicate slashes in dates, decimal marks in credit fields, dollar signs in money amounts, suppression of leading zeros, and so forth. Some common edit masks are:

    z   suppress leading zeros

    ^   print whatever is in the field

    !   put a decimal in this position, regardless of the defined number of decimals

    ^^/^^/^^ put in slashes to separate month, date, and year

Data Sets

A data set is a group of related data elements (fields). A data set typically defines the contents of a record. A data set combines several pieces of information; for example, the STU-D data set includes the data elements STU-STREET, STU-CITY, STU-ST, STU-ZIP, CLVL-GPA, HI-SCHL, as well as others. Every data set has a key data element that identifies the record. Some

data sets have more than one key data element. Key data elements are used as pointers or links to other data sets.

There are three types of data sets:

- An **automatic master data set** contains only one data element and that data element is a key to a detail data set in the same database. The key data element points or links to detail data sets where the remainder of the data resides. An automatic master record is created by the computer whenever a new record is added to a related detail data set. Use an automatic master data set for faster searches when primary selection criteria is based on a specific value of the key or a range value of the key. Automatic master data sets are identified by the suffix "A" in the data set name. For example, YRQ-A is an automatic master that locates all classes for a specific year/quarter.
- A **manual master data set** can contain other data elements in addition to the key element. A manual master data set can have only one key element and can contain only one record for each value of that key element. Manual master data sets link to detail data sets. Manual masters are identified by the suffix "M" in the data set name. For example, STU-YRQ-M is a manual master that contains quarterly biographic data for a student. There is only one record for a student for each year/quarter and those records are identified by the key data element STU-QTR.
- A **detail data set** can contain many data elements in addition to one or more data elements which act as key elements when linking from master data sets. Detail data sets contain descriptive information related to records in master data sets. A detail data set is identified by the suffix "D" in the data set name and is differentiated from a manual master in one of the following ways:

- There could be more than one record for a particular key data element. For example, the UNUSUAL-ACTN-D data set can contain more than one unusual action record for a student.
- There might be more than one key data element associated with the records in a detail data set. For example, records in the STU-D data set can be identified by SID (which is unique for each student) or by ABBR-LAST-NAME (which is not unique; for example, there could be many students with the last name "Smith").

Databases

A database is made up of related data sets (and the data elements associated with each data set). For example, the SM database consists of many data sets containing student and course-related data elements.

Obtaining Data Dictionary Information

To obtain hardcopy reports of the Data Dictionary, order the following jobs through the Job Scheduling System:

- **Data Elements by Element Abbreviation (DD0002J)**. This is a long report of approximately 1000 pages that contains all the data elements in all the FMS, PPMS, and SMS databases. You will want to coordinate running this job with your computer services department.

- **File Element Listing (DD0011J).** When you order this report, you select the database for the data sets (files) that you want.

Working with Database Information

You can create DataExpress procedures which:

- Work with single or multiple data sets
- Work with single or multiple databases
- Work with combination of files and data sets

The computer accesses the information in data sets either:

- Sequentially (in consecutive order, one item at a time).
- Randomly (in random order based on a key piece of information).

Data that is stored in the computer can be either:

- Global data. This is static information that changes only occasionally. A global item is widely accessible.
- Detail data. This is dynamic information that changes frequently and/or there is more than one occurrence of it.