

## Files, File Management and File Manipulation

---

### Import

Import almost any file type from almost any source, using IDEA's ImportAssistant to guide you. For more complex files, variable length records or multiple record types, IDEA provides an Advanced Record Definition Creator. IDEA will also seamlessly import XML files, specific accounting packages via downloadable import plug-ins and reports printed to files and PDFs (see Report Reader below).

If **IDEA Server** is installed, files can be directly imported to the Server or uploaded from IDEA.

### Report Reader

This powerful IDEA component offers tools for importing data from complex systems. Rather than dealing with raw data, you can create a report using system tools, save the report to a file (PRN or PDF), then use Report Reader to describe and capture the data in IDEA. You can view the entire report on screen, preview the IDEA database before importing and use simple “point and choose” tools to identify the data you need.

### Bad Data

When a file is imported and Date, Numeric, Time, or Virtual fields contain invalid data, a bad data database is generated. In the imported database, the invalid field data is marked in red as “Error”. If a change in a field type or the application of an equation to a virtual field results in invalid data, data is also marked as “Error”.

### Export

IDEA databases can be exported in all common formats for word processors and spreadsheets, plus HTML and XML. IDEA can also create PDF and RTF files for databases, reports and results.

### File and Client Management

IDEA uses a File Explorer window to provide easy, industry-standard file management. This “fly-out” window displays information about all IDEA files in the current folder, or the current Project if IDEA Server is installed, in either a hierarchical or sorted list. IDEA files stored locally are shown with blue icons; IDEA Server files are shown with green icons. The File Explorer toolbar gives easy access to file management functions, including the ability to flag a file.

### Visual Connector

The Visual Connector allows you to generate a single database from several other databases that share common or “key” fields. To create a visual connection between databases, you select a primary database, and then connect databases that contain matching records. The relationship that the visual connector creates between the databases is one to many, meaning that the primary database can have several matches in connected databases. All records in the connected databases that match records in the primary database as well as all other records in the primary database can be included in the output database.

## Join

The Join Database option is used to combine fields from two databases into a single database for testing or test for data which matches or does not match across files. Files can be joined or matched if they contain a common link (referred to as the “key”) e.g. part number if joining the pricing and inventory files.

## Append

The Append Databases option is used to append or concatenate two or more files into a single database for audit testing. For example, you may append 12 monthly payroll files to produce a database of all payroll transactions for the year. The database could then be summarized by Employee to produce year-to-date gross, net, tax, deductions, etc. Up to 32,768 files can be appended into a single database.

## Compare

The Compare option is used to identify differences in a numeric field within two files for a common key. Files could be compared at two points in time, e.g. the payroll at the beginning and end of the month to identify changes in salary for each employee. You can also compare a numeric field from different systems, e.g. the quantity of inventory on the inventory master file versus the quantity of inventory on the inventory count file.

## Data Functions

---

### Direct Extraction

Extractions, or exception testing, is the most frequently used function in IDEA, used to identify items which satisfy a specific characteristic, such as payments more than \$10,000 or transactions before a given date. The extraction criteria are entered using the Equation Editor and all records satisfying the specified criterion are output to a new database. You can perform a single extraction on a database, or up to 50 separate extractions with a single pass through the database.

### Indexed Extraction

This extraction option allows you to limit the scope of data for which IDEA searches in the database. An indexed extraction saves time when reviewing large databases. You can select and index for the search, rather than have IDEA search through the entire database.

### Key Value Extraction

This extraction option allows you to generate a series of secondary databases quickly from common values found in a main database. Key value extraction does not require you to create equations to perform the extraction. A key is an index created for information in a database. A key value is a piece of information indexed by the key.

### Top Records Extraction

This extraction option allows you to generate a database of the first specified records for each unique key, for example, the top five sales by salesperson.

### Equation Editor

The IDEA Equation Editor allows you to create simple or very complex extractions; ease of use enhancements include integrated on-screen help, “intelli-sense” for function and field names, pop-up Fields window and built-in interactive equation tester.

### @Functions

@Functions are used to perform more complex calculations and exception testing. IDEA provides almost 100 functions which can be used for date arithmetic, text manipulation and conversion and numerical, financial and statistical calculations. IDEA functions begin with the '@' symbol, very similar in style and operation to functions found in Microsoft Excel.

## Sort

The Sort option is used to create a new database physically sorted in the specified order. Sorting can significantly improve performance of certain functions.

## Search

Search is a simple way to find text within the fields of a database without using an equation to specify the criteria. By entering the text you are looking for, Search looks for matches within the specified fields of one or more databases stored in the same working folder (local databases) or project (if you have IDEA Server). Search also enables keyword searches, the use of wildcards, and proximity searches. IDEA presents search results in the fly-out Search Results window at the base of the Database window, and optionally, as an extraction database.

## Add fields (Field Manipulation)

You can insert and delete fields without doing an extraction or import – use the Field Manipulation task to view field definitions, add or delete fields, change field properties, for example, field name and type, or add tags for IDEA Smart Analyzer, an optional component. The tag names are also accessible by IDEAScript to allow users to create their own standardized tests.

## Action Fields

This field type allows you to set up relationships across multiple files; for example, create an action field on “Customer Number” in a customer database to link to related invoices. Action fields can also be used to run IDEAScripts or other applications.

## Group Records

You can use the Group Records task to organize matching records in indexed fields (keys) into expandable/collapsible groups.

## Chart

The Chart Data option can be used to graph data files or test results, in line, bar, stacking bar, curve, scatter, pie or area charts. The Charting Assistant will guide you through the steps for creating a chart. Chart options include titles, 3D effects, legends, colors and patterns, and grid styles. Charts can be printed, saved to as a Bitmap file or copied into any other Windows application via the clipboard. You can also choose to base your chart on “live” or snapshot data.

## Analysis Functions

---

### Field Statistics

Statistics can be generated for all numeric and date fields within a database. For each numeric field, values such as net, maximum, minimum and average values as well as numbers of debit, credit and zero value items are provided. For each date field, statistics provide information such as the earliest and latest dates and daily and monthly analyses of numbers of transactions. Statistics for time fields include latest and earliest time, most common hour, minute and second, records before 6 am or after 6 pm. Field statistics can be exported to Text, Excel, HTML and PDF.

### Stratification

Numeric Stratification, Character Stratification and Date Stratification are powerful tools used to total the number and value of records within specified bands. Examples of use include analyzing items by postal code or alphanumeric product code or fixed assets by date of acquisition.

### Summarization

The Quick Summarization function is used to accumulate the values of numeric fields for each unique key where there is a single field in the key. The Key Field Summarization function is used where there is a one or more fields in the key. Summarization results can be graphed and users can drill down on each unique key. Summarization includes statistics – average, minimum, maximum and variance – and an option to calculate percentage when creating a result file.

## Aging

The Aging function is used to age a file from a specified date for up to 6 user defined intervals. For example, the outstanding accounts could be aged at the year-end in order to determine provisions required against bad debts. The Aging function produces an Age Analysis report and two optional databases – Detailed Aging and Key Summary.

## Duplicates

IDEA can identify duplicate items within a database where up to 8 fields of information are the same e.g. duplicate account numbers, duplicate addresses or duplicate insurance or benefit claims. Duplicate Key Exclusion identifies duplicates but only where a specified field is different. This helps reduce the number of legitimate duplicates, e.g. standing orders.

## Gaps

You can search a file for gaps in numeric or date sequence, or alphanumeric sequences with a user defined mask. For date gaps, you can choose to ignore weekends or user specified holidays. Like many other IDEA functions, you can apply criteria before your search, e.g. look for gaps in check numbers, where the check amount is greater than \$1,000. You can also modify the increment – look for gaps in multiples of 10, for instance.

## Benford's Law

Applying Benford's Law to digital analysis can identify possible errors, potential fraud or other irregularities. The Law states that digits and digit sequences in a dataset follow a predictable pattern. The analysis counts digit occurrences of values in the database and compares the totals to the predicted result according to Benford's Law.

## Pivot Table

A Pivot Table allows users to create multi-dimensional, multi-variable analysis of large data files. When creating a pivot table in IDEA, the user drags and drops the required data fields onto the location on the pivot table where the user wants the information to appear. New enhancements include usability including export to Excel, ability to rearrange and sort columns and rows, top 'n' view of data, ability to display and use "page" data. Data can be shown as a "% of row", "% of column" and "% of total".

## Advanced Statistical Methods for IDEA

(Available August 31, 2007 as a free module that users "under maintenance" can download from our website.)

1. Correlation, also called Correlation Coefficient, is a measure of the strength and direction of a linear relationship between two random variables. Correlation, whose absolute value is less than or equal to 1, is computed using the Pearson product-moment correlation coefficient.
2. Trend Analysis is used to measure whether the data in a selected Numeric field is showing an upward or downward trend over time and then to fit the best fitting straight line to the data to forecast values into the future on the assumption that the trending pattern will continue.
3. Time Series is used to calculate the trend (upward or downward) and the seasonal pattern of data in one Numeric field over time and to predict values into the future on the assumption that the trend and seasonal pattern will continue. The statistical approach taken when designing the Time Series task was to use a method called decomposition and trend.

## Sampling

IDEA offers five sampling methods together with the ability to calculate sample sizes based on parameters entered and evaluate the results of sampling tests.

## Systematic

This method allows you to extract a number of records from a database at equal intervals, by entering the required number of records or specifying the selection interval.

## Random

With this method, you enter the sample size as well as the range of records from which the sample is to be extracted. Then, using a random number seed, IDEA generates a list of random numbers and selects the appropriate records associated with these numbers.

## Stratified Random

This method allows you to extract a random sample with a specified number of records from each of a series of bands. This task requires the database to first be stratified into a series of numeric, character or date bands. IDEA then presents you with a table displaying the number of records within each band. You enter the number of sample records you want IDEA to extract at random from each band.

## Monetary Unit

With this method, you choose to test monetary units on an interval (or random item within each interval) basis; the items which contain those monetary units are identified. Any items greater than the sampling interval should be extracted from the population and subject to separate testing, leaving just those items below a specified high value amount to be statistically evaluated.

## Classical Variables

Classical Variables Sampling techniques are appropriate when several errors are expected in the sample database. It can also be used for accounting estimates such as obsolete inventory.

## Productivity

---

### Database View

IDEA provides multiple display options for viewing and formatting data as well as zoom, powerful right mouse functions and “find all records like...”. From within the “fly-out” Properties window, you can view Control Totals, Criteria and Indices and easily switch between views of Database (Data, History and Field Statistics), Results or Comments. A check mark next to any item in the Properties window indicates that it is active in the current view.

### History

The History View of the Database Window maintains an audit trail or log of all operations carried out on a database, presented in an easy-to-use expandable list. Each test or function performed automatically generates IDEAScript code, which can be copied to the IDEAScript editor. IDEAScript is a Visual Basic compatible programming language – see below for more information.

### IDEAScript

IDEAScript is an object oriented programming language, compatible with Microsoft's Visual Basic for Applications™. IDEAScripts, which are often called macros, can be created either by recording a series of steps, by copying from other scripts, by copying from the History, or by typing into a Macro Window or a combination of these. Scripts can be attached to the Tools menu or to a Toolbar Macro Button or run simply from the Tools – Macro menu. IDEAScript has a number of supplied tools: the Language Browser, the Dialog Editor and the Debug Tools to help create and complete a script.

### Help and Bubble Help

IDEA's Help system is easy to use and comprehensive. It is organized into three sections, Getting Started – overview information on the product and how it can be used, Navigating IDEA – information on individual IDEA menu items, windows and toolbars, and Working with IDEA Tools – detailed information on using Equation Editor, Report Reader and IDEAScript. Bubble Help instant pop up help on toolbars and dialogs providing immediate information.