

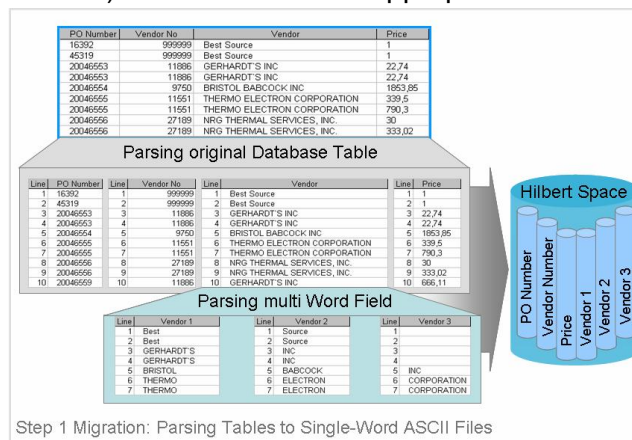
# HILBERT<sup>®</sup> MIGRATION & QUANTIFICATION OVERVIEW



Migration, Quantification and Sort are the preparatory steps the Hilbert Machine performs to place data from structured or unstructured sources into the Hilbert Space. These steps are fundamental and essential to the speed and flexibility advantages the Hilbert Technology provides.

## Migrating Tables into Single-Word Files

The Migration phase converts each column of a database table into single-word ASCII text (*American Standard Code for Information Interchange; defines a unique number per character*) or numeric files as appropriate. This phase also parses text fields which contain more than one word into multiple single-word files. In migrating an unstructured text file, each word is handled individually and stored, along with its position within the text, in single-word ASCII files analogous to a database table information. Having such small files promotes rapid and efficient accessing and processing because a larger quantity of closely related data can co-exist in the computer memory at the same time. This set of ASCII files is the starting point for the Hilbert Engine, and is used as reference data. All single-word ASCII files which belong to one table will be grouped and the group they belong to will be stored in the header information of the existing file.



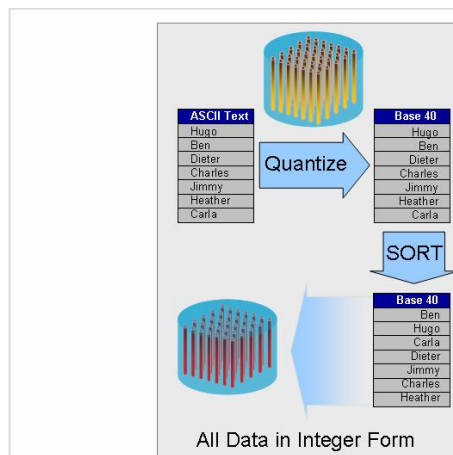
Step 1 Migration: Parsing Tables to Single-Word ASCII Files

Step 1 Migration: Parsing Tables to Single-Word ASCII Files

Step 1 Migration: Parsing Tables to Single-Word ASCII Files

## Quantify all Data to Integer Values

The quantification process converts or quantifies numbers and text data to unique integers (Polynomial Integer Quantities – PIQs), sorts them for ultra high-speed search capabilities and loads them into Hilbert Space. The quantification process transforms the ASCII text into numbers by utilizing a Base-40 algorithm (representing 0 ... 9 and A .. Z plus four extra characters), Z plus four extra characters), which allows a unique and economical representation of all data whether originally in numeric or text format.



Quantizing ASCII Information to Numbers in Hilbert Space

## Why Hilbert?

We are the only solution provider for ultra high-speed access, manipulation, storage and analysis of large volumes of structured and unstructured data.

Speed, Flexibility and Scalability have been the design criteria of the Hilbert solution.

The Hilbert solutions are offered as embedded Hilbert Engine technology, standard products or industry specific solutions.

Bjorn Gruenwald, CEO and Founder of Hilbert Technology Inc.

This is the default base; quantification can be configured to use any base, as may be needed to satisfy the requirements of the data and the character set used (for Arabic or Asiatic languages Hilbert may take more than 40 characters to represent all their text and numeric information). A simple yet powerful mathematical concept allows the Hilbert Architecture to support any and all language data. This numeric representation allows complex but highly efficient mathematical operations to handle text information with as much speed as numerical data.

### Quantification in Base-40

While any base can be used to represent data as integers, Base-40 was chosen to be the default base for the Hilbert Engine's quantification method. The Hilbert Base-40 symbols include the usual decimal digits, the 26 uppercase letters of the alphabet, and four punctuation symbols. This symbol set is rich enough to represent data in a 'language' understandable to humans, and yet lean enough to translate efficiently into a 'language' understandable to computers. Internally, computers follow binary or 'machine language' instructions to do all storage and calculations of data. Computers only understand the Base-2 symbols 0 and 1, because in a computer's world, either the electricity is on, or it is off.

The Hilbert Engine uses, for example, a BASE-forty System:

00 01 02 03 04 05 06 07 08 09 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39  
 0 1 2 3 4 5 6 7 8 9 A B C D E F G H I J K L M N O P Q R S T U V W X Y Z # ! @ ^

M A R Y

$$\text{MARY}_{40} = 22 \times 40^3 + 10 \times 40^2 + 27 \times 40^1 + 34 \times 40^0 = 1,425,114_{10}$$

$$= 1,425,114_{10} = 10101101 11110110 11010_2 \leftarrow \text{apx. 50\% reduction} \rightarrow$$

$$\text{MARY} = 4D 41 52 59 00_{16} = 01001101 01000001 01010010 01011001 00000000_2$$

Line	Vendor	ASCII	Line	Vendor	Base 10
1	Best		1	Best	727549
2	Best		2	Best	727549
3	GERHARDT'S		3	GERHARDT'S	107263720149988
4	GERHARDT'S		4	GERHARDT'S	107263720149988
5	BRISTOL		5	BRISTOL	47868719381
6	THERMO		6	THERMO	3014060104
7	THERMO		7	THERMO	3014060104

Step 2 Quantification: Text to Integers

## Example

As an example of 'Base Notation,' which is used when talking about base systems,,  $10001_2 = 17_{10}$ , means the Base-2 (or binary) number 10001 is equivalent to the Base-10 (or decimal) number 17.

The diagram above contrasts the 'machine language' generated from a Base-40 representation of the ASCII text 'MARY,' with the 'machine language' generated from a Base-16 representation of the same ASCII string.

The ASCII text 'MARY' converts to Base-40 as,  $MARY_{40}$ . This number is equivalent to a Base-10 (decimal) value of  $1425114_{10}$  and a Base-2 (binary) number  $10101101\ 11110110\ 11010_2$  as shown.

The same text 'MARY,' converts to Base-16 (hexadecimal), as  $4D\ 41\ 52\ 59\ 00_{16}$  and finally into a binary number that is much longer than the binary number generated from the Base-40 representation— one of the optimizations in Hilbert.

## Sorting for Ultra High-Speed Search

After the ASCII-text single-word files are quantified, the single-number files are sorted. This allows the Hilbert Engine to do binary searching, a highly efficient method of finding a specific number in a large set of sorted numbers. Binary searching permits ultra high-speed searches over large data volumes, and enhances access and analysis performance.

## About Hilbert

Pennsylvania based Hilbert Technology Inc. is an international provider of business intelligence (BI) solutions for large and medium size enterprises worldwide. The offering is based on the revolutionary, patented Hilbert Engine technology for the ultra high-speed access, manipulation, storage and analysis of large volumes of structured and unstructured data. The Hilbert solutions are offered as embedded Hilbert Engine technology, standard products or industry specific solutions. Organizations in public services, law, government, finance, communications, whole- and retail sales, transportation & tourism and chemical & pharmaceutical can employ Hilbert solutions to gain unparalleled speed in access and analysis over large data volumes.

## Notes

This document is provided for information purposes only, and the contents hereof are subject to change without notice. This document is not warranted to be error-free, nor is it subject to any other warranties or conditions, whether expressed orally or implied in law, including implied warranties and conditions of merchantability or fitness for a particular purpose. We specifically disclaim any liability with respect to this document, and no contractual obligations are formed either directly or indirectly by this document. This document may not be reproduced or transmitted in any form or by any means, electronic or mechanical, for any purpose, without our prior written permission.

Hilbert Technology Inc.  
P.O. Box 330  
Newtown, PA 18940  
USA  
Web: [www.hilbertcompany.com](http://www.hilbertcompany.com)  
E-Mail: [info@hilbertcompany.com](mailto:info@hilbertcompany.com)  
Tel: +1 (212) 252 1600  
Fax: +1 (212) 252 1615