

# New Database Design Cuts the Middleware

A data-management startup says its information engine offers a better way of managing RFID data collection than the use of data-filtering middleware.

By Jonathan Collins

June 22, 2005—When it comes to storing and analyzing the vast amounts of data RFID creates when tagging shipments and items, companies could learn a lot from video games, according to Berlin-based data-management venture [Pile Systems](#).

Pile has developed a new database design—dubbed an information engine—that it says will drastically cut the cost of data storage by eliminating data duplication. The information engine will also store data in a way that enables it to be restructured to suit any query a user might run.

The information engine operates similar to how computer games generate screen images from a multitude of viewpoints in real time. "A computer game does not store single-image frames. Instead, it stores relationships in [a] form of code that, in real time, generates the images that you see on the screen. Otherwise, the number of possible frames in an interactive computer game would be so large that storing each possible frame would require hard disks of cosmic dimensions," says Peter Krieg, Pile Systems' president and CEO.

Instead of storing it in a database, the Pile information engine virtualizes data, saving a reference or link to a single instance of that data rather than each instance of the data itself. Because the reference takes up less storage memory than the data itself, the use of references cuts the capacity required to store information. "It's a web of relations," says Krieg.

As duplicate data is entered into the engine and stored, the engine saves the data reference. The data can then be rebuilt and outputted by the engine when required. The larger the amount of duplicated data entered into the engine, the greater the storage savings.

Jim Crawford, an analyst at [Retail Forward](#) in Columbus, Ohio, calculates that if [Wal-Mart](#) stored every RFID read of every tagged item on every shelf, it would generate nearly 8 terabytes of data per day. "There are some very deep problems for RFID where it comes into contact with the existing database systems," says Krieg. Given the large amount of data RFID systems are expected to create, RFID users are a key target for Pile's new software, according to the company.

By enabling more data to be stored on far less hardware, Pile believes its software presents a better method of managing RFID data collection than data-filtering middleware that allows only relevant data to be stored in corporate databases. Pile maintains that one key problem with the filtering approach is that data may be discarded that might later be determined to be useful. By storing data in a virtualized form, the company reports, its information engine enables data to be queried in a range of ways—unlike traditional relational and object relational databases, which can require huge IT resources to enable new key fields to be added or the data restructured.

"RFID introduces individual-transaction tracking," says Krieg. "Banks and telephone companies have experience with that, but Wal-Mart doesn't yet because it just hasn't been feasible [to use query the data easily]. Now it is." Wal-Mart can't use the same type of system banks and telephone companies have been using, because traditional transaction systems are dependent on being set up for the kind of data queries they are expected to solve. What's more, new and ad hoc queries can be extremely costly.

Pile believes, however, the flexibility of its information engine customers could allow complex ad hoc queries that would help drive the usefulness of RFID networks. For example, says Krieg, a retail customer equipped with a computing device could access the company database to see the products he usually buys, and while he is there, the retailer could also query the database to make purchasing suggestions and offer promotions. "Such interactive applications can be supported by RFID, but at the same time can drive the complexity of the database systems through the roof of computability and affordability," he explains.

Krieg claims his firm tested its engine against a traditional suffix-tree database used for gene analysis. Suffix trees are good for string-matching applications, such as those that arise when working with DNA sequences. The results reportedly showed that Pile's engine enabled the storage of 50 times the data of the existing system within 2GB of RAM and was several times faster at carrying out pattern-recognition queries.

Pile maintains that by reducing the computing power and resources required to manage and analyze vast amounts of data, its information engine would help drive the widespread deployment of RFID and other systems once deemed impossible to implement because of the vast computing requirements demanded by huge volumes of data.

Pile aims to license its information engine to database and application developers. The company says it has a prototype of its information engine ready but no applications—for RFID or for any other purpose. A development kit is available from the company's Web site, and Pile plans to make an open-source version of its development kit available this summer.

Copyright ©2005 RFID Journal, Inc. All Rights Reserved