WHITEPAPER

INGRES

INGRES VECTORWISE FOR ISVs

INGRES vectorwise

TABLE OF CONTENTS:

EXECUTIVE SUMMARY

Access to better analytics improves your customers' business. Your customers will pay for a solution that creates business value.

Existing analytic databases often require specialized skills, are costly to implement, or are marketed by startup companies that do not provide the enterprise-level support that you want for your customers. Ingres VectorWise is a new analytic database that is different:

- Customers report 10x-70x performance gains over other databases.
- Deliver solutions quickly with Ingres VectorWise. Fast performance requires no special tuning or schema design. Ingres VectorWise exploits the 95% of modern chip capability that other business software does not take advantage of.
- Ingres VectorWise is supported by Ingres, a global company supporting over 10,000 customers running mission-critical database applications.

This paper introduces the Ingres VectorWise technology and provides you concrete steps to enable your customers to take advantage of the power of Ingres VectorWise.

INTRODUCTION

"In an ever-changing climate of risk and fluctuating market conditions, having the right information to make the right decisions is the way to outperform the competition…" Study after study confirms the truth of this statement made by Accenture in their 2007 CIO Study: better information – enabled through better analytics – drive better business decisions.

As an Independent Software Vendor (ISV) it is important to:

- Provide a product that makes your customers successful.
- Deliver a solution that gives you strong competitive advantage over other players in the market.
- Use components that deliver optimum cost/benefit for your business.
- Secure the investment in your technology to ensure your solution benefits as underlying technology components such as computer chips improve.
- Support standards so that you are flexible to adopt and support better technologies as they become available.

Ingres VectorWise is an Ingres database that has been extended with a high performance analytic engine and column-based storage. Ingres VectorWise is the only relational database that unlocks powerful performance features in modern chips such as SIMD[1] processing, out-of-order execution, large chip caches and hardware-accelerated string-based operations.

Complex code structures in other relational databases prevent these products from transparently taking advantage of these performance features, and a complete rewrite of the code would be required to fully take advantage of these features.

[1] Single Instruction, Multiple Data, see http://en.wikipedia.org/wiki/SIMD

Ingres VectorWise customers have reported tremendous performance gains of 10x-70x compared to other relational database technology on the same hardware.

INGRES VECTORWISE TECHNOLOGY OVERVIEW

Ingres VectorWise is the first relational database to fully exploit the power of modern CPUs. This section provides an overview of the Ingres VectorWise breakthrough technology.

In-cache Vector-based Processing

Ingres VectorWise processes arrays – vectors – of data at the time. Modern CPUs support instruction sets to efficiently work with so-called SIMD instructions making data processing much more efficient. Traditional database software has become increasingly complex over the years and a complete rewrite would be required to take advantage of SIMD instructions.

Most database software that is commercially available today was initially written in the 1970s and 1980s when large servers had few megabytes of RAM. Access to RAM was fast, access to disk was slower, and access to tape was really slow. Today the amount of CPU cache alone is often greater than the amount of RAM in a server back in the 1970s or 1980s. Today access to the CPU cache is fast, access to RAM is slower, and access to disk is really (relatively) slow. Traditional database software has not changed to adopt these changes. Ingres VectorWise has. Ingres VectorWise was written to use the chip's cache as processing memory. RAM is used as a high performance disk buffer.

By taking advantage of these and other performance features in modern CPUs Ingres VectorWise processes data much faster than existing relational database technology. The following sub sections explain how Ingres VectorWise uses technologies that have been proven in analytic databases to enable relatively low-cost configurations and deliver an extreme price/performance ratio.

Updateable Column Store

Most analytic database queries only access a subset of all columns across the tables accessed in the query. Column-based storage, as opposed to row-based storage, is the most efficient storage model for these types of queries. The use of column-based storage results in less I/O when retrieving data from disk and is the first step in the Ingres VectorWise RDBMS to address the limited bandwidth issue.

Many analytic queries perform scans across many rows. At the storage level this type of operation translates into a sequential read from disk for a column-based store. Sequential reads are far more efficient and provide much better throughput on spinning disks than random reads. Many RDBMSs will translate table scans into less efficient random reads, due to various reasons including the row-based storage model, small block sizes and/or the implementation of parallel processing in conjunction with the layout of the data on disk.

Traditionally the Achilles heel of a column-based store is small incremental changes to the data as opposed to large data changes applied in bulk. Ingres VectorWise addresses this

challenge using a patent-pending approach called Positional Delta Trees (PDTs). The PDTs are in-memory structures that store changes to the data that is optimally stored on disk. Queries efficiently merge the PDT with the data from disk to produce a consistent and correct image of the data. PDTs are periodically merged and stored persistently to the table. Note that although PDTs reside in memory and will get lost if the database would crash all transactions will be recoverable from the transaction log.

Compression

Column-based databases achieve better compression ratios than row-based databases. Ingres VectorWise implements aggressive compression schemes to dramatically lower the amount of I/O while limiting storage space requirements. Ingres VectorWise automatically chooses the most optimal compression algorithm on a per-column basis. Decompression comes at almost no cost because it is directly integrated in the vector-based processing. It is far more efficient than previous speed-optimized compression libraries such as LZOP.

Since Ingres VectorWise effectively uses a portion of the server's memory as a buffer for disk I/Os it becomes feasible to run with part if not all of the database's data in-memory. A server with a large amount of memory that is running a relatively small Ingres VectorWise database can be configured to run entirely in-memory (using the disk as the persistent store).

Indexing

Ingres VectorWise supports two types of indexes. First of all Ingres VectorWise automatically maintains basic storage indexes on every column containing minimum and maximum values per block. These indexes are extremely small (using the default settings the space utilization for the storage index is less than 0.1% of the column size). The storage indexes enable filtering of blocks of data that will not contain values of interest to a query leading to lower I/O requirements and more efficient data processing since less data will have to be filtered out by the CPUs.

The second type of index Ingres VectorWise supports is the clustered index. In some cases data in a table is typically accessed through a particular column (or set of columns) in which case it would make sense to order the data on that column (or set of columns). Ingres VectorWise supports this approach through the CREATE INDEX command.

Ingres infrastructure

Ingres provides utilities to manage the database, as well as various drivers to connect to the database. Being built into the Ingres database Ingres VectorWise takes advantage of this infrastructure. You can use the tools and utilities to manage Ingres VectorWise from a familiar desktop environment. The available drivers enable you to quickly get started using your tools and products against Ingres VectorWise with minimal risk.

CERTIFY INGRES VECTORWISE

Ingres VectorWise supports ANSI SQL and uses the Ingres drivers to connect to the database. Business Intelligence tools or products that generate standard ANSI SQL and support standard ODBC or JDBC connectivity should run on Ingres VectorWise with little to no modification. Applications that already support the Ingres database should have even fewer challenges to run on Ingres VectorWise.

Benefits

Your customers want to take advantage of the benefits Ingres VectorWise provides:

- Superior performance by unlocking performance features in modern CPUs.
- Ease of use with no requirement for expert tuning or special schema design.
- Enterprise-class support from a vendor with a proven track record to support mission-critical database applications.

Business Intelligence tools or products that perform analytic queries against a relational database will respond faster running against Ingres VectorWise. End users will be able to work more interactively and increase their productivity. What does it mean to an end user if queries return 10 to 70x faster?

Your customers want to use a relational database that performs fast analytic queries. Tools or products that perform data loads into Ingres VectorWise enable customers to take advantage of data analytics at the speed of thought.

Requirements and Prerequisites

The sub sections below consider three scenarios: (1) it is impossible to make changes to the product at this point in time, (2) some, but not very many changes are possible, or (3) the product can be modified to take full advantage of Ingres VectorWise. Each scenario describes requirements and/or prerequisites.

Obviously applications that satisfy the prerequisites are not guaranteed to support Ingres VectorWise. Any solution must be thoroughly tested against Ingres VectorWise.

No changes to the product

Some products may not be able to change in the short term. For example, the release cycle for the product prevents changes to be implemented soon. A product that cannot change to adopt Ingres VectorWise must at least satisfy the following prerequisites:

- The product must generate standard ANSI SQL.
- Standard ODBC, JDBC or .NET connectivity must be supported out-of-the-box and connections using these can be defined.
- Temporary tables are not required for the product to function[2].

[2] Ingres VectorWise will support temporary tables by the beginning of calendar year 2011. The application may create regular tables that act as temporary tables as long as the application takes care of removing the tables at the end of their use.

- If data is loaded into Ingres VectorWise then either a copy from a data file or create/insert into table as select from within Ingres VectorWise must be used[3].
- Secondary indexes are not required for the product to function correctly.
- The Ingres VectorWise database can run on x86-64 Linux[4,5].

Some changes to the product

Some products can introduce some, but not many or large changes. For example, the product uses a model with plug-ins that can be added dynamically, or a product patch/update is due in the short term that can introduce minor changes to support Ingres VectorWise. A product that can make some changes to adopt Ingres VectorWise must satisfy the following prerequisites:

- The product must generate standard ANSI SQL.
- Standard ODBC, JDBC or .NET connectivity must be supported.
- If temporary tables are required for the product to function correctly then it must be possible to use regular tables instead which have to be cleaned up explicitly by the product.
- If data is loaded into Ingres VectorWise then either a copy from a data file or create/insert into table as select from within Ingres VectorWise must be used.
- Secondary indexes are not required for the product to function correctly.
- The Ingres VectorWise database can run on x86-64 Linux.

Major changes to the product

Some products can introduce major changes to support Ingres VectorWise because a major update is coming up and development resources are available to implement changes.

Following are requirements that must be implemented to support Ingres VectorWise.

- Generate standard ANSI SQL to run against Ingres VectorWise.
- Use ODBC, JDBC or .NET to connect to Ingres VectorWise.
- Use only copy from a data file or create/insert into table select from within Ingres VectorWise to populate any tables in Ingres VectorWise.
- Eliminate any secondary indexes on tables that the product may require for other relational databases.

In addition to the requirements there is one major prerequisite:

- Ingres VectorWise database can run on x86-64 Linux.

---

[3] Ingres VectorWise will support create table as select from Ingres (classic) tables by the beginning of calendar year 2011. That release will also enable batch apply through the JDBC driver so that inserts/updates/deletes can more efficiently be applied directly.

[4] Other platforms will be supported in the future. Microsoft Windows will be the next platform to be supported, planned to be available no later than the beginning of calendar year 2011.

[5] The X86-64 architecture is provided by Intel and AMD. Most hardware vendors, including IBM, HP, Dell and Oracle Sun provide servers are using this chip architecture. Although Ingres VectorWise will take advantage of very recent performance features in CPUs, it will run fine on slightly older hardware as well (the benefit you achieve over other relational database technology may be slightly less impressive on older hardware).

EMBED INGRES VECTORWISE

There is a large market for analytic applications that are either sold directly to customers or provided as a hosted Software as a Service (SaaS) offering to customers. Analytic applications that use a relational database to perform analytics should consider embedding Ingres VectorWise.

Benefits

Ingres VectorWise provides you a number of benefits.

- Ingres VectorWise unlocks the 95% of modern chip technology that other databases don't utilize. The result is far superior analytic query performance compared to other relational databases running on the same hardware.
- Fast performance is delivered out-of-the-box and does not require expert tuning or special schema design. You can now develop your solution much quicker and at lower cost.
- Ingres is a global company with a proven track record to support mission-critical database applications. Ingres provides 24x7 follow-the-sun support in multiple languages.
- As an open source database company that pioneered the New Economics of IT, Ingres provides attractive subscription-based pricing. Ingres VectorWise gives you a very attractive price/performance ratio that you can use to offer your customers an extremely powerful solution at an attractive price while still earning good profits.

Your customers expect you to deliver a good product. Make sure to test your solution against Ingres VectorWise.

Requirements and Prerequisites

The very minimum change an application must undergo to support Ingres VectorWise is to swap out any current database solution with Ingres VectorWise. This requires changes to the installation and packaging, and possibly the way the application is populated for initial use.[6]

The following sub sections describe 3 different scenarios: (1) applications that cannot introduce any changes other than the very minimum to support Ingres VectorWise, (2) applications that can make some changes to support Ingres VectorWise, and (3) applications that can be re-architected to take maximum advantage of Ingres VectorWise. Every scenario states the requirements and/or prerequisites. Irrespective of whether the application satisfies the prerequisites or implements the requirements, make sure to thoroughly test Ingres VectorWise before selling an application that uses Ingres VectorWise.

Future versions of Ingres VectorWise will limit the number of prerequisites to adopt Ingres VectorWise. If Ingres VectorWise is not a good fit today then it may well be in the future. Please provide your feedback to your Ingres representative or use the online forum at http://community.ingres.com/forum/vectorwise/.

[6] An application that migrates from Ingres (classic) to Ingres VectorWise may not require many changes.

No changes to the application

Some applications can undergo no changes other than the database change to Ingres VectorWise. The reason may be the length of the release cycle and/or migrations from older versions of the application to the new version. Following is a list of prerequisites that the application must fulfill:

- The current database is Ingres or supports clean ANSI SQL that is compatible with Ingres.
- The application uses ODBC, JDBC or .NET to connect to the database.
- The application only makes use of data types supported by Ingres VectorWise. For a list of the supported data types please refer to the Ingres VectorWise User Guide.
- Ingres VectorWise database can run on x86-64 Linux, either directly, inside a virtual machine, or in the Cloud.
- The application does not rely on secondary indexes.
- Any tables that require primary indexes are only loaded in bulk.
- Every statement being a separate transaction does not introduce a problem.
- 24x7 availability is not a requirement. For example, the database can be shutdown or set into read-only mode for a backup.
- Disaster recovery for the application, if provided, is implemented using a dual load strategy.
- Any updates to the data in the database take place in batch or in bulk.
- Application changes that require changes to the tables are rare or don't occur[7].
- The application does not rely on the use of temporary tables. If tables have to be created temporarily, then the application explicitly creates and drops the tables.

Some changes to the application

Some applications can be modified to adopt Ingres VectorWise. However, major changes are impossible (at this point in time) due to either a release cycle and/or migration concerns in conjunction with the risk of introducing a new technology. Following is a list of pre-requisites for these types of applications.

- Ingres VectorWise database can run on x86-64 Linux, either directly, inside a virtual machine, or in the Cloud.
- The application uses ODBC, JDBC or .NET to connect to the database.
- The application only makes use of data types supported by Ingres VectorWise. For a list of the supported data types please refer to the Ingres VectorWise User Guide.
- Every statement being a separate transaction does not introduce a problem.
- Any secondary indexes the application uses can be eliminated if database query performance can deliver results fast enough without the use of these indexes.
- Tables that require indexes can be initially loaded in bulk and for subsequent (batch) updates the application can be changed to use the CALL VECTORWISE statement.
- Application changes that require table changes can be implemented using a create/drop table approach. This implies that downtime is acceptable to apply such an upgrade/patch.

[7] Ingres VectorWise 1.0 only supports very limited alter table statements. Restrictions will be lifted in future releases. To add/drop columns will be supported by the beginning of calendar year 2011.

In addition to the prerequisites there are requirements that the application must fulfill once the changes have been introduced.

- The application must generate ANSI SQL.
- Temporary tables, or tables that act as such, have to be created as regular tables and be dropped explicitly at the end of their use.
- Any disaster recovery strategy for the application is implemented using a dual load approach.

Major changes to the application

Some applications can be re-architected to take full advantage of Ingres VectorWise. Examples of such applications include applications that have yet to go to market, as well as existing applications that have a relatively long release cycle or SaaS applications that can be fully tested using all available data causing little risk to the operations when the switchover takes place.

The re-architected application must satisfy the following requirements:

- The application generates ANSI SQL to work with Ingres VectorWise.
- ODBC, JDBC or .NET are used to interact with the database.
- Only use data types supported by Ingres VectorWise. For a list of the supported data types please refer to the Ingres VectorWise User Guide.
- The application makes use of bulk or batch loads. Application developers may choose to rely on the automatic triggering of the CALL VECTORWISE statement, or explicitly code it into the application. Indexed tables must be incrementally loaded using the CALL VECTORWISE statement.
- Tables that act as temporary tables are dropped explicitly by the application.
- Disaster recovery is implemented using a dual-load strategy.

There is only one major prerequisite:

- The Ingres VectorWise database can run on x86-64 Linux, either directly, inside a virtual machine, or in the Cloud.

Finally, a re-architected application may consider:

- Use (heavily) denormalized tables. Depending on the workload the application may take maximum advantage of column-based storage and storage indexes making queries against a large denormalized table extremely efficient. Of course, to use any other schema definition is fine too.
- Create separate staging tables for data loads that are used to incrementally refresh the application tables. Use the CALL VECTORWISE command to optimize data storage for the application tables, while keeping the size of the staging tables small by emptying the tables once the application tables have been updated.
- Analyze the access pattern to the application tables to identify whether it makes sense to index a table. If there is no predominant access path then it probably does not make much sense to index the table in Ingres VectorWise.

INGRES PARTNERSHIP PLUS PROGRAM

Ingres Corporation invites you to take part in our Partnership Plus Program. The program lists various activities that you can choose to participate in. As you participate in more activities you move up through the levels of the Ingres Partnership program from Silver to Gold to Platinum and earn bigger discounts on our software and services.

Ingres is committed to its partners and wants to help you be successful. As soon as you join the Partnership Plus program you will be assigned a Business Development Manager who will help you get the most out of the program. If you run into technical challenges then Technical managers are available to assist.

Ingres offers you a free half-day quick start to help you get started with an Ingres VectorWise evaluation. The hands-on quick start provides you:

- An introduction into the Ingres VectorWise architecture.
- Product installation and configuration in your environment.
- Initial table creation and how to load data into Ingres VectorWise.
- An overview of basic database operations such as startup, shutdown and running queries.
- How to connect your application to Ingres VectorWise.

For more information about the Ingres Partnership Plus program please visit http://www.ingres.com/partners.

GET STARTED USING INGRES VECTORWISE TODAY

The best way to experience the value of Ingres VectorWise for your customers is to try the technology. You don't require a large investment and there is little risk involved in getting started. An evaluation version of Ingres VectorWise can be downloaded (http://www.ingres.com/vectorwise) at no cost. You can get started quickly on your own commodity server, using a virtualized environment or by using resources in the Cloud[8] at low cost to you.

Here are some ideas to get started:

- Use your tools or products to connect to Ingres VectorWise, load some data, and experience the improvements to query performance.
- Use another product – perhaps open source – to access Ingres VectorWise and identify the benefits you can achieve. From there, decide how to move forward with Ingres VectorWise.
- Offload some of the data out of your application into Ingres VectorWise and run queries you captured from your application. Evaluate the performance of Ingres VectorWise and compare to your current solution.
- Maybe you can identify a module of your application that could be moved into Ingres VectorWise for reporting purposes to entice your customers with better reporting performance.

[8] For more information, see http://community.ingres.com/wiki/Ingres_VectorWise_Amazon_EC2_Images

- Encourage some of your most engaged customers, or ones who have reported performance challenges, to work with you to test Ingres VectorWise. Help them become successful while learning this exciting new technology.
- Apply a combination of these ideas.

WHY INGRES

Your most important goal as an ISV is to ensure your customers are happy. In addition you want to run a profitable business. Here are some key points to remember about a partnership with Ingres.

- Ingres has been around for over 25 years and supports more than 10,000 customers around the world running mission-critical database applications. Ingres provides a professional 24x7 follow-the-sun support model. You and your customers can rely on solid product support if and when you need it, wherever you are located.
- Ingres is committed to make its partners successful.
- Ingres pioneered the New Economics of IT, enabling its customers to cut costs using an open source software model while providing high quality professional support based on subscriptions to the software. Take advantage of this model to make your business more profitable.
- The IT landscape changes constantly. Vendors you treated as partners may have turned into competitors due to mergers or acquisitions. To support a competitor's product involves risks. Ingres's focus has been and is the database business. Ingres VectorWise does not change this.
- Ingres Corporation is very flexible to work with you to implement requirements you may have. Ingres VectorWise is a new product that will be updated frequently in the next few years. Take advantage of this exciting opportunity to get your requirements included in Ingres VectorWise.

Learn more about Ingres at http://www.ingres.com.

Why Ingres VectorWise

Ingres VectorWise is a new relational database that performs analytic queries much faster than any other relational database. As an ISV you want your customers to be happy. Your customers can improve their business with better analytics. Adopt Ingres VectorWise and achieve:

- Faster performance: customers have reported 10x to 70x performance gains over existing relational database solutions running on the same hardware.
- Better use of existing hardware resources: Ingres VectorWise unlocks the 95% of chip functionality other database software does not utilize.
- Quick value: Ingres VectorWise is easy to use, requires no expert tuning and no special schema design.

- Rapid return on investment: Ingres provides an attractive subscription-based model that enables you to provide your customers a powerful solution at an attractive price.
- Security: Ingres supports more than 10,000 customers worldwide running mission-critical database applications.

Datamatics Global Technologies GmbH (DGTG), a subsidiary of Datamatics Global Services Limited (DGSL), tested Ingres VectorWise for one of their existing applications that is currently sold using the Oracle Database. Michael Thuleweit, Managing Director for Datamatics in Frankfurt, said about this experience:

> *"We ported a business application from Oracle to Ingres VectorWise and were astounded by the substantial performance increases, which were up to 70x… With Ingres VectorWise Database, we get the best of both worlds; we can now offer our customers a lower cost and better performing analytic database."*

dbConcert is a leading professional services firm focused on delivering world-class implementations of open source software solutions, including: Financial Trading Systems, Business Intelligence, Customer Relationship Management and Database Management. Steve Ferrando, dbConcert CEO, said about their experience with Ingres VectorWise:

> *"Ingres VectorWise performance has been outstanding and it was so easy to get up and running. Position analysis was 10 to 50 times faster than the leading commercial databases in our testing and with linear query performance from 12 mm – 1.2 billion financial records. These are drastic improvements in terms of speed."*

Roy Hann at Rational Commerce, a leading systems integrator, said about his testing with Ingres VectorWise:

> *"The speed is blindingly fast right out of the box and it eliminates layers of work that had to be done in the past. Ingres VectorWise uses automated compression and indexing which eliminates the need to spend extra hours tuning and tweaking our data. This means we are now able to cut the time of our business intelligent engagements with our customers by up to 50%."*

Enable your customers to take advantage of the power of Ingres VectorWise. Learn more at http://www.ingres.com/vectorwise.

About VectorWise

VectorWise is a technology spin-off that resulted from 6 years of research at the CWI research institute. CWI has a proven track record in advancing computer science with projects like Python and MonetDB, the first analytical columnar database in 1994. The unique VectorWise technology has been incorporated inside the proven Ingres relational open source database, and is brought to market worldwide by Ingres in the Ingres VectorWise database product. The VectorWise innovation continues through an expanding community that includes many of the leading computer scientists from around the world.

About Ingres Corporation

Ingres is a leading open source database management company. We are one of the largest open source companies in the world and the pioneer of the New Economics of IT, providing open source solutions at a dramatically reduced cost compared with proprietary software vendors. As a leader in the New Economics of IT, Ingres delivers low cost and accelerated innovation to more than 10,000 customers worldwide.

Ingres Corporation
500 Arguello Street, Suite 200
Redwood City, California 94063
USA
Phone: +1.650.587.5500

Ingres Europe Limited
215 Bath Road
Slough, Berks SL1 4AA
United Kingdom
Phone: +44 (0) 1753 559550

Ingres Germany GmbH
Ohmstrasse 12
63225 Langen
Germany
Phone: +49 (0) 6103.9881.0

Ingres France
7C Place Du Dôme
Immeuble Elysées La Défense
92056 Paris La Défense Cedex
France
Phone: +33 (0) 1.72.75.74.54

Ingres Australia
Level 8, Suite 1
616 St. Kilda Road
Melbourne, Victoria, 3004
Australia
Phone: +61 3 8530.1700

For more information, contact ingres@ingres.com

INGRES

WP-316A