WINTER CORPORATION

WHITE PAPER

XTREMEDATA dbX

Product Review

The Large Scale Data Management Experts

WinterCorp

SPONSORED RESEARCH PROGRAM

W I N T E R   C O R P O R A T I O N

XTREMEDATA dbX

Product Review

May 2010

WinterCorp

245 FIRST STREET, SUITE 1800
CAMBRIDGE MA 02145
617-695-1800

visit us at www.wintercorp.com

©2010 Winter Corporation, Cambridge, MA. All rights reserved.

Executive Summary

THIS WHITE PAPER REVIEWS XtremeData's dbX product architecture and the business intelligence (BI) analytics environment the company aims to address with dbX. Research for the paper included interviews with representatives of three companies that have experience with dbX—Sandia National Laboratories, a large U.S. credit services organization, and Noetic Partners—to understand their respective requirements and assessments of dbX's capabilities.

The XtremeData dbX analytics appliance integrates the database, server, operating system, and storage in a single system. dbX combines a highly parallel appliance architecture with the ability to execute selected database functions using FPGA (field-programmable gate array)-accelerated software and high-speed data movement using InfiniBand network fabric. The goal is to deliver high-performance data loading and ad hoc query execution at a relatively low cost and minimal energy consumption.

With dbX, XtremeData is focusing on "deep analytics," an emerging area of BI analytic requirements. Users of deep analytics are typically small groups of skilled analysts and statisticians that need access to extremely large volumes of raw, detailed data; access to the full dataset in addition to samples or summaries of data; ad hoc, custom analytics; and good performance and scalability in the presence of large volumes of data and unpredictable data access patterns.

dbX features and design approaches that support a deep analytics environment include:

- Fast data loading without requiring advance data modeling/partitioning information from users
- High performance query processing achieved through (1) execution of selected database operations using FPGA-accelerated software rather than the processor, and (2) dynamic redistribution of data and load balancing among data nodes at run time, employing techniques to reduce skew
- Low cost per TB of user data capacity

The three organizations profiled in this white paper conducted proof-of-concept (POC) tests on dbX and cited all of these factors as key to the decision to implement dbX in their deep analytics environments.

In the opinion of WinterCorp, organizations that need a platform for deep analytics will want to consider XtremeData's dbX. A key component of the evaluation will be a realistic POC that tests for data loading and query performance on the anticipated volume of analytic data and range of query complexity to assure the organization that the platform chosen best meets the specific requirements for deep analytics.

Copyright © 2010, WINTER CORPORATION, Cambridge, MA. All rights reserved.

Table of Contents

Executive Summary 3
Table of Contents 4
1 Introduction 5
Purpose and Methodology for This Report 5
2 Targeting Deep Analytics 6
2.1 Overview 6
2.2 Speed of Implementation 7
2.3 Access to the Full Dataset 7
2.4 Performance and Scalability 7
2.5 Cost Effectiveness 8
3 XtremeData's dbX 8
3.1 Architecture 8
3.2 Functional Components 9
3.3 Design Features of dbX for Deep Analytics 10
3.4 Futures 12
4 Customer Cases 12
4.1 Sandia National Laboratories 12
The Business Challenge 13
The XtremeData Solution 14
Summary 15
4.2 Large U.S. Credit Services Organization 15
The Business Challenge 16
The XtremeData Solution 17
Summary 18
4.3 Noetic Partners 18
The Business Challenge 18
The XtremeData Solution 19
Summary 20
5 Conclusions 20

LIST OF ILLUSTRATIONS

Figure 1: Characteristics of a deep analytics environment 6
Figure 2: dbX models and node resources 9
Figure 3: A multirack dbX configuration 10
Figure 4 XtremeData aims to achieve a 1GB per second sustained rate of data movement and processing across components 12

Copyright © 2010, WINTER CORPORATION, Cambridge, MA. All rights reserved.

1 *Introduction*

In today's world, a business intelligence (BI) solution is often required to satisfy a wide range of user and analytical needs. At one end of the spectrum is operational BI, where hundreds or thousands of users may need concurrent and fast access to BI data and analytics that support business operations. In the middle is BI as most people know it—a data warehouse environment that supports reporting, business analytics, end user queries, etc. At the other end of the spectrum is *deep analytics*, where a relatively small group of skilled analysts needs ad hoc access to very large volumes of data for complex analysis.

In a deep analytics environment, datasets can be hundreds of terabytes or petabytes in size and may represent many years of historical data or a long time series. A dataset may be a temporary one created for a specific analytical purpose. Users are typically small groups of skilled analysts and statisticians that need ad hoc access to the dataset and the ability to do customized data exploration and analysis. Examples of deep analytics applications include scientific simulations, finding patterns in historical financial transactions, and finding patterns in a network of relationships among people, events, and/or places.

Deep analytics presents a distinct set of functional requirements. These include:

- Speed of implementation – It should be possible to introduce new data into the analytics environment quickly;
- Access to the full dataset – The analytics environment needs the capacity to support the full dataset in addition to samples or summaries of the data; and,
- Good performance and scalability on ad hoc queries in the presence of large volumes of data and unpredictable data-access patterns

PURPOSE AND METHODOLOGY FOR THIS REPORT

This WinterCorp White Paper reviews the architecture of XtremeData's dbX, its ability to handle the requirements of a deep analytics environment, and the experiences of three XtremeData customers.

This White Paper was sponsored by XtremeData.

In developing this paper, WinterCorp operated as an independent industry expert, interviewing XtremeData customers and employees; reviewing product documentation; and critically reviewing product design, measurements and evidence in order to arrive at the descriptions and conclusions presented here. XtremeData was provided an opportunity to comment on the paper with respect to facts. WinterCorp has final editorial control over the content of this publication.

Copyright © 2010, WINTER CORPORATION, Cambridge, MA. All rights reserved.

2 *Targeting Deep Analytics*

2.1 OVERVIEW

As a new player in the data warehousing and analytics appliance market, XtremeData is focusing on solutions that address an emerging area of analytic requirements. This area represents a class of business intelligence users and applications that use sophisticated analytics and need ad hoc access to very large amounts of data under circumstances that often bump up against the limitations of existing data warehouses within their organizations. XtremeData describes this analytics environment as "deep analytics." The characteristics of deep analytics are quite different from those of both business and operational analytics (see *Figure 1*).

Datasets can be hundreds of terabytes or petabytes in size. The dataset might be previously archived historical data that, if it could be "unarchived" and analyzed, could be used to identify useful patterns of customer behavior. An example would be purchases or events that occur infrequently within a larger set of data, such as the purchase of a car or a washing machine within data on purchases of all types of products. Another dataset of interest might be a temporary one compiled for a particular purpose, loaded, analyzed and then unloaded from the data warehouse. Other examples include long time series (e.g., recording clinical trial patient data over many years), event data (e.g., recording RFID logs or click stream logs) and external data sources.

The analytics environment involves data exploration and analysis to answer the question at hand or provide the "aha" moment. The analytics can be described as both *broad* and *deep*—broad in the sense that a dataset can span years of history and deep in the sense that there is often a need to analyze raw data, not summaries or samples. The analyst may not know what he or she is looking for in the beginning, so aggregating data in advance does not help—and may actually mask an interesting pattern waiting to be discovered. The analyst often uses custom analytics and each analytics project may have very different requirements. Therefore, both the analytics and access to the data can be considered ad hoc.

Typical users are skilled analysts, engineers, scientists, and statisticians who need to make sense of very large and complex data environments. Examples include risk managers, algorithmic traders, quantitative analysts, customer insight specialists, vehicle designers, PhDs/academics and researchers. It is also important to point out that the number of users in any group involved in deep analytics is usually relatively small, ranging from a few users to 40 or 50 users. This is very different from the operational BI environment, where hundreds or thousands of users need access to BI data and analytics. Thus, the impetus to implement a data warehouse platform that supports deep analytics often comes from an individual analytics group or department. In addition, the need for a high degree of concurrency control and high availability are often less important than ad hoc query performance.

Figure 1: Characteristics of a Deep Analytics Environment

Users	Skilled analysts and statisticians Relatively small user groups (ranging from a few to 40 or 50)
Datasets	Extremely large volumes of raw, detailed data User has access to the full dataset in addition to samples or summaries of data Datasets may be permanent or temporary
Analytics Environment	Ad hoc, custom analytics Unpredictable data access patterns

Copyright © 2010, WINTER CORPORATION, Cambridge, MA. All rights reserved.

Deep analytics applications include:

- *Scientific simulations* – An example is weather and climate analysis and forecasting;
- *Finding patterns in historical financial transactions* – Examples include tick-stream (trading) analysis, modeling and analysis of financial instruments, portfolio analysis and mortgage analysis. The knowledge gained is used to develop current or future trading strategies, risk management assessments and strategies, customer service policies, etc.;
- *Finding patterns in a network of relationships among people, events, and/or places* – This type of analysis can yield critically important information, such as security intelligence or enabling an organization to better understand its customers; and,
- *Analyzing medical research and clinical trials* – This is a good example of a long time series (recording statistics, such as cholesterol and blood pressure, over time for a very large number of patients) and making the data available over many years for analysis.

The three customer case studies in *Section 4* provide a more in-depth description of some of these applications. Here is a high-level summary of the common requirements of a deep analytics environment.

2.2 SPEED OF IMPLEMENTATION

A major requirement of deep analytics is fast implementation—how fast the data environment can be up and ready for analysis. Waiting months for IT to complete a requirements analysis, design the data model, and create the data structures does not fly in these applications. Once the need is recognized, time is of the essence and it should be possible to create the database and load the data for analysis in a short period of time. Important features that support this include the ability to bring datasets into the data warehouse as needed, very fast loading and unloading of data, and the ability to provide good performance for queries without having to modify the existing data model or partition the data in a particular way.

It is also desirable to create an analytics environment that does not demand strong data-modeling skills from the user. Queries do not necessarily have to be known beforehand in making decisions about physical placement of data. The lifetime of the database is often relatively short and projects may last weeks or a few months. Therefore, the user's tolerance for spending significant time in a data-modeling exercise may be low.

2.3 ACCESS TO THE FULL DATASET

Access to the complete dataset is also important for deep analytics. In many cases, the analyst starts by analyzing a sample of the full dataset. This can limit the scale of the problem so it can be solved in a smaller data environment. However, if the sample analysis yields interesting results, the user may want to load and analyze the full dataset, which could be many years of detailed history as described above. The data warehouse environment needs the capacity to support the full dataset in addition to samples or summaries of the data.

One issue here can be the fact that the information density of the full dataset is low. That is, it may not be clear in advance that there is value to be gained from analyzing a very large dataset in terms of return on investment (ROI), and this can affect an organization's ability to justify spending the time and money to create an analytics platform for the data.

2.4 PERFORMANCE AND SCALABILITY

One of the major challenges in this analytics environment is to provide good performance and scalability in the presence of large volumes of data and unpredictable data-access patterns.

Copyright © 2010, WINTER CORPORATION, Cambridge, MA. All rights reserved.

2.5 COST EFFECTIVENESS

Another challenge is bringing a level of performance to this analyst user community that enables them to get their work done at a price point they can afford considering that the data volumes can be very large. The need to minimize energy consumption is also important as this affects operating costs and addresses a growing concern in today's data centers.

3 *XtremeData's dbX*

XtremeData's dbX aims to address the needs of users that need a platform for deep analytics. The dbX analytics appliance integrates the server, operating system, storage and database engine in a single system. dbX combines a highly parallel appliance architecture with the ability to execute selected database functions using FPGA (field-programmable gate array)-accelerated software and high-speed data movement using InfiniBand network fabric. The appliance is designed to deliver high-performance data loading and ad hoc query execution at a relatively low cost and reduced energy consumption.

3.1 ARCHITECTURE

dbX is based on the following core architectural components:

- Commercial, off-the-shelf Linux blade servers (such as HP Proliant DL185 or DL385) that each support two multicore processors and direct-attached SAS disks for distributed data storage;
- InfiniBand high-speed interconnect among components with the ability to move data at 2GB per second bi-directionally;
- PostgreSQL (Postgres) open-source database management software; and,
- FPGA-accelerated software using XtremeData's patented In-Socket Accelerator (ISA) technology implemented on FPGAs from Altera.

XtremeData has customized these components for dbX in several significant areas. First, the company re-engineered its implementation of Postgres. While XtremeData retained the Postgres front end (standard ODBC and JDBC interfaces, language, parser, optimizer, etc.) with some modifications, it completely replaced the back-end database engine with XtremeData's own parallel query execution engine to take full advantage of the dbX architecture. In addition, XtremeData executes many database functions in the FPGA instead of the general-purpose processor to improve performance. According to XtremeData, the FPGA can execute these functions faster than the processor while using one-third the power.

XtremeData has programmed the FPGA to take advantage of both parallel processing and pipelining. This makes the FPGA scalable in both width (parallelism that handles concurrent processing of data) and depth (pipelining that handles sequential processing of steps in the query process). The key, according to XtremeData, is to provide multiple, parallel pipelines of data-processing steps with no bottlenecks.

Second, the FPGA fits on a carrier board that is pin-compatible with the processor on the blade server. This enabled the company to replace the second processor slot on the server's motherboard with the FPGA, making the FPGA a master with the same access to resources as a processor. The FPGA is directly connected to both memory and the I/O subsystem, avoiding the need to involve the processor when moving data to/from either of these resources. XtremeData considers this "hybrid" computing environment one of its distinct advantages in the marketplace. The company's approach is to use commodity components to the extent possible to keep pricing affordable and

Copyright © 2010, WINTER CORPORATION, Cambridge, MA. All rights reserved.

then layer on proprietary technologies and components to achieve the required high level of query performance.

3.2 FUNCTIONAL COMPONENTS

A one-rack dbX system has a *head node* and up to 16 *data nodes*. Each additional rack has a *coordinator node* in addition to the data nodes. As racks are added, the total number of data nodes and volume of user data supported are not linear because as the configuration grows, the switch gets bigger and takes up additional node slots in the first rack. Each node has its own server with processors(s), 32 GB of memory and disk drives. Each data node also has an FPGA as described above (see *Figure 2*). All data nodes are peers.

The smallest dbX configuration is a half-rack system with eight data nodes that supports 30TB of user data. The largest configuration (in theory, as XtremeData has not yet built one), is a 66-rack system with 1024 data nodes that supports 3.8 petabytes of user data. User data is roughly one-third of total disk capacity.

Figure 2: dbX Models and Node Resources

Model	dbX 1008	dbX 1016	dbX 1028	dbX 1060	dbX 11024
Rack	1/2	1	2	4	66
Data Nodes	8	16	28	60	1024
User Data	30TB	60TB	105TB	225TB	3.8PB

Head Node	
CPU	2xAMD 6 Core
Memory	32 GB
Disk	16x 300GB SAS
FPGA	—
Interconnect	InfiniBand, GigE

Data Node	
CPU	AMD 4 Core
Memory	32 GB
Disk	12x 1TB SAS
FPGA	XD2000
Interconnect	InfiniBand, GigE

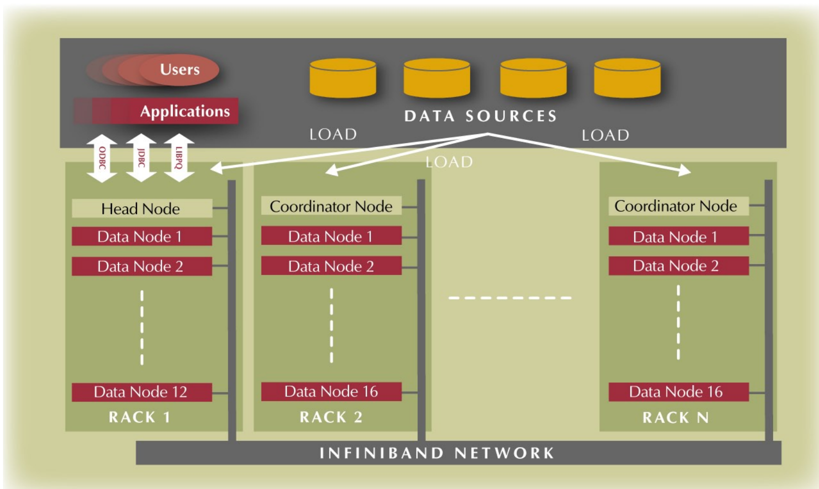
The head node runs the front-end processes for dbX (i.e., the front end of Postgres), including user, tool, and administrator interfaces; user sessions; the SQL processor; and system-level merge operators (min, max, average, etc.). Essentially, the head node receives query requests from users/applications, optimizes the queries, sends the query plans to the data nodes for execution, and then assimilates and sends query results back to the user. In a single-rack system, the head node also coordinates data loading/unloading (see *Figure 3*).

Each data node contains the parallel database execution engine and performs the back-end query execution. The processor manages the overall execution of the query in the data node, sending specific steps of the query plan (e.g, joins) to the FPGA. The processor also handles indexed and sequential scans of data, some SQL processing such as execution of user-defined functions (UDFs), and all inter-node communications via Ethernet (GigE). The FPGA handles key functions in the SQL query plan (data load/unload, scans, filters, joins, sorts, group by, order by, and aggregations).

Copyright © 2010, WINTER CORPORATION, Cambridge, MA. All rights reserved.

Each data node also houses 12 direct-attached 1TB SAS drives, each of which contains a slice of user data. dbX supports hash, range, and round robin data partitioning. If the user does not specify a partitioning method, dbX does the following: for tables with a primary key, dbX uses it for partitioning and hashes on it; for tables with no primary key, dbX uses round robin to eliminate data skew among nodes. dbX implements B-tree indexing and supports unique and multicolumn indexes.

Figure 3: A Multirack dbX Configuration



The coordinator node provides rack-level coordination and execution support as an offload to the head node, including staging of data (load/unload, backup/restore), parallel fast load of data (bypassing the head node), and rack-level merge operators (min, max, average, etc.).

XtremeData has also addressed resiliency in dbX in several areas. These include support for dual-parity RAID 6 to provide protection for up to two disk failures on one controller; support for GigE as a backup interconnect for InfiniBand (GigE is also used for internal messaging among components); data-mirroring and hot spares in case of data node failure; and a hot spare in case of head node failure.

3.3 DESIGN FEATURES OF dbX FOR DEEP ANALYTICS

This section highlights the features and design approaches of dbX that support a deep analytics environment.

Fast data loading. Fast loading of large amounts of data without requiring any advance data modeling or partitioning information from the user is a key design goal for dbX. Accomplishing this addresses several deep analytics requirements: speed of implementation, access to the full dataset when needed, and high performance and scalability.

While the user can specify a partitioning method, dbX does not need that to load data and distribute it across all the data nodes. Data is loaded in parallel across all data nodes in the system. In a single-rack system, the head node divides the data to be loaded into "chunks" for each data node, and each data node then loads its data. In a multi-rack system, loading is also done in parallel across racks. The head node chunks the data for itself and each coordinator node. The head and coordinator node(s) then manage the data loaded in parallel by each of their own data nodes.

Copyright © 2010, WINTER CORPORATION, Cambridge, MA. All rights reserved.

High-performance query execution. High-performance query execution is one of the most important requirements of a deep analytics environment. dbX is designed to address this in three significant ways.

First, XtremeData takes a distinctive approach to improving query performance with its patented "SQL in Silicon"™ feature in which dbX executes selected database operations using FPGA-accelerated software rather than in the processor.

A second important feature in dbX is dynamic data redistribution with the related goal of load balancing among data nodes at run time. There are two issues here. One is the need to redistribute data dynamically among nodes at run time for efficient execution of joins on non-co-located data. The second is the need to effectively balance the load at run time to ensure all nodes are involved in the join operation to maximize the benefits of parallelism. (XtremeData describes load balancing among nodes as a critical factor that can limit the scalability of parallel processing systems.) dbX automatically assesses both of these needs at every intermediate step of query execution.

dbX gathers statistics at each branch in the query tree and dynamically repartitions the data as required to (1) co-locate data for joins and (2) ensure that every node participates regardless of data skew or cardinality. To do this, the FPGA, in coordination with the processor, builds statistics and a sample histogram on the join key in question. dbX uses the statistics to redistribute the data directly to the appropriate data nodes so that join data is now co-located on every node.

Another data-redistribution performance strategy is to broadcast the data on one side of the join to all data nodes. dbX may do this for Cartesian products; one side of the join is broadcast and the other is partitioned.

According to XtremeData, the InfiniBand interconnect pathways are key to enabling dbX to dynamically repartition data and maintain fast query performance. XtremeData's philosophy is that careful partitioning of data is not sufficient to maintain high performance. Eventually there will be a join that cannot be performed efficiently using static partitioning. So dbX has been designed to move data as fast as possible between nodes in a "just in time," highly pipelined fashion.

A third design goal is maintaining a consistent speed of data movement and processing among components. XtremeData aims to achieve a 1GB per second sustained rate across components. This is based on the speed of reading data per node (12 disks per node reading at 80 to 85MB per second per disk). The overall speed is also affected by the fact that the rated bandwidth between specific components is often substantially higher than that (see *Figure 4* below) and the fact that the data rate drops below 1GB/second under certain circumstances (e.g., a disk is both reading and writing data, the query is very complex, etc.). The goal is effective system utilization, with all components busy and data moving through at a constant rate of 1GB/second per node. Ideally, this makes the cost of reading data the same whether the data is local or remote (i.e., reading local data at 1GB/second is the same as moving the data to a different node at 1GB/second).

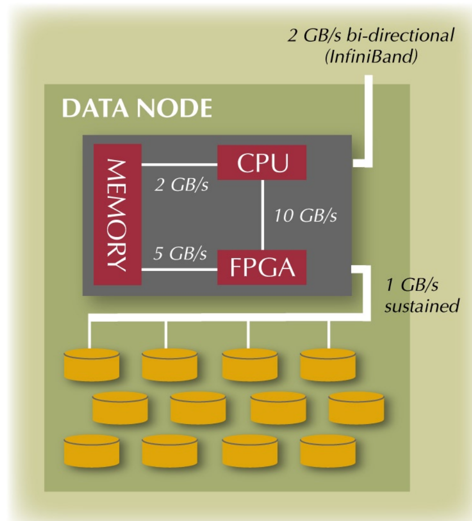
Low cost. dbX currently lists at $20,000 per TB of user data capacity. For example, a half-rack system with 30TB of user data costs $600,000; a full rack with 60TB costs $1.2 million. It is important to note that this pricing is based on uncompressed data. XtremeData expects the cost per TB of user data to decrease once dbX supports data compression.

Prospective buyers need to be aware that per TB prices are based on the vendor's estimate of a typical workload. The actual list price of a system will depend on how a customer's workload varies from this average.

Copyright © 2010, WINTER CORPORATION, Cambridge, MA. All rights reserved.

Energy efficiency. According to XtremeData, FPGA power requirements are only one-third that of the processor: 30 to 40 watts for the FPGA versus 120 watts for the processor. Therefore, using an FPGA instead of a second processor not only enhances performance but also saves 80-90 watts per server. The total saved could be hundreds of kilowatts in a very large database environment with many nodes. This reduces heat generation in addition to power consumption.

Figure 4: XtremeData Aims to Achieve a 1GB per Second Sustained Rate of Data Movement and Processing Across Components



3.4 FUTURES

XtremeData has identified areas for future dbX enhancements. One is implementing new functions in the FPGA. A key candidate is data compression, which would increase the data capacity of existing dbX systems and drive the cost per TB of user data even lower than it is now. The company is targeting a 3-4x reduction in data space required. The compression algorithm is column-based, will use different types of compression depending on the data type, and will work against all general data types, including char and varchar. The FPGA will handle both compression and decompression.

XtremeData also plans to increase the number of FPGAs in each node and support 2TB disk drives.

Another area for future development is handling upgrades in place. For a customer going from one rack to two, XtremeData will build two racks on site and redistribute data from the first rack to the second one.

4 Customer Cases

As part of the research for this report, WinterCorp interviewed the representatives of three organizations that have experience with XtremeData's dbX platform. The goal was to understand the business challenges these companies faced, why they chose XtremeData, the benefits they achieved or expect to achieve, and their assessment of XtremeData's capabilities.

These customers also provided valuable insight into the types of analytical applications that XtremeData is targeting with dbX, that is, "unconstrained analysis and exploration" of very large datasets.

4.1 SANDIA NATIONAL LABORATORIES

Sandia National Laboratories is a U.S. Department of Energy facility created in 1949 to support national security needs through the development of science-based technologies. Historically, the primary mission has been the safety and surety of the U.S. nuclear stockpile. This focus has shifted over time toward "closer to home" aspects of national security, including homeland security, infrastructure and sustainable energy.

The lab is a government-owned, contractor-operated facility managed by Sandia Corporation, a Lockheed Martin company. Sandia has over 11,000 full time staff and contractors located in

Copyright © 2010, WINTER CORPORATION, Cambridge, MA. All rights reserved.

Albuquerque, New Mexico and Livermore, California plus a number of smaller offices around the world.

WinterCorp interviewed Dr. Andrew T. Wilson, Senior Member of Technical Staff, in the Visualization and Data Analysis group for this white paper.

The Business Challenge

Analyzing extremely large datasets is a major challenge for the Visualization and Data Analysis group. According to Wilson, "In some ways, visualization is currently the easy part. The harder part is handling large datasets and getting them into a form where we can usefully analyze them. This is what prompted Sandia to consider a parallel data warehouse/appliance environment." Wilson described two application areas of particular concern to Sandia.

The first is *scientific simulations*. When designing a new car, it might be possible to crash-test 10 cars in a lab. But if one needs to do this several thousand times, it is much more cost-effective to conduct a finite element simulation. A more relevant example for Sandia is global climate forecasting. A climate model is typically run as a set of many similar variations of a given scenario in an effort to synthesize a consensus and likelihood on what the simulation says is going to happen in various locations. Will regional temperatures rise or fall? How will patterns of precipitation change? How quickly will these changes occur? For each likely outcome, what is the best estimate of its likelihood with associated uncertainties? An analysis of the data from one of these simulations uses a database to store both raw data and summary statistics. The raw data (i.e. columns in the main data table) comprises a few hundred different numeric parameters for each data point (tens of thousands to tens of billions) in the simulation domain. The summary statistics include the mean, variance, median, min, max, and 1st and 3rd quartiles (which bound the middle 50% of data). The organization then writes a visual analysis application to extract this summary information (usually 10 to 100 times smaller than the original dataset) and further distill it to produce a summary that can be presented to a skilled user. The user then generates ad hoc queries to analyze the data with the ability to drill all the way down to the original, detailed data. The role of the database here is fast access to exactly the subset of data requested and fast summaries of data, with summaries calculated on the fly if necessary.

Wilson says this type of query fits fairly well into SQL with some custom code on the application side. Primary operational requirements for scientific simulations include scans, range queries, and aggregates (typically group bys and window functions).

In terms of data volume, "The weather and climate folks consider 100GB to be fairly small and 200TB to be more like what they want. These numbers continue to grow without bound. The best projections in the climate community, for example, suggest repositories of hundreds of exabytes within the next ten years. Some security missions today generate close-to-petabyte datasets and we expect to have petabyte-scale datasets within the next 12 months. So these are not traditional business analytics datasets; they deal with very simple schemas and gargantuan amounts of data."

The second application is *graph analysis*. In this environment, a graph is defined as a generalized network of data objects where any object can be related to any other object (and frequently is). A network of social relationships among people is a good example of a graph data structure. Here, components include nodes or vertices (each person) and edges or arcs (the connection between two people/nodes). The graph is a set of pairs (e.g., person A and person B are friends, person C and person D lived together, etc.) that are interconnected in a potentially very complex network. One might want to find a path through the graph or find a set of similar relationships.

Copyright © 2010, WINTER CORPORATION, Cambridge, MA. All rights reserved.

A very simple example of graph analysis would be querying a publication database (e.g., PubMed) to find groups of five or more authors who publish together all the time. Another common graph-based query is subgraph isomorphism. Here, the user defines an "example" graph representing a pattern of interest or social relationships for further investigation and asks the system to find all instances of that in the graph.

This differs from the typical database paradigm where, even in a very large data warehouse, there might be millions of objects but only a hundred or fewer relationships. In a graph environment, there can be billions or trillions of objects in the network and the rule of thumb is that each one has at least 10 connections or relationships (i.e., the number of relationships is *always* greater than the number of objects by a factor of ten or more). This clearly escalates the database scalability requirement. As Wilson says, "'Challenge' is a polite word for this problem."

Graph analysis essentially combines graph computing with analysis of very large datasets. Sandia's existing solutions all have serious limitations. Supercomputers, good at graph computing, do not handle very large datasets well because the I/O is "painfully slow compared to computational power." Commercial relational databases can handle large datasets, but they have difficulty with graph computing, the relatively simple but computationally difficult algorithms that must be implemented in SQL and UDFs. "In particular," according to Wilson, "a shared-nothing parallel architecture becomes a real handicap when doing large graph calculations because it is impossible to partition the data in a way that effectively partitions the processing."

Here are other characteristics of the Sandia analytics environment:

- *Database longevity* – Sandia needs to build and use both long-term (months or years) and temporary (a few days) petabyte-scale databases.
- *Frequency of updates* – Sandia typically has new data coming in daily that is loaded overnight. The organization does not need to react to new data more quickly than that.
- *Interactive vs. long-running queries* – Sandia sees the need for both. "For some analyses, it is OK that they run for a day. Sometimes we need an answer in 30 seconds."
- *Data compression* – This is not of significant value; Sandia's data involves a lot of floating point and varchar fields which do not compress well in Sandia's experience.
- *User community/concurrency* – Each analytics community is typically 10 or fewer users and there are a relatively small number of queries running at one time.
- *Analytics tools* – Sandia writes its own analytics tools using a combination of JavaScript, Python, and C++. Sandia uses SQL to retrieve the data and does analysis in the application. If some of the analysis can be pushed down into SQL, Sandia will take advantage of that.

Sandia decided to experiment with data warehouse appliances that use FPGA-accelerated software for selected database functions in an effort to achieve the required performance on large datasets for its target applications.

The XtremeData Solution

Sandia recently completed two proof-of-concept (POC) tests with XtremeData and these demonstrated that dbX could improve performance by a factor of two to three over the incumbent solution. As a result, Wilson submitted a proposal for funding to acquire an XtremeData single-rack system, designed to house 60TB of user data, on which to conduct further performance tests.

Performance is key for Sandia. As Wilson put it, "Essentially, a parallel database appliance is the access layer, or index, to a dataset too large to fit into memory on almost any machine we have.

Copyright © 2010, WINTER CORPORATION, Cambridge, MA. All rights reserved.

Because we are doing very simple calculations on very large datasets, a solution like XtremeData that pushes the computation down close to the data has distinct advantages. It is the opposite of the way traditional supercomputers work, where a huge amount of compute power is concentrated on a relatively small amount of data with limited I/O requirements."

In the case of conventional parallel DBMSs, Wilson cites the overhead of indexing as a major concern in both performance and storage requirements. Sandia's data warehouse group estimates that index structures require three times again the storage space of the original data. This is not a problem for a 100GB dataset, but it can be impractical for a 100TB+ dataset. "The ability to throw many disks at the data using little or no indexing and still get good performance is a huge win for databases with FPGA-accelerated software."

Another issue for Sandia is the ability to design the database to accommodate all the different data sources involved, the many types of questions that might arise on the same dataset, and all of the possible relationships for a set of objects with the ability to easily extract the data structure needed for a particular analysis. "This is a hard problem. We don't have the luxury of designing the database to accelerate specific classes of questions or analysis. We often have to create a graph on the fly in response to a request."

Cost is also an important consideration. Wilson described a target of 200TB for $2 million or less, or $10K per TB. (The *ideal* would be $2K to 4K per TB, enabling Sandia to explore datasets of 500TB or more for the same cost). Achieving the current target would bring significant value to Sandia's mission and propel the research program ahead three years. While currently available systems are not in this price range, they have improved performance to the point where a query that would have taken months previously now takes only 10 minutes.

Sandia does not have full redundancy on its incumbent system because of cost, but it is very important that the system continue to run when a disk fails.

Wilson describes these additional features of XtremeData's dbX that are needed in Sandia's environment:

- *Support for user-defined types (UDTs) and functions (UDFs)* – This is an important feature. Many of the most common operations in both graph analysis and simulation data analysis run orders of magnitude more quickly in UDFs than when they are expressed solely in SQL.
- *Index support* – An example is range queries where a simple index can provide significant performance benefit. However, indices are not a make or break capability for Sandia and there is no need for "bullseye" indexing to pull out individual rows of data.
- *Fast data load and unload* – dbX supports direct parallel loading and unloading of data via multiple streams using InfiniBand and a coordinator node per rack.

Summary

Wilson is satisfied with the relationship with XtremeData. XtremeData is clearly targeting applications like Sandia's, applications that are heavily customized and different from traditional business intelligence analytics. XtremeData is also willing to enable Sandia to do its own onsite maintenance and support. This is critical in cases where a machine is located in a secure area.

4.2 LARGE U.S. CREDIT SERVICES ORGANIZATION

This organization is a leading provider of consumer credit and information management services to thousands of businesses and hundreds of millions of consumers worldwide. Products and

Copyright © 2010, WINTER CORPORATION, Cambridge, MA. All rights reserved.

services include credit risk and portfolio management, online access to and analysis of a wide variety of consumer financial data, and marketing.

WinterCorp interviewed both the vice president and a senior manager in the Data Quality Services group for this white paper.

The Business Challenge

The primary function of Data Quality Services is protecting and monitoring data quality. This group is responsible for analyzing and ensuring data quality, developing data-transformation business rules, and finding opportunities to improve the performance of the data. A related function is to identify ways to leverage the data for analytic purposes. An analytics example would be providing loan-portfolio management for a large business customer (e.g., overall trend analysis, account activity, account origination, scoring of accounts nationally and regionally, etc.).

Over the past five to seven years, the organization has been upgrading all of its business systems. Several components are key to the mission of Data Quality Services. One is the core consumer credit database managed by a relational DBMS on the mainframe. This database contains detailed financial data on hundreds of millions of consumers. The largest table has data on four billion accounts with two billion update transactions each month (equivalent to almost 67 million transactions per day). Another component is a data warehouse, called the research investigation database, managed in a column-oriented Unix RDBMS. This is the group's primary analytics platform. Data from the core credit database is copied daily to the data warehouse and analyzed using SQL and the Ab Initio ETL (extract/transform/load) and data manipulation toolset.

While this architecture meets current needs, the organization wants to grow the data warehouse to meet future needs by providing more data, more dynamic access, and more analytic capabilities. One alternative is to add capacity to the existing system. Another is to add other systems/platforms to build out the desired functionality. Cost is a driving issue.

Here are some major goals for this "growth" effort:

- *Leverage additional data sources and more detailed data* – The current data warehouse largely mirrors the production database (although it doesn't contain all of the data). The group would like to present different views of the data as well as include more transaction-level data.
- *Expand analytics functionality* – One example cited by the group's vice president is better analysis of matching results. "When we get a request for financial data, we may only get a name and address with no unique key. We would like to be able to better analyze our 'hit rates' on identifying the correct consumer to monitor this and find opportunities to tune the algorithm." Another example is to develop risk management tools for mortgage securitization products.
- *Provide access to historical data* – The current data warehouse is essentially a daily snapshot of the core credit database plus various scores and attributes built on this snapshot for each account or person. Access to two years of history (e.g., loan payments, calculated scores) would be useful to monitor overall trends and to build more risk analytics that can tap into these historical trends to better predict financial reliability and behavior.

The data warehouse has about 14TB of data today and as stated above, this is a subset of the data in the core credit database. According to the senior manager, "14TB could easily increase by two or three times [28-42TB] if we had the budget and systems to include all of the data sources and current detailed data that we would like to have from the core credit database. Most of the time

Copyright © 2010, WINTER CORPORATION, Cambridge, MA. All rights reserved.

we can be effective taking a 10% random sample of the core credit database for our analysis needs. Adding 24 months of historical data pushes the size to hundreds of terabytes." With this increase in scale, the objective is to find a cost-effective system going forward that can manage this data volume while processing up to 100 million record updates per day.

The current user population of 20 represents a wide range of skill sets. Seven are core team members that use both SQL and Ab Initio (which requires sophisticated users) to run very large, complex queries. Others include database administrators, experts on testing and matching algorithms, and general data acquisition/data quality analysts. As the platform grows, the goal is to expose it to other business users in the analytic services area who design risk models on behalf of customers. These users currently come to Data Quality Services for data extracts and work with small subsets of the data (e.g., 10% samples of the credit database). Giving them direct, self-service access to more data would be valuable and would increase the overall number of users to 30 or 40.

The XtremeData Solution

The organization conducted a proof of concept (POC) test on a beta version of XtremeData dbX in early 2009. XtremeData had contacted the company to introduce the product and garnered interest in the overall design approach. The credit services company described as "new and interesting" XtremeData's combination of "off-the-shelf hardware and software, proprietary technology (the ability to execute some database functions in the FPGA)" and its claims of fast query performance and low cost. XtremeData brought the POC system onsite and here is a brief summary of the POC experience and results:

- *Ease of installation* – The company described the overall installation as "easy." According to the vice president, "We put the appliance in the data center, plugged it in, connected it to our network and were up and running and ready to load data within five minutes." The POC machine was a prototype eight-node system that could accommodate 8TB of user data (1TB per node).
- *Physical database design* – "We gave XtremeData the schema and once the database structure was created, there was no need to specify any partitioning or run any statistics prior to executing queries. No DBA involvement was required, which surprised us. We had no idea how XtremeData partitioned the data."
- *Preparation and execution of test queries* – The organization ran a series of data load and query execution tests with a mix of query types. The simplest queries requested specific data in one table (e.g., maxcount, extract a row based on an exact match in a single column). Moderately complex queries included additional columns and tables, group by aggregations, summaries, etc. The most complex queries joined two or three of the largest tables (e.g., four billion rows, 800 million rows, 300 million rows). The organization did no optimization or tweaking of queries to get them to run faster other than to make syntax changes to get the queries to run. This provides an example of XtremeData's stated approach to POCs in which it simply runs existing queries and does not do any special query optimization.
- *POC results* – XtremeData's dbX achieved the most dramatic performance results on very large, complex queries (in general, a factor of 10 or better performance). For example, on one large, long-running query, the incumbent system took 1.5 hours and dbX took four minutes, an improvement of more than 22x. On the largest query tested, the current system took 4.5 hours and dbX took 15.5 minutes, an improvement of more than 17x. (This query is one the

Copyright © 2010, WINTER CORPORATION, Cambridge, MA. All rights reserved.

organization usually saves for the weekend in the current environment.) On medium-complexity queries, the existing system and dbX were in the same ballpark. On very small queries that involved one or two columns, the incumbent system beat dbX by a factor of as much as seven times because the incumbent system benefits from a column storage model and sophisticated indexing.

According to the vice president, "For the most part, dbX met or exceeded our expectations." (As a note, at the time of the POC, there were limitations on the beta system tested. It could run only one query at a time and could not load data and execute a query concurrently. XtremeData has since addressed these issues and also stated it has rerun the POC tests.)

Based on the POC results, the organization has requested funding to acquire an eight-node dbX system. If further testing goes well, the plan is to run the existing data warehouse system and dbX in parallel, keeping the easier queries where they are and moving a subset of the data; newer views of the data; and larger, more complex queries to dbX. For example, the organization would add new analytics data (e.g., 5-10 years of mortgage data) and monthly snapshots of historical financial data to the dbX platform. Thus, dbX would be a complementary platform for building out the desired extended capabilities described above with dbX eventually replacing the existing system over a year or two.

Summary

The credit services company cites the following as advantages for XtremeData: dbX's FPGA-accelerated implementation of key database functions as a credible source of performance improvement; XtremeData's value proposition in terms of price/performance; and dbX's ability to balance processor, I/O, and storage capacity so the customer doesn't have to worry about this. The company also mentioned the opportunity to work with XtremeData and a pre-production version of dbX to provide suggestions concerning further development of the dbX product.

XtremeData has been "very helpful, bending over backwards to ensure that the POC process was acceptable and contributing personnel to get it done quickly and smoothly. The resources are knowledgeable, always accessible, and they have been great to work with."

4.3 NOETIC PARTNERS

Noetic Partners is a consulting firm whose partners have extensive experience in building financial services systems for Fortune 50 companies. Founded in 2009, the organization delivers market data and reference data strategies and systems to assist client organizations in managing a wide range of entities, events, products, geographies, etc. It also focuses on data warehousing and advanced analytics systems for large financial companies.

Current clients include a large global bank and a large global stock exchange group. Noetic is also working with a partner to develop its XtremeData offering, described in more detail below.

WinterCorp interviewed Justin Magruder, president and managing director, for this white paper.

The Business Challenge

Magruder described two areas in which Noetic Partners believes XtremeData can provide a solution to significant business problems.

All of Noetic's customers face massive data management challenges, including high volumes of data, lots of customers and transactions, high throughput needs, and very complex decision-support requirements. These companies are also saddled with legacy technologies, often from

Copyright © 2010, WINTER CORPORATION, Cambridge, MA. All rights reserved.

multiple vendors, and see the need to get more information out of their terabytes and petabytes of data. Noetic is working with these customers to find more cost-effective ways to accomplish this. XtremeData is one alternative given its cost, performance and scalability claims.

Noetic is also working with a partner and XtremeData to build an advanced analytics system for the financial industry. This system integrates XtremeData's dbX with data feeds to create a consolidated database of financial transaction activity. According to Magruder, "The volume of consumer and institutional financial transactions today is massive, and very few environments can record the full stream and keep it available for analysis for more than one or two days."

Few companies have been able to build a truly comprehensive financial transaction data warehouse that is directly available to users for ad hoc analysis. The scalability and performance of near real time analytics is expected to have high value.

The XtremeData Solution

Noetic Partners' approach is to build an architecture using XtremeData's dbX that enables the system to capture financial data feeds as quickly as possible and make the data available to analysts and business managers in near real time. This involves the following:

- *Leverage the FPGA for ETL processing* – Magruder describes the ability of dbX to process key database functions in the FPGA as critical to its query performance. Noetic is developing software to leverage the dbX FPGA to support ETL (extract, transform and load) processes.
- *Load transactions intraday* – These are indexed at load time and made available immediately for analysis in a very high-performance query environment.
- *Include every record from every source* – Noetic plans to include all transactions as they become available.

According to Magruder, the cost to purchase and collect various types of data today can range from thousands to millions of dollars. Historically, analysts have been constrained by the volume of data, performance and time to market for new analytics, such as building joins between two or more complex data sets. Noetic intends to provide its database in a scalable and high-performance environment using XtremeData's dbX.

Magruder says that "In the past, organizations have been constrained by platform resources so they try to keep file sizes small. Most analysts spend 50 to 75% of their time scrubbing the data and getting it into a usable condition. We want to take that overhead away and accelerate the decision-making process for the analyst. Organizations in the past have been unable to accomplish this because of cost and scalability issues. The key is to make a highly scalable, ad hoc analytics environment available on demand at a significantly lower price/performance point. There is a lot of potential here."

Magruder gave as an example an experience he had at a previous employer. "This organization had hundreds of terabytes of data and hundreds to thousands of users and was constantly trimming the size of the database to make it more available. The processor was operating at 110% and user queries were often knocked off the box." XtremeData's claim of "unconstrained" access to data and query capability changes the equation and the game for Noetic. Noetic cites the need for its customers to model millions of instruments and billions of related events, and to build massive joins across multibillion-row tables.

When asked what convinced him that dbX could handle this price/performance requirement, Magruder responded, "First is cost per terabyte. We have priced it several different ways and

Copyright © 2010, WINTER CORPORATION, Cambridge, MA. All rights reserved.

dbX is cheaper. Second is scalability and performance. We have observed several proof of concept (POC) tests with current XtremeData customers and have conducted our own performance and stress testing. So we have first-hand experience with the performance of dbX and believe it has the necessary query performance."

Magruder estimates the data storage requirements for the Noetic Server are a few hundred terabytes per year with data compression and about a petabyte per year without compression. "History is very important to analysts, academics, and risk managers. They are looking for signals to anticipate what's going to happen next. The more history they have, the more patterns they see."

Summary

Noetic is in the pre-production testing cycle and is planning to use the dbX 1008 with 8 nodes and 30 TB of user data. The company plans to offer its product either as leased access to a shared environment or as feeds to the customer's own dbX box.

5 *Conclusions*

XtremeData is a new player in the data warehousing and analytics appliance marketplace. The company has introduced an interesting and distinctive product design at a relatively low cost. dbX is specifically designed to serve the needs of deep analytics, an application environment where XtremeData sees a significant opportunity to address unmet user requirements. Key goals for dbX are fast loading of data and high-performance query execution in the presence of large volumes of data and unpredictable access patterns. Major dbX design features aimed at supporting these goals include execution of selected database functions using FPGA-accelerated software, and dynamic redistribution of data and load balancing among nodes at runtime, employing techniques to reduce data skew. XtremeData has also designed dbX with cost and energy efficiency in mind.

All three of the customers cited these factors as key to their plans to implement dbX in their deep analytics environments.

An important concept here is that the dbX appliance can function as a complementary adjunct to the data warehouse. dbX is not designed to replace an enterprise data warehouse or support thousands of concurrent users, but rather to add value and functionality by offloading the execution of deep analytics.

In the opinion of WinterCorp, organizations that need a platform for deep analytics will want to consider XtremeData's dbX, recognizing that the company is early in the process of building a customer base. An in-depth evaluation should include a careful and realistic test of the customer's specific requirements. The proof-of-concept should test for data loading and query performance on the anticipated volume of analytic data and range of query complexity, thus providing the prospective user with evidence concerning the behavior of each platform considered, including XtremeData. The POC should also test the ability of the platform to handle multiple concurrent queries where expected. This will assure users that they have chosen a platform that best meets their needs.

Copyright © 2010, WINTER CORPORATION, Cambridge, MA. All rights reserved.

WinterCorp is an independent consulting firm that specializes in the performance and scalability of terabyte- and petabyte-scale data management systems throughout their lifecycle.

Since our inception in 1992, we have architected many of the world's largest and most challenging databases in production today. Our consulting services help organizations define business-critical database solutions, select their platforms, engineer their implementations, and manage their growth to optimize business value.

With decades of experience in large-scale database implementations and in-depth knowledge of database products, we deliver unmatched insight into the issues that impede performance and the technologies that enable success.

WinterCorp

245 FIRST STREET, SUITE 1800
CAMBRIDGE MA 02145
617-695-1800

visit us at www.wintercorp.com

©2010 Winter Corporation, Cambridge, MA. All rights reserved.