

White Paper

IBM DB2 Universal Database and SGI Altix: Identity Resolution and Business Intelligence for an Uncertain World



Table of Contents

1.0 Executive Summary	3
2.0 Extracting Certainty from Mountains of Data	3
3.0 DB2 Entity Analytics on SGI Altix	3
4.0 Overview: IBM DB2 Entity Analytic Solutions	5
5.0 Entity Analytics in Government	6
5.1 Intelligence and Homeland Security	6
5.2 Other Government Uses	6
6.0 Entity Analytics in Business	7
6.1 Identity Resolution for Improved Customer Relations	8
6.2 Relationship Resolution to Identify Target Customers	8
7.0 Anonymous Resolution for Improved Government/Business Cooperation.....	8
8.0 SGI Altix with Intel Itanium 2 Processors	8
9.0 Altix and IBM DB2: A Superior Platform for Business Intelligence Applications	9

1.0 Executive Summary

Both governments and companies are faced with significant challenges that require improved identity resolution. Uncovering potentially hidden relationships between individuals and organizations has also become increasingly important. For example:

- **Homeland Security.** Governments must identify individuals with ill intent in real time.
- **Public Transportation.** Real-time pre-boarding checks of people entering aircraft, long-distance trains, and ships must be accomplished while maintaining personal privacy and complying with federal regulations.
- **Hiring.** Job candidates with criminal histories or unwanted associations (competing companies, conflicts of interest, appearance on watch lists) must be quickly identified and screened out.
- **Fraud.** People and organizations involved in fraudulent activity (health insurance fraud, credit card fraud, etc.) must be identified even when they use aliases to hide their illicit activities.
- **Compliance.** Companies must identify people whose backgrounds put employers in danger of non-compliance with recent statutes (Sarbanes-Oxley, Basel II, etc.)
- **Data Sharing.** Federal, local and state government agencies must share sensitive information to collaborate on joint missions and public safety efforts while simultaneously protecting personal privacy.

This whitepaper describes a unique solution capable of meeting these challenges: IBM DB2 Entity Analytics Solutions (EAS) running on the SGI Altix platform. This solution combines the unique capabilities of IBM EAS software for identity and relationship resolution with the proven performance of SGI Altix running a standard Linux operating environment.

2.0 Extracting Certainty from Mountains of Data

Intelligence and security experts around the globe are faced with a daunting task. They must carefully sort through huge volumes of electronic data, often in real time, to discriminate friend from foe—all while striving to protect the personal privacy of their citizens. And they must do this every single day of the year with no mistakes and without unnecessarily impeding the flow of people, transactions, and materials that is the lifeblood

of the modern world. To meet this challenge, governments are turning to IBM DB2 Entity Analytic Solutions (EAS) running on SGI Altix.

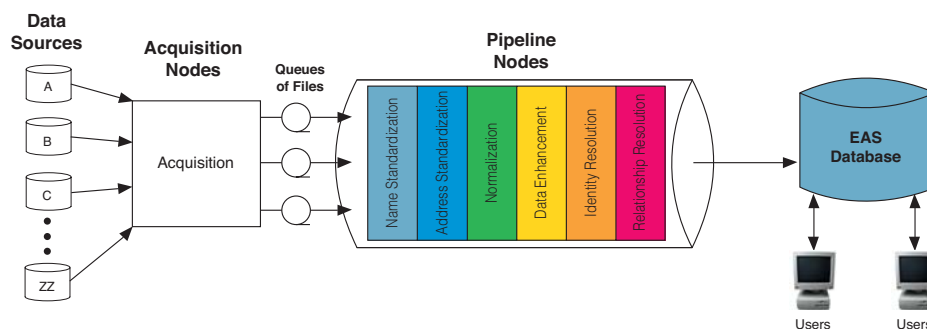
IBM DB2 EAS offers a revolutionary approach to answering the most basic yet most formidable question that governments—and businesses—face: Who am I dealing with? Traditional approaches to answering this question fall short of modern requirements because they neglect important hidden or non-obvious relationships. EAS extends identity profiles to encompass non-obvious relationships, dramatically expanding the capabilities of identity-based applications and offering a level of security and certainty never before possible.

IBM DB2 EAS is a powerful solution that requires a powerful platform. The unique NUMAflex architecture of SGI Altix delivers the balanced compute and I/O capabilities necessary to meet the needs of EAS and other demanding database applications without sacrificing the proven economic advantages of Linux and Intel processors. The combined strength of SGI Altix, IBM DB2 Universal Database (UDB), and IBM Entity Analytic Solutions offers significant performance benefits for both government and commercial customers whose missions depend on effective identity and relationship resolution.

This was demonstrated during a February, 2005 benchmark in which IBM and SGI delivered record performance with DB2 UDB and DB2 EAS running on SGI Altix (see EAS on Altix Technical Whitepaper at <http://www.sgi.com/products/software/eas/>). SGI Altix and IBM DB2 UDB—running EAS or related business intelligence applications—are capable of real time results that can help intelligence, security and business experts reduce uncertainty and increase the speed of decision making to get the job done.

3.0 DB2 Entity Analytics on SGI Altix

IBM DB2 EAS is built as a series of “nodes” or processes and designed to scale massively with a flexible acquisition architecture and pipelining of key EAS processing steps. Multiple data sources can feed into one Acquisition Node, or one very large source system can feed many Acquisition Nodes.



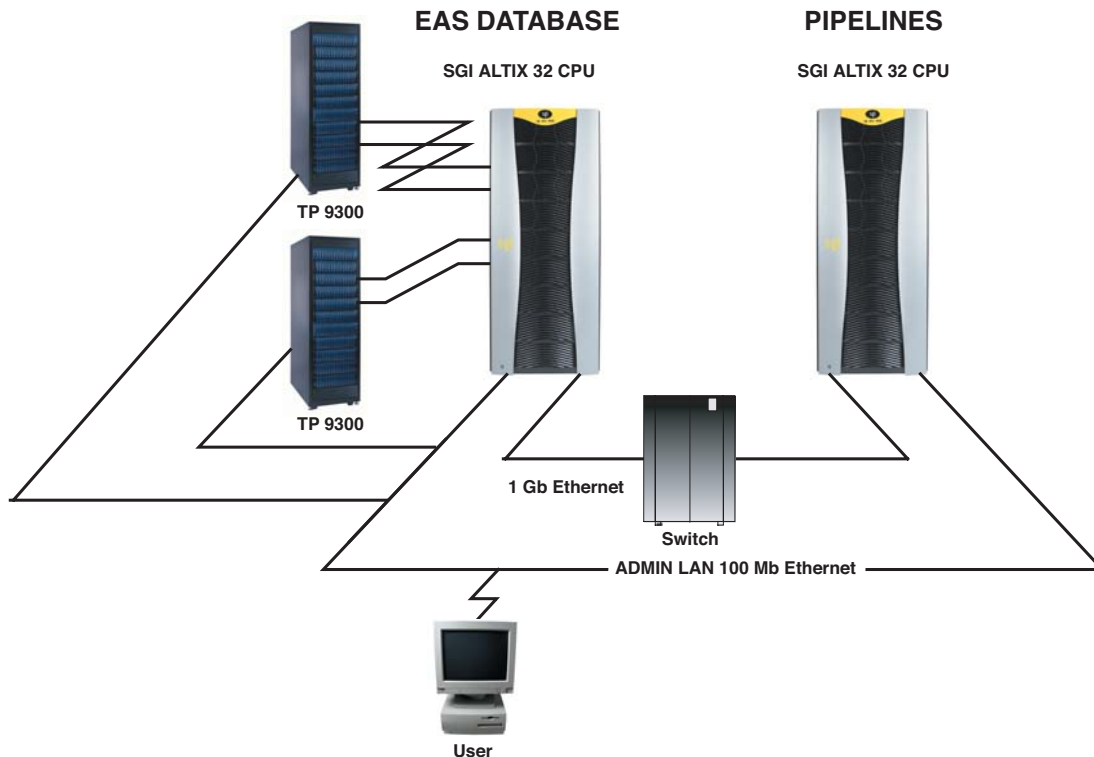
The EAS solution stack (EAS and DB2) can scale to accommodate the huge, rapidly growing data sets and heavy workloads of the most demanding government and business applications. All that is needed is a hardware platform with the scalability and performance capabilities required to keep the software “fed” and running at peak throughput.

SGI Altix is one such hardware platform. The SGI Altix NUMAflex architecture is designed to deliver superior scalability and performance in terms of compute, memory, and I/O without sacrificing the benefits of a standard Linux environment, making it extremely well suited for EAS. SGI Altix dramatically reduces the time required to run the EAS database and pipeline nodes by accommodating extremely large data sets in a single, system-wide, shared-memory space. By combining massive memory capacity (up to 60TB of globally addressable memory) with superior I/O (up to 1TB/day/processor demonstrated ingest with the NUMAflex architecture), SGI Altix can significantly decrease the time needed to assimilate new data and execute complex queries. The combination of Altix and EAS enables governments and businesses to achieve a new level of perpetual analytics to guard against threats and capitalize on opportunities.

A typical EAS deployment might utilize two SGI Altix systems (or a single, partitioned system). One system runs the EAS database with IBM DB2 UDB, while the other executes the DB2 EAS application pipelines. Pairing the highly scalable IBM DB2 database with SGI Altix allows DB2 to take full advantage of the robust capabilities of Intel Itanium 2 processors and SGI's global shared memory, while pipeline processes achieve enhanced throughput as a result of the low latency and high bandwidth of the Altix NUMAlink interconnect.

Global shared memory allows direct and efficient access to all data in the system's memory even for extremely large processor counts. Database workloads take full advantage of this feature to accomplish more work in memory, thereby reducing time-consuming I/O for higher database throughput.

For pipelined workloads, clustered systems pass copies of data in the form of messages over cluster interconnects. SGI Altix significantly streamlines these operations by replacing external cluster connections with NUMAlink connections that provide internal bidirectional bandwidths of 6.4GB/sec per link and latencies of 1 microsecond. By comparison, Infiniband achieves bidirectional bandwidth of only 1.6 GB/sec with latencies of 3.5 microseconds.



4.0 Overview: IBM DB2 Entity Analytic Solutions

IBM DB2 EAS is a unique suite of middleware solutions designed to resolve individual identities to the depth and degree of accuracy required to conclusively determine who is really who. EAS can manage identities in real-time as new information is received, including the impact of both obvious and non-obvious relationships.

The seemingly simple questions of "Who is Who?" and "Who Knows Who?" cut across a wide variety of business problems. When run on highly-scalable hardware such as SGI Altix, DB2 EAS technology provides solutions to these age-old problems with unparalleled speed and accuracy.

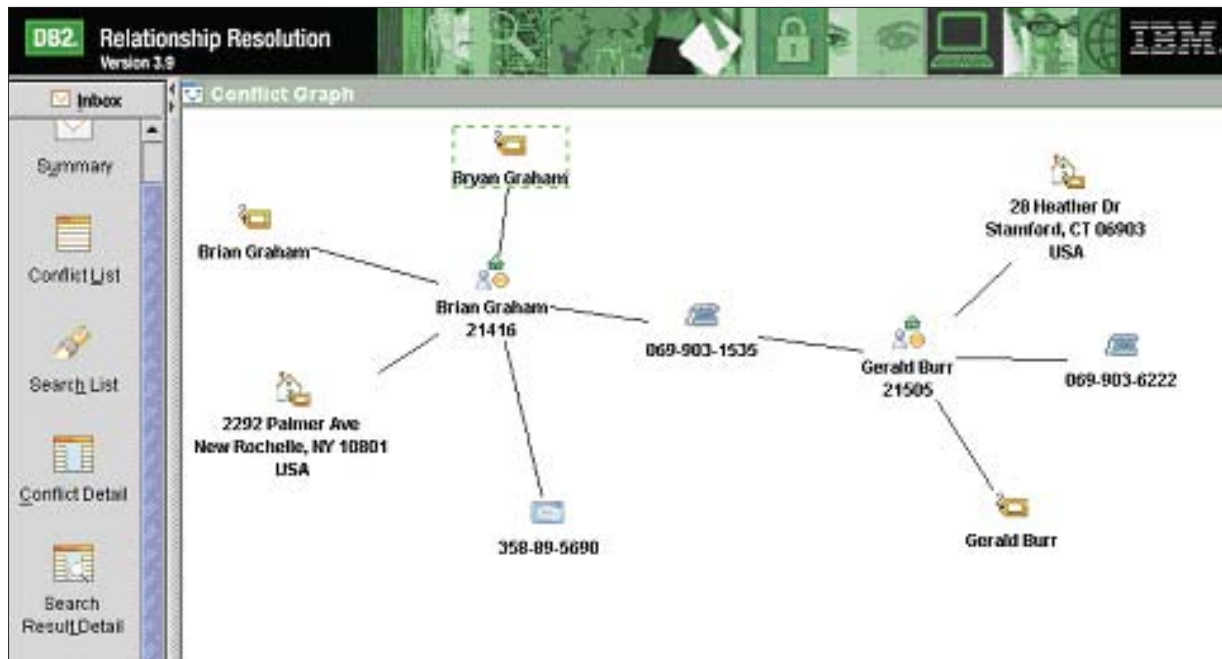
IBM DB2 Identity Resolution is an enabling technology that helps any organization solve problems related to recognizing true identity. Identity Resolution answers the question "Who is Who?" The Identity Resolution process determines whether multiple records, which appear to describe different individuals or organizations, actually correspond to a single resolved identity, despite name variations or deliberate attempts to confuse or misrepresent identity.

DB2 Identity Resolution acquires data from source systems in real time and then conducts a sophisticated five step process:

- **Name standardization:** Identifies root names and variations using proprietary algorithms to compare names and evaluate similarities. Includes 39 international name families.

- **Address verification and correction:** Compares, verifies, and corrects addresses using U.S. and international address databases.
- **Normalization:** Applies data-driven quality rules to addresses, phone numbers, and date of birth, social security numbers, and other significant numbers. Incomplete and invalid entries are rejected and generic values—such as a travel agent phone number given in place of a customer phone number—are tagged.
- **Data enhancement:** Identities are enhanced by adding outside data from an unlimited variety of sources including IBM's extensive partner network of data providers.
- **Identity Resolution:** Compares new information with existing information to determine whether the new information is about a new or existing entity.

IBM DB2 Relationship Resolution extends the value of identity information to include "Who Knows Who?" It begins where most solutions leave off, extending a resolved identity to identify and include non-obvious relationships among individuals and organizations. This unique capability can integrate thousands of internal and external data sources to enable an organization to look beyond the bounds of data directly attributable to an individual to uncover social, professional, value, and criminal networks. DB2 Relationship Resolution seeks out and integrates non-obvious relationships between customers, employees, vendors, and any internal and external source to determine potential value or potential danger.



DB2 Relationship Resolution extends the capabilities of DB2 Identity Resolution by assembling networks of associations, events, and entities using the following data types:

- **Individual attributes:** names, titles, identification numbers, etc.
- **Locations** including a physical address, IP address, or geo code coordinate
- **Facilities** such as factories, airports, hotels, schools, or warehouses
- **Organizations:** cells, gangs, clubs, associations, etc.
- **Documents** such as passports, drivers licenses, e-mail, frequent flyer accounts
- **Money:** cash, wire transfers, money orders
- **Accounts:** bank, checking, credit, savings, services, etc.

DB2 Relationship Resolution allows investigators to identify potential threats in real time without violating privacy rules. DB2 Relationship Resolution cross references data from sources which can reside in multiple formats (DB2, Oracle, SQL Server, flat files, etc.) and other publicly available data sources such as public records and the Office of Foreign Assets Control (OFAC) watch list to determine both obvious and non-obvious relationships. The software is designed to flag suspect relationships and send real-time alerts based upon a user-defined rules engine.

DB2 Anonymous Resolution allows organizations to share data anonymously to maximize the value of information without compromising security or intellectual property rights. Data is transformed using an indecipherable and irreversible format that can be shared to resolve identities and spot relationships across organizational boundaries. Real-time alerts are issued and sent to appropriate parties. Anonymous Resolution enables an organization to compare customer lists with partners, to cross-reference identity records with watch lists, or to determine the amount of customer overlap in a merger and acquisition process, allowing organizations to share only pertinent data with a level of security and privacy never before available.

5.0 Entity Analytics in Government

DB2 EAS software running on SGI Altix offers significant benefits for intelligence and homeland security applications as well as other government uses. The following examples indicate a few of the ways these technologies might be applied to solve important government problems.

5.1 Intelligence and Homeland Security

Identity Resolution

With the increase in security consciousness in recent years, governments at all levels (national, regional/state, local) have

an increasing need to uniquely identify individuals for a variety of purposes such as entry into government buildings. With the many recent, highly publicized cases in which innocent people have been detained because they had the same name as someone on a government watch list, it is clear that better technology is needed. DB2 EAS Identity Resolution can help in these situations by rapidly distinguishing between individuals with similar or identical names.

Relationship Resolution

DB2 Relationship Resolution can be employed in law enforcement and intelligence to examine anomalies and inconsistencies as well as uncover networks of relationships hidden in multiple databases. For terrorists and others wishing to enter a country under false pretenses, one of the first steps is to obtain a valid driver's license. This identifying document then gives them the ability to rapidly develop a complete cover. Consider a foreign-born person applying for a license in Kansas. In the standard process, the individual's identifying information might be compared to a variety of watch lists such as OFAC Watch and other intelligence lists. If no matches are found, the individual is granted a license. With Relationship Resolution, the individual is mapped to a previous address in Belgium. That address was shared by an individual who once shared a phone number with an individual who appears on a government watch list, triggering further investigation and rejection of the application.

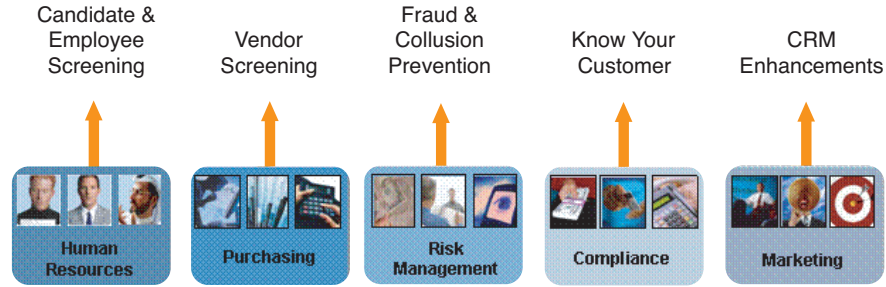
5.2 Other Government Uses

Identity Resolution for Entitlement Programs

In a more mundane but no less important area, governments can apply DB2 Identity Resolution and SGI Altix to improve efficiency and reduce fraud in entitlement programs. Programs such as social security, government-sponsored healthcare and welfare are a large and growing part of budgets in most developed countries. Fraud and double payouts waste billions of dollars annually. DB2 Identity Resolution running on SGI Altix accurately resolves identity information across multiple data sets, despite ambiguities and obfuscations, to increase the efficiency of entitlement and similar programs.

Relationship Resolution for Government Hiring

Media reports of government officials with serious conflicts of interest and improper relationships with private industry are on the rise. Government agencies across the board are trying to do a better job of investigating candidates prior to hiring, but this is often a laborious, manual process. DB2 EAS Relationship Resolution on SGI Altix has the ability to substantially streamline these background investigations.



For example, consider the U.S. Food and Drug Administration (FDA). Before hiring, the FDA needs to ensure that candidates don't have any improper connections to major pharmaceutical companies. Using Relationship Resolution, the agency could quickly uncover not only direct ties to drug-related companies, but indirect connections that might be cause for concern through friends and family.

6.0 Entity Analytics in Business

IBM DB2 Entity Analytic Solutions on SGI Altix also provide significant advantage to a wide variety of businesses. The DB2 EAS toolset, running on a scalable and high performance platform, has broad applications for any organization that requires:

- **Fraud Detection.** Every day brings new reports of identity theft, consumer data theft, and other forms of fraud. Ability to rapidly identify individuals, companies, and associated networks of relationships can dramatically reduce a company's exposure to embarrassing and potentially costly white collar crimes.
- **Insider Threat Detection.** In today's fluid job market, threats from company insiders can be at least as significant as those from outsiders. DB2 EAS on SGI Altix allows businesses to quickly identify individuals who may pose a threat and to rapidly identify suspicious relationships such as inappropriate employee/customer and employee/vendor connections that might signify conflicts of interest or criminal intent.
- **Customer Relationship Management.** Global Enterprises frequently have multiple business units doing business with the same customers, often with little or no awareness of different activities. Duplicated efforts, over-communication, and inefficiency are the result. DB2 EAS on SGI Altix allows businesses to consolidate information from disparate databases to create a single, rationalized customer list to streamline customer communication and eliminate redundancies.
- **Regulatory Compliance.** Recently tightened compliance guidelines such as the Sarbanes Oxley Act, Gramm-Leach-Bliley Act, and Basel II require companies to substantially improve knowledge of customers and vendors and tighten a

variety of monitoring and reporting activities. DB2 EAS on SGI Altix can help companies meet these requirements with greater confidence.

- **Enhanced Due Diligence.** Company mergers and acquisitions require careful screening of customer lists, vendors, employees, etc. DB2 Anonymous Resolution on SGI Altix can enable companies investigating a merger to identify red flags without prematurely releasing valuable corporate data.

DB2 EAS solutions running on SGI Altix are attracting interest in a variety of important markets:

- **Financial Services.** Due to increased regulation in a variety of areas as well as other recent developments, financial services organizations are broadly applying DB2 EAS capabilities to improve business processes and compliance. Identity theft and credit card fraud are among the major concerns that can be combated using DB2 Identity Resolution, while DB2 Relationship Resolution can help dramatically reduce money laundering linked to criminal and terrorist activities.
- **Insurance.** Insurance companies can use DB2 EAS to identify and reduce insurance fraud and improve customer management.
- **Healthcare.** Private healthcare faces the same problems as government entitlement programs. Improved management and analysis of customer information with DB2 EAS on SGI Altix can help healthcare companies identify fraud and improve payout processes.
- **Retail.** Large retailers face a variety of threats including fraud and collusion from insiders and vendors. DB2 EAS on SGI Altix can help retailers better screen employees and vendors, identify conflicts of interest and possible collusion and also improve monitoring of customer activities. For example, any relationships between slip and fall victims and employee witnesses can be quickly uncovered.
- **Hospitality.** The hospitality industry faces a number of unique challenges that can result from globally distributed properties and highly mobile customers. DB2 EAS on SGI Altix provides

the tools to allow hotels and casinos to immediately recognize customers anywhere in the world to increase service while reducing potential threats. The hospitality industry also benefits significantly from improved screening of employees and vendors.

6.1 Identity Resolution for Improved Customer Relations

Companies of all sizes have difficulty keeping track of customers, as anyone who routinely receives multiple copies of the same catalog (sent to slightly different names or addresses) can attest. DB2 EAS on SGI Altix can enhance a company's knowledge of its customers, quickly clean up customer lists and consolidate multiple databases to improve the productivity of customer contacts and increase efficiency.

For example, consider a large financial institution with multiple divisions offering many different products: checking and savings accounts; CDs; credit cards; IRAs; personal and commercial loans, etc. Each division maintains a separate customer list. The same individual may show up differently in different lists, and when that individual deals with a new division of the company they start from scratch. For instance, a loan application may require the person to complete lengthy questionnaires to provide information that the company already stores elsewhere.

DB2 Identity Resolution can resolve identities and provide the institution with a clean and comprehensive client database. As a result, redundant customer mailings can be reduced or eliminated, and any division can access all of a client's information as needed using standard business tools. The level of customer service naturally improves, because any employee looking at a customer's records can see and understand the complete scope of the customer's history with the company.

6.2 Relationship Resolution to Identify Target Customers

Most of us receive mountains of junk mail, e-mail, and other solicitations that are of absolutely no interest to us. This occurs because companies have difficulty identifying appropriate target customers. EAS Relationship Resolution can help a company extend its web of prospects without resorting to mass mailings or other low-return marketing strategies.

For example, a nationwide retailer with millions of customers and billions of sales transactions might begin by rationalizing its customer list using EAS Identity Resolution. Starting with that clean list, the company can then apply EAS Relationship Resolution to identify friends and family members that share the same interests as existing customers to undertake a targeted direct mailing with a much higher hit rate than would be achievable using traditional techniques.

7.0 Anonymous Resolution for Improved Government/Business Cooperation

Privacy concerns are increasingly hampering government-to-business and business-to-business relationships. Individuals don't want their governments or the companies they do business with sharing their personal information widely. This need for privacy obviously reduces cooperation between organizations and interferes with legitimate security concerns.

EAS Anonymous Resolution improves cooperation between organizations of all sorts while enhancing privacy by enabling information to be shared in an encrypted form to enable discovery without disclosure. For example, a government might want to watch for terrorists at hotels in major metropolitan areas. Obviously, just handing out a watch list to every hotel is not sufficient (and not desirable since the information contained on watch lists is considered sensitive), but hotels are understandably reluctant to share their guest lists with the government due to privacy concerns.

Anonymous Resolution can be used to convert hotel records into undecipherable text strings using a one-way hash function. The government can then compare this information to similarly encoded watch lists to find matches without the need to obtain the guest list in a readable form. When a match is detected, the government can simply request the record that corresponds to the match, not the entire guest list. As a result, security can be improved without unnecessarily compromising privacy.

8.0 SGI Altix with Intel Itanium 2 Processors

SGI offers a number of Altix platform options to accommodate any size EAS deployment. Each Altix 330 or Altix 1330 cluster compute node supports from 1 to 16 processors and up to 128GB of memory in a single system. The SGI Altix 1330 is a fully integrated cluster solution that can be deployed to tackle demanding workloads such as EAS.

For the highest level of performance, the Altix 4700 system can accommodate up to 512 processors under one instance of Linux and as much as 60TB of globally addressable memory. The SGI NUMALink 4 interconnect can be used to connect Altix 4700 nodes in a supercluster configuration that supports thousands of processors and hundreds of terabytes of memory—powerful enough to address the most challenging compute problems.

A key element of the Altix architecture is the 64-bit Intel Itanium 2 processor, which delivers extensive computational resources, including large register sets, and multiple, parallel execution units for integer, multimedia, floating point, branch, and load

and store operations. By choosing the Itanium 2 processor family for Altix, SGI takes advantage of the outstanding performance of Itanium today, and protects customer investments through the ability to upgrade over time according to the Itanium 2 roadmap.

Itanium 2 processor-based platforms deliver exceptional scalability and availability for IBM DB2 solutions, at much lower costs than traditional, RISC-based platforms. Based on Explicitly Parallel Instruction Computing (EPIC), Intel Itanium architecture supports highly parallel processing, large memory addressability, and innovative, compiler-based optimization that greatly improve performance for database operations. The parallel architecture includes a large number of execution registers that enable efficient, simultaneous processing for up to six instructions. Since the compiler optimizes the software for parallel throughput, the processor can focus all its resources on fast execution, making optimal use of advanced features such as the large on-die cache; predication, which reduces branch delays; and speculation, which preloads essential data for faster processing.

“The proven performance of IBM DB2 EAS running on SGI Altix and Intel Itanium 2 processors provides governments and businesses with a new toolset to identify and protect against a broad range of threats and to capitalize on a wide range of opportunities,” says Jason Waxman, Director, Intel Server Platforms Marketing. “New and incoming data can now be compared in real time against a huge variety of existing data sources to provide previously unprecedented perpetual analytics capabilities.”

9.0 Altix and IBM DB2: A Superior Platform for Business Intelligence Applications

The level of performance achievable by executing the highly scalable IBM DB2 Universal Database on SGI Altix can also be applied to a wide range of Business Intelligence problems beyond EAS. “The IBM DB2 database running on Linux offers great economy, manageability, performance, availability and scalability,” says W. Thomas Stanley, National Director, Intelligence & Department of Homeland Security at SGI. “These characteristics are a great match for the proven capabilities of SGI Altix. Together, the combination allows governments and businesses to deploy powerful, highly scalable databases for EAS, data warehousing, or other business intelligence needs.”

IBM DB2 UDB and advanced DB2 solutions such as DB2 EAS and DB2 Intelligent Miner on SGI Altix create a powerful platform for real time business intelligence for increased certainty, faster and better decision making, reduced risk, and greater operational efficiency.



Corporate Office
1500 Crittenden Lane
Mountain View, CA 94043
(650) 960-1980
www.sgi.com

North America +1 800.800.7441
Latin America +55 11.5509.1455
Europe +44 118.912.7500
Japan +81 3.5488.1811
Asia Pacific +1 650.933.3000

©2005 Silicon Graphics, Inc. All rights reserved. ©2005 Intel Corporation. All rights reserved. Silicon Graphics, SGI, Altix, and the SGI logo are registered trademarks of Silicon Graphics, Inc., in the U.S. and/or other countries worldwide. All other trademarks mentioned herein are the property of their respective owners. Other names and brands may be claimed as the property of others. Intel, the Intel logo, and Itanium are trademarks or registered trademarks of Intel Corporation or its subsidiaries in the United States and other countries. Performance tests and ratings are measured using specific computer systems and/or components and reflect the business goals, metrics, and design of the enterprise environment and client systems under test. Any difference in business goals, software design, infrastructure, or configuration may affect actual performance.