



Uncover key entity insights from big data with  
advanced linking technology from LexisNexis®.

LexisNexis® Data Analytics Supercomputer



Risk Solutions  
Government

Government agencies look to LexisNexis® for answers to essentially the same challenge: the need to ingest, transform, process and analyze massive amounts of data in order to get the information needed to make informed decisions. All agencies also have essentially the same problem: extracting that information can be expensive due to the power, space and cooling considerations of high performance computing systems.

## Immediately analyze massive amounts of data

### Your challenge

Conventional high performance computing platforms and database technologies can't handle the variety and volume of data you need to analyze to uncover key connections and relationships. Tasked with ingesting, linking and analyzing this amount of complex data, conventional technologies are brought to a grinding halt. They're not scalable and they're not fast enough. They don't allow you to see the whole picture—the entirety of the data sets you need to analyze.

### Unparalleled capacity, linking technology and analysis

The LexisNexis® Data Analytics Supercomputer (DAS) platform is a High Performance Computing Cluster (HPCC) designed to handle the diverse data sources, variety of data types and amount of data you need to analyze in support of your mission. The DAS supports up to petabytes of data on a massively parallel system. Ingest, sort, link and query huge data sets—link in hours, and conduct complex, multifaceted queries in seconds.

Benchmark tests show that the DAS sorts one terabyte of data in half the time with 44 percent of the hardware and 10 percent of the code when compared against Hadoop, the 2008 Terasort World Champion. Better performance on a smaller hardware footprint means lower hardware costs. The DAS also has lower operating costs and integrates easily with existing systems via a service-oriented architecture (SOA) interface.

It's fast, it's accurate and it's scalable. And, it has a lower total cost of ownership than other high performance computing platforms.

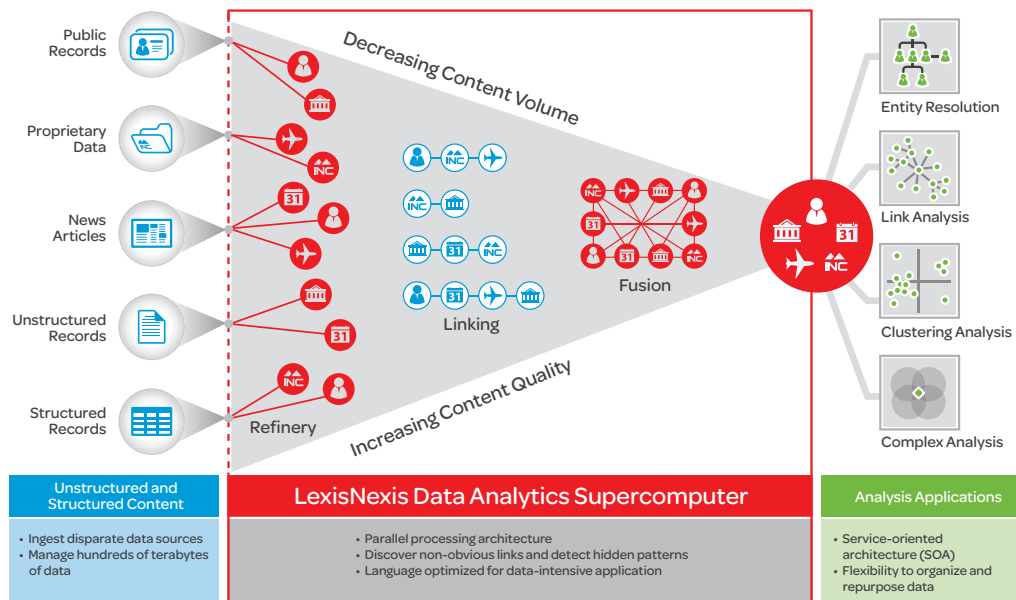
## Pose Complex Questions, Uncover Key Relationships

The LexisNexis Data Analytics Supercomputer enables government agencies to perform complex data sorting and obtain specific results. For example, find all male persons aged 35 to 45 who lived in an apartment in Seattle from 2005 to 2007, owned a car with the letters "O" and "H" in any position on the license plate number, and may go by John or Jim or Joe or Jose Smith.

Or, search hundreds of terabytes of netflow data and find all communications between any two computers that occur at regular intervals with relatively small standard deviation.

### Architected for speed and scalability

DAS architects focused on one key issue: designing a system that instantaneously delivers answers regardless of the amount of data being analyzed. By focusing on how best to ingest and analyze terabytes of data in hours instead of days, DAS architects have completely reengineered the traditional approach to data analysis solutions. Instead of requiring that data be transferred to a middle tier where algorithms for analyzing data are typically stored, LexisNexis engineers realized it would be infinitely more scalable to move the algorithms to the data. The result was a world-class HPCC platform that leverages a distributed, parallel processing platform architecture developed to enable the execution of the algorithm at the disk where the actual data is stored. Our engineers also developed an extensible, declarative, non-procedural language (Enterprise Control Language, or ECL) to handle the complexities of large-scale data management and query processing. Leveraging the speed from its parallel processing architecture and the power of its declarative programming language, the DAS refines, links and fuses large amounts of data from disparate sources. The result: complex analysis and queries at speeds many times faster and programming 10 times more efficient than anything else available in the market today.



## Uncover connections and relationships

### Massive data capacity

Ingest and analyze the entirety of a data set, from hundreds of terabytes to petabytes. The DAS enables a “one to many” search capacity that allows full table scans instead of only analyzing a portion of the full data set.

### Powerful ingest and refresh

The DAS provides an exclusive Extract, Transform, Load and Link (ETL<L>) feature—a powerful linking capability that links and clusters all associated information around a unique entity.

### Delivery

The Data Delivery Engine provides fast access to data and services via Simple Object Access Protocol (SOAP)-based Web Services and is designed to provide access to many users concurrently.

### Fast and easy implementation

The DAS platform is pretested and preconfigured for your system prior to delivery, making its implementation into your workflow fast and seamless—as quickly as 30-60 days.

### Adaptability

The DAS allows adaptability to new questions and data, providing the flexibility to adapt to changing mission requirements and information sources at the time of ingestion.

## Just How Fast Is It?

LexisNexis® Special Services Inc. and Sandia National Laboratories performed competitive tests using the PageRank algorithm (PageRank is a link analysis algorithm that assigns a numerical weighting to each element of a hyperlinked set of documents, such as the World Wide Web, with the purpose of “measuring” its relative importance within the set). In competitive PageRank performance tests conducted by Sandia, results show that a 400-node LexisNexis Data Analytics Supercomputer performed 10 times faster than the next fastest system. The speed and accuracy is derived from the DAS High Performance Computing Cluster (HPCC) technology that enables it to process, analyze and find links and associations in high volumes of complex data. In addition, the DAS scales linearly from 10s to 1000s of nodes to handle petabytes of information, supporting millions of transactions per day.

### Accuracy

Enterprise Control Language (ECL) is a flexible, declarative programming language that enables users to express complex queries with fewer lines of code than other conventional programming languages. Clusters and best-in-class LexisNexis® advanced linking technology algorithms provide for disambiguation and reduction of false positives.

## Case Study: From 26 Days to 8 Minutes

### The Challenge

To reduce losses on existing loans, a leading national credit bureau used a complex algorithm to predict the likelihood that a customer would default within the next 30 days. However, its traditional mainframe system took 26 days to run the algorithm on its entire population, leaving only a few days to act. The credit bureau's alternative was to only process a five percent sample.

### The Solution

Using the DAS, the customer was able to process the algorithm for the entire population in just eight minutes.

### The Results

An increase in effectiveness and coverage 20 times of the prediction model resulted in significantly greater savings in time and money for the credit bureau and its customers.

## Flexibility

Linear scalability allows nodes to be added and capacity to be scaled as data needs change. Nodes can be repurposed for Data Refinery or Data Delivery. SOA-compliant architecture facilitates seamless integration with third-party applications. What's more, the DAS does not require new sources to conform to specific data schemas, enabling multiple data formats to be quickly loaded and indexed for analysis.

## Lower total cost of ownership

The DAS utilizes commercial off-the-shelf (COTS) hardware, which significantly lowers its total cost of ownership. The scalability of the DAS allows you to expand the system as data volumes dictate.

## LexisNexis Data Analytics Supercomputer (DAS)

### System architecture

- Parallel processing architecture
- Optimized network system configuration

### DAS software suite components

- Data refinery engine
- Rapid data delivery engine
- Service-Oriented Architecture (SOA): Web Services access via Simple Object Access Protocol (SOAP)

### DAS client tools

- Query builder, attribute migration tool, DAS command line, (DAS command line tool set)

### Hardware components

- Rack-mounted "off-the-shelf" servers and network switches delivered with DAS software installed, preconfigured and pretested
- Number of nodes (processing and storage units) configurable based on data and computational load

### Operating system

- Linux operating system

## For more information:

Call 800.291.3670 or visit [lexisnexis.com/government](http://lexisnexis.com/government)

### About LexisNexis Risk Solutions

LexisNexis Risk Solutions ([www.lexisnexis.com/risk](http://www.lexisnexis.com/risk)) is a leader in providing essential information that helps customers across all industries and government predict, assess and manage risk. Combining cutting-edge technology, unique data and advanced scoring analytics, we provide products and services that address evolving client needs in the risk sector while upholding the highest standards of security and privacy. LexisNexis Risk Solutions is part of Reed Elsevier, a leading publisher and information provider that serves customers in more than 100 countries with more than 30,000 employees worldwide.

Our government solutions assist law enforcement and government agencies with deriving insight from complex data sets, improving operational efficiencies, making timely and informed decisions to enhance investigations, increasing program integrity, and discovering and recovering revenue.

