HPCC Systems:
Powerful Open Source Big Data Analytics Platform

Born from LexisNexis® Risk Solutions deep data analysis history, HPCC Systems® helps businesses of all sizes find the answers they need by making data easier to process, analyze, and understand.

HPCC Systems platform is a set of easy-to-use software components enabling developers and data scientists to process and analyze data at any scale. With a strong commitment to the open source community, the HPCC Systems platform is available free of licensing and service costs.

HPCC Systems is cost-effective, comprehensive, fast, powerful, and scalable.

Ultimately, it makes managing big data easier.

Open source. Easy to use. Proven.

The HPCC Systems stack comprises of:

ETL - Extract, Transform, and Load your data using a powerful programming language (ECL) specifically developed to work with data.

Data Management Tools - Data profiling, Data Cleansing, Snapshot Data Updates and consolidation, Job Scheduling and automation are some of the key features.
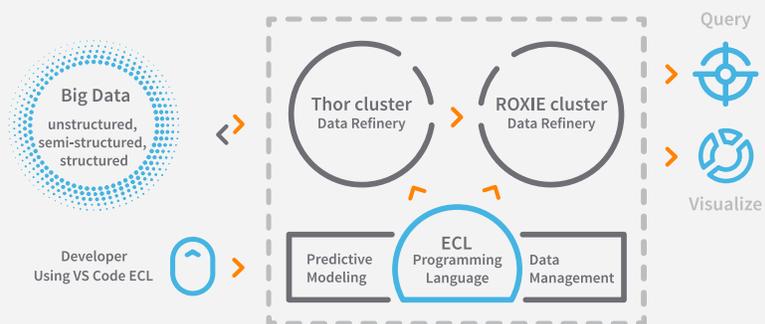
Query and Search - An indexed based search engine to perform real-time queries. SOAP, XML, REST, and SQL are all supported interfaces.

Predictive Modeling Tools - In place (supporting distributed linear algebra) predictive modeling functionality to perform Linear Regression, Logistic Regression, Decision Trees, and Random Forests.

HPCC Systems:

Big Data
unstructured, semi-structured, structured

Thor cluster
Data Refinery

ROXIE cluster
Data Refinery

Query

Visualize

Developer
Using VS Code ECL

Predictive Modeling

ECL Programming Language

Data Management

HPPC Systems. Its easier:

- Easier to update
- Easier to learn
- Easier to program
- Easier to integrate data
- Easier to manage clusters

End to end big data in a massively scalable super computing platform

Features:

- Standard hardware, operating system, and protocols
- High redundancy and availability
- Practical tools and extensions
- Efficient programming with a declarative, modular, extensible language
- End-to-end configuration
- Optimized distributed file systems (DFS)
- Massive scalability and performance

Thor - Big Data ETL Made Easy

Thor executes data manipulation scripts to transform and analyze raw data.

Known as "Thor" after the hammer-wielding god of thunder, this cluster is designed to execute big data workflows including extraction, loading, cleansing, transformations, linking, and indexing.

Thor uses a master-slave topology in which slaves provide localized data storage and processing power, while the master monitors and coordinates the activities of the slave nodes and communicates job status information.

Middleware components provide name services and other services in support of the distributed job execution environment.

Manage data more quickly and efficiently:

Analytics

Transform

Thor Cluster
Data Refinery Engine

Profiling

Cleaning

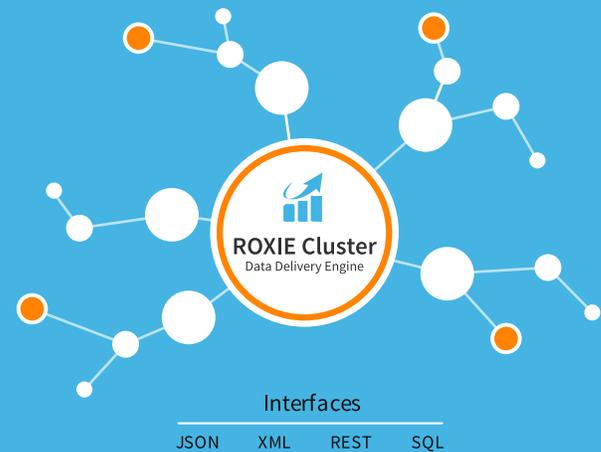# ROXIE – High Performance Real-Time Queries

ROXIE is an index based search engine to perform real-time queries.

The ROXIE rapid data delivery cluster provides separate high-performance online query delivery for big data. ROXIE (Rapid Online XML Inquiry Engine) utilizes highly optimized distributed B-tree indexed data structures conceived for high concurrent use.

Each ROXIE node runs a Server process and an Agent process. The Server process handles incoming query requests from users, allocates the processing of the queries to the appropriate Agents across the ROXIE cluster, collates the results, and returns the payload to the client.

Queries may include joins and other complex transformations, and payloads can contain structured or unstructured data.

Perform real-time queries with highly concurrent delivery:

ROXIE Cluster
Data Delivery Engine

Interfaces

JSON   XML   REST   SQL

HPCC SYSTEMS®

About HPCC Systems®
HPCC Systems® (www.hpccsystems.com) from LexisNexis® Risk Solutions is a proven, open source solution for Big Data insights that can be implemented by businesses of all sizes. With HPCC Systems, developers can design applications with Big Data at their core, enabling businesses to better analyze and understand data at scale, improving business time to results and decisions. HPCC Systems offers a consistent data-centric programming language, two processing platforms and a single, complete end-to-end architecture for efficient processing. For more information, please visit www.hpccsystems.com

LexisNexis®
RISK SOLUTIONS

About LexisNexis® Risk Solutions
At LexisNexis Risk Solutions, we believe in the power of data and advanced analytics for better risk management. With over 40 years of expertise, we are the trusted data analytics provider for organizations seeking actionable insights to manage risks and improve results while upholding the highest standards for security and privacy. Headquartered in metro Atlanta USA, LexisNexis Risk Solutions serves customers in more than 100 countries and is part of RELX Group, a global provider of information and analytics for professional and business customers across industries. For more information, please visit www.risk.lexisnexis.com

For more information, call 877.316.9669 or visit www.hpccsystems.com