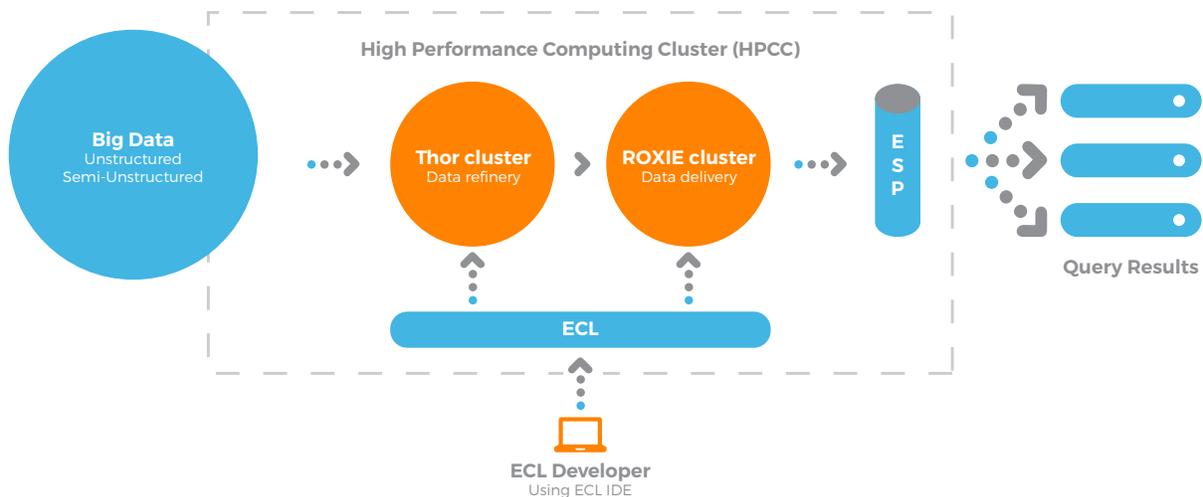# HPCC SYSTEMS ®

Open Source and Enterprise-Proven Big Data Analytics Platform

www.hpccsystems.com

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

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

HPCC Systems incorporates a software architecture implemented on commodity shared-nothing computing clusters to provide high-performance, data-parallel processing and delivery for applications utilizing Big Data. The HPCC Systems platform includes system configurations to support both parallel batch data processing (Thor) and high-performance data delivery applications using indexed data files (ROXIE). It also includes a high level and implicitly parallel data-centric declarative programming language for parallel data processing, called Enterprise Control Language (ECL).

High Performance Computing Cluster (HPCC)

Big Data
Unstructured
Semi-Unstructured

Thor cluster
Data refinery

ROXIE cluster
Data delivery

ESP

Query Results

ECL

ECL Developer
Using ECL IDE

HPCC SYSTEMS COMPONENTS:

THOR - Data Refinery Cluster designed to execute big data workflows including extraction, loading, cleansing, transformations, linking and indexing.

ROXIE - Rapid Data Delivery Cluster provides separate high-performance online query delivery for Big Data delivery. ROXIE utilizes highly optimized distributed B-tree indexed data structures and has been conceived for high concurrent use. A typical 10 node cluster can process thousands of concurrent requests and deliver them in fractions of a second.

ECL - Enterprise Control Language is declarative, data-centric, distributed processing language for Big Data.

ESP - Enterprise Services Platform provides an easy to use interface to access ECL queries using XML, HTTP, SOAP (Simple Object Access Protocol) and REST (Representational State Transfer).

Identify trends and anomalies with transformative graphs

Many complex data problems require a series of advanced functions to solve them. With HPCC Systems technology, complex data challenges can be represented naturally with a transformative data graph.

- The nodes of the data graph can be processed in parallel as distinct data flows.

- The ECL IDE turns code into graphs that facilitate the understanding and processing of large scale, complex data analytics.

- Each section of the graph includes information such as function, records processed or skew.

- Each node can be drilled into for specific details.
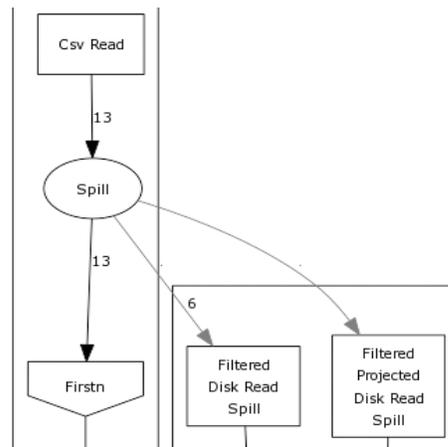
End to end big data in a massively scalable super computing platform

The HPCC Systems technology was designed by data scientists. The programming language, ECL, is a declarative, collaborative and extensible, high-level language that allows the programmer to describe the desired outcome instead of programming tedious and ambiguous scripting.

```
log_out_init := project(log_init,
                        transform(layout_logout,
                        self := left,
                        self := []));

// Create error log
outerrorfile := join(log_seq,
                     log_out_init,
                     left.linenum = right.linenum,
                     transform(recordof(log_seq),
                        self := left),
                     left only,
                     hash);
```

ENTERPRISE CONTROL LANGUAGE (ECL) is:

DECLARATIVE: describes the what, not the how.

FOCUSED: Higher level code means a reduced programming burden and shortens time to delivery.

EXTENSIBLE: As new attributes are defined, they become primitives that other programmers can use.

IMPLICITLY PARALLEL: Parallelism is built into the underlying platform. The programmer does not need to manage it.

MAINTAINABLE: Designed for long-term, large scale, enterprise use.

COMPLETE: Provides for a complete programming paradigm.

HOMOGENEOUS: One language to express data algorithms across the entire HPCC Systems platform, including a data ETL and high speed delivery.

HPCC SYSTEMS®

For more information call 877.316.9669 or visit hpccsystems.com

HPCC Systems Community Edition is a free version supported by an active community of developers and enthusiasts via the online discussion forums. HPCC Systems Enterprise Edition is a subscription-based offering for enterprises. It includes software, services and support to deliver the highest levels of reliability, security, and uptime proven over the years in mission-critical industrial settings.

LexisNexis®
RISK SOLUTIONS