

AGENDA: 2019 HPCC Systems® Community Day

Challenge Yourself –
Challenge the Status Quo



Training Workshop and Community Day October 15-16, 2019

LexisNexis® Risk Solutions will host the 6th annual HPCC Systems Summit Community Day on Wednesday October 16, 2019 in Atlanta, Georgia at the Georgia Tech Hotel and Conference Center. We are offering a new pre-event workshop, “**Rocking Big Music Data using ECL Essentials,**” to our external Community Day participants! The workshop will take place the day before on Tuesday, October 15 at the same venue.

What will you learn?

This class is for attendees who want to understand the HPCC Systems platform and learn ECL to build powerful data queries. The four hour workshop will take students through the data extraction, transformation, and delivery cycle using HPCC Systems and ECL. Code examples and hands-on lessons will be included. A public dataset based on the Million Song Dataset, and other auxiliary data will be used.

Course Date/Time: October 15, 8:30AM - 12:30 PM, Lunch provided.

Course Price: \$50 (Going to Community Day on October 16 as well? If yes, then select the **2-DAY pass** and enjoy a discount off the \$50 Community Day registration fee. Attend both days for \$75!)

Class Prerequisites - REQUIRED: Pre-installation of ECL IDE (or VS-Code with the extension for Mac users) and connection to AWS cluster or HPCC Systems VM.

Instructor: Bob Foreman, Senior Software Engineer, LexisNexis Risk Solutions

Topics include:

- Introduction to HPCC Systems Architecture
- Data Extraction and Spraying to the Cluster
- Configuring the ECL IDE or VS-Code environment
- Introduction to the ECL Watch
- ECL Essentials and Syntax
- Defining RECORD and DATASET
- Repository Code Organization
- The Open Data Model of HPCC Systems
- Four Fundamental Data Definitions and Filtering
- Data Aggregation Functions
- TABLE Function and Cross-Tabulation Reports
- Data Duplication Processing: DEDUP and ROLLUP
- Data Cleansing using PROJECT
- Using the Standard Function Library
- Creating Custom ECL Functions
- Lookup Tables and JOIN
- Indexing a Dataset for ROXIE
- Writing and Publishing a ROXIE Query

Course includes a printed manual and thumb drive (including data, code and documentation).

Event Details:

hpccsystems.com/hpccsummit2019
#HPCCSUMMIT



2019 HPCC Systems® Community Day

Challenge Yourself – Challenge the Status Quo

Wednesday, October 16, 2019

Time	Topic
7:00AM – 8:30AM	Registration and Breakfast
8:30AM – 9:00AM	Welcome & Keynote
9:00AM – 10:15AM	Track 1: HPCC Systems in Industry Featuring talks from a research lab on a bio-tech industry use case, and our own success stories from RELX IADP and LexisNexis Risk Solutions.
10:15AM – 10:30AM	Break - Exhibits
10:30AM - 12:00PM	Track 2: HPCC Systems in Academia Presentations include research findings and project outcomes from our HPCC Systems Academic Community featuring Clemson University, NC State University, Florida Atlantic University and NSU University School.
12:00PM - 12:15PM	Community Awards Ceremony
12:15PM – 1:00PM	Lunch
1:00PM - 1:30PM	Interactive Expo - NEW THIS YEAR! Robotics showcase featuring American Heritage School and NSU University School, Hands-on demos, and 1:1 Q&A with our HPCC Systems experts
1:30PM - 3:00PM	Track 3: HPCC Systems Breakouts - see descriptions on pages 3-4
3:00PM - 3:15PM	Break - Exhibits
3:15PM - 4:40PM	Track 4: HPCC Systems Roadmap Tech Talks Our platform team will share an update on the latest features included on the roadmap
4:40PM - 4:55PM	Community Website: Virtual Ribbon Cutting
4:55PM - 5:00PM	Closing & Adjourn

Event Details:

hpccsystems.com/hpccsummit2019

#HPCCSUMMIT



2019 HPCC Systems® Community Day

Challenge Yourself – Challenge the Status Quo

Wednesday, October 16, 2019
Breakouts

Time	Descriptions
1:30PM - 3:00PM	Track 3: HPCC Systems Breakouts Sessions include a deeper dive into specific technical topics and components of the HPCC Systems platform.
1:30PM - 2:10PM	Rotation 1
	Theme 1: System Enhancements
	Workunit Analysis Tool: The Workunit Analyser examines the entire workunit to produce advice that both novices and experienced ECL developers should find useful. The Workunit Analyser is a post-execution analyser that identifies potential issues and assists users in writing better ECL. Speaker: Shamsar Ahmed, LexisNexis Risk Solutions
	Leveraging the Spark-HPCC Ecosystem: Join us for an introductory walk-through of using the Spark-HPCC Systems ecosystem to analyze your HPCC Systems data using a collaborative Apache Zeppelin notebook environment. Speaker: James McMullan, LexisNexis Risk Solutions
	Theme 2: Usability Improvements
	Dapper Tool - A Bundle to Make your ECL Neater: Have you ever written a long project for a simple column rename and thought, this should be easier? What about nicely named output statements? Yeah they bother me, too. Oh, and DEDUP(SORT(DISTINCT()))? There is a better way! Learn how Dapper can help! Speaker: Rob Mansfield, Proagrica
	DataPatterns - Profiling in ECL Watch: DataPatterns.Profile() has been evolving since the last time you may have seen it. It does more. It looks better. It has been integrated into the ECL standard library and into ECL Watch. Learn what this data profiler can do for you and how its built-in visualization easily summarizes the results. Speaker: Dan Camper, LexisNexis Risk Solutions
2:10PM - 2:20PM	Room Change Break

Event Details:

hpccsystems.com/hpccsummit2019

#HPCCSUMMIT



Wednesday, October 16, 2019
Breakouts

Time	Descriptions
1:30PM - 3:00PM	Track 3: HPCC Systems Breakouts continued
2:20PM - 3:00PM	Rotation 2

Theme 1: Novel Applications

Leveraging Intra-Node Parallelization in HPCC Systems: Our research focuses on the parallel implementation of the set similarity join (SSJ) operator. This operator finds all pairs of records which have a similarity above a defined threshold using a similarity measure such as Jaccard. Our goal is a robust approach for executing SSJ that does not overutilize memory and exploits CPU parallelization as much as possible. This approach requires data sharing between tasks/threads which is not foreseen in HPCC Systems so far. In this talk, we describe how we implemented multi-threaded user-defined functions for the SSJ operator, implementing a C++ plugin for HPCC Systems, and how we visualized the relevant system parameters.

Speaker: Fabian Fier, Humboldt University of Berlin

Expanding HPCC Systems Deep Neural Network Capabilities: The training process for modern deep neural networks requires big data and large computational power. Though HPCC Systems excels at both of these, HPCC Systems is limited to utilizing the CPU only. It has been shown that GPU acceleration vastly improves Deep Learning training time. In this talk, I will explain how HPCC Systems became the first GPU accelerated library while also greatly expanding its deep neural network capabilities.

Speaker: Robert Kennedy, Florida Atlantic University

Theme 2: Cloud Enablement

Docker Support: Learn how to package the HPCC Systems Platform in a Docker container and deploy it locally, and build HPCC Systems Platform AMI followed by an AWS deployment.

Speakers: Xiaoming Wang & Godson Fortil, LexisNexis Risk Solutions

Progress Towards the Cloud: General discussion of the B1 datacenter migration toward the public cloud. Includes a high-level overview on decisions, planning, and methodology along with success stories and patterns.

Speaker: Jon Burger, LexisNexis Risk Solutions