Tuesday October 9, 2018

Time	Topic	Presenter
8:30am - 8:35am	**Welcome**	**Flavio Villanustre, LexisNexis Risk Solutions**
8:35am - 10:30am	**HPCC Systems in Industry: Real World Use Cases**	**Track 1**
8:35am - 8:55am	**Platinum Sponsorship Keynote: Driving Innovation with Artificial Intelligence**	**Badhri Krishnamoorthy, Cognizant**
	The fourth industrial revolution is here – powered by algorithms and fueled by data. Artificial Intelligence is pushing its way to the center of tomorrow's business value chain, disrupting business models, delivering engaging products and promising experiences like never before. To uncover unmet customer needs and establish market differentiation companies must take an AI-first, human-centric approach to create real value. Intelligence must begin with humans – empathizing with and designed for actual people – not just for customers, but also for the millennial workforce. Our Platinum sponsor, Cognizant, will deliver a keynote on how to innovate to survive and prepare to thrive in this growing age of artificial intelligence (AI).	
8:55am – 9:10am	**Gold Sponsorship Keynote: Soaring Through Emerging Technologies in the Big Data Era**	**Prasad Joshi, Infosys**
	If you are thinking emerging technologies and innovation, some of the things that come to your mind are: Chatbots, BlockChain, Adaptive Systems, Infosys Sense platform, AR/VR, BlockChain, Incubation–as–a-Service (IaaS) and more. In this talk, Prasad will share his experiences of working with clients in Emerging Technology and Innovation, and some of the work the Infosys team is doing in the Big Data space.	
9:10am - 9:30am	**Prepaid Banking on Steroids – Managing Massively Scalable Datasets with Ease**	**Jeff Lewis, Sutton Bank**

#HPCCSummit
Event Details: hpccsystems.com/hpccsummit2018
Livestream: http://bit.ly/2018hpccsummit

	DataDriven Approach gives Sutton Bank cutting-edge advantage over other players in the market. An HPCC Systems based platform FinanSeer developed by DataSeers makes smaller regional banks take on larger players in the market with a key advantage in the market space which revolves around speed and accuracy of data handling. HPCC Systems has automated some of the most trivial tasks that's have haunted the banking industry for many years and has posed serious problems in scaling the business in the prepaid world. Jeff Lewis, SVP of Sutton Bank, will explain how the HPCC Systems Based Solution made them leap ahead of competition and increase their efficiency ten-fold.	
9:30am - 9:50am	**HPCC Systems vs SAS: The Final Countdown**	**Luke Pezet, Archway Health Advisors**
	Archway Health shares their experience with using HPCC Systems alongside SAS for supporting a bundled payments program solution in the health industry.	
9:50am - 10:10am	**Integrating CRM and Sales Systems to Drive ROI and Increase Sales Projection Accuracy by 10x**	**David Dasher, CPL Online**
	CPL Online was formed in 2010 and specialises in bespoke digital services and products as well as e-learning training for the hospitality sector in the UK. In 2015, CPL Online introduced a series of added-value reporting and data-analysis functions for their clients thanks to the open source Big Data platform, HPCC systems®. These developments have allowed their clients to gain a better understanding of their workforce and benefit from significant cost savings. Since then as well as expanding those features they have now fully integrated their CRM / Visual Studio Team Services / Accounts into the same BIG Data platform to create analytics that run our business. In this presentation David Dasher, Chief Technical Officer, CPL Online, will explore how they use data from multiple external sources to build a Realtime P&L that can not only show sales per product to date but profitability and direct costs apportioned per product and as a business. The presentation will focus on the challenges CPL faced and how they have gone from 'struggling with SQL' to 'flourishing with HPCC Systems'. He will also explore the way in which CPL Online have built algorithms around the data allowing them to process and analyse vast amounts of data in real time. David will show real life examples of how this unique data is used to track user trends, such as spotting unusual or suspicious training activity and also highlighting the best performing staff as well as building internal systems to help various teams run the business.	
10:10am - 10:30am	**How HPCC Systems is Building the next generation Credit Bureau**	**David Wheelock, Mauricio Nunes de Oliveira, Robert Berger & Lucas Sobrinho, LexisNexis Risk Solutions**

	According to a Brazilian bureau market research study in March 2018, a total of 61.7 million Brazilians, which represents a staggering 40.5% of the country's population over 18 years old, are late on bill payments. The high delinquency causes interest rates charged in Brazil to be among the highest in the world, and they apply for everyone, regardless of credit history. A significant reason for this situation is that Brazil lacks a unified, national credit bureau that can provide the banks with an accurate indication of credit risk. Enter HPCC Systems. In this presentation, we will show you how HPCC Systems has been used to build a credit bureau from scratch in Brazil, making use of state-of-the-art ECL to automate not only file ingestion and profiling, but also to automatically generate ECL for the complete data pipeline processing. Additionally, we will show how we used HPCC Systems to easily create and test analytics attributes to deliver risk related products, such as credit and fraud scores. These scores will enhance the ability of credit grantors in Brazil to leverage the data available, enabling them to better distinguish the good payers from the bad ones, and ultimately allowing for lower interest rates that increase access to credit and foster economic growth.	
10:30am - 10:45am	**Break - Poster Presentations**	
10:45am - 12:00pm	**HPCC Systems in Academia: Beyond the Classroom**	**Track 2**
10:45am - 11:10am	**Deep Content Learning in Traffic Prediction and Text Classification**	**Jingqing Zhang, Imperial College of London**
	In this talk, Jingqing will introduce recent advances at the Data Science Institute, Imperial College London, and focus on a general framework named Deep Content Learning. Two recent projects will be discussed as examples. In the traffic prediction project, we released a new large-scale traffic dataset with auxiliary information including search queries from Baidu Map app and proposed hybrid models to achieve state-of-the-art prediction accuracy. The other project on zero-shot text classification integrated semantic knowledge and used a two-phase architecture to tackle the challenging zero-shot learning in textual data. The integration of TensorLayer and HPCC Systems will be discussed in the talk.	
11:10am - 11:35am	**Parallel Distributed Deep Learning on HPCC Systems**	**Taghi Khoshgoftaar & Robert Kennedy, Florida Atlantic University**
	The training process for modern deep neural networks requires big data and large amounts of computational power. Combining HPCC Systems and Google's TensorFlow, Robert created a parallel stochastic gradient descent algorithm to provide a basis for future deep neural network research, thereby helping to enhance the distributed neural network training capabilities of HPCC Systems.	
11:35am - 12:00pm	**Autonomous Agricultural Robot: Is the Machine Uprising Coming Sooner Than You Think?**	**Taiowa Donovan & Robotics Team, American Heritage School**
	Hear how HPCC Systems is being used by a team of high school students to build an autonomous robot for the agricultural industry to help provide time sensitive data to a farm management system.	
12:00pm - 1:00pm	**Lunch**	

1:00pm - 1:15pm	**Community Award Ceremony: Poster Winners Announced, Community Recognition Award**	**Flavio Villanustre & Vijay Raghavan**
1:15pm - 3:15pm	**HPCC Systems Roadmap Tech Talks**	**Track 3**
1:15pm - 1:35pm	**Data Patterns - A Native Open Source Data Profiling Tool for HPCC Systems**	**Dan Camper, LexisNexis Risk Solutions**
	Data profiling is a technique used to uncover information about a source of data. Information such as the shape or accuracy of the data is extremely useful during data discovery (when you're exploring a new dataset) or when verifying that updated data appears to be a valid replacement for old data. DataPatterns, an open sourced ECL bundle for HPCC Systems, offers a native function macro for data profiling that is easy to use and supports a number of options for tuning the profile result. This talk will briefly explore the bundle's profile feature and options.	
1:35pm - 2:00pm	**Making IoT Data Actionable Using Predictive Analytics**	**Dan Camper & Hicham Elhassani, LexisNexis Risk Solutions**
	This is a proof-of-concept where an HPCC Systems cluster is used to gather current IoT device data from opt-in subscribers. The cluster's architecture and collected data will be described in the presentation, as well as the additional datasets (e.g. property characteristics, weather, etc.) brought in to enhance the data for analysis using predictive analytics for potential applications in the insurance industry.	
2:00pm - 2:25pm	**Learning Trees - Decision Tree Learning Methods**	**Roger Dev, LexisNexis Risk Solutions**
	Decision Tree based Machine Learning algorithms are among the most powerful and easiest to use. The new Learning Trees bundle from HPCC Systems provides a robust library of tree-based methods including Random Forests, Gradient Boosted Trees, and Boosted Forests. How do these algorithms work, and which are likely to provide the best results? This talk provides details of various Tree-Based learning methods and insight into the data science involved.	
2:25pm - 2:50pm	**A First Look at HPCC Systems 7.0, Innovation in Action**	**Gavin Halliday, LexisNexis Risk Solutions**
	The latest version of the platform contains improvements to functionality, usability and interoperability. This talk gives an overview of the changes and explains how you might find them useful.	
2:50pm - 3:15pm	**Innovation with Connection, The new HPCC Systems Plugins and Modules**	**James McMullan, LexisNexis Risk Solutions**
	The HPCC Systems platform team continues to expand interoperability with third party systems, which increases the platform feature-set and facilitates custom solutions. James will share an update on the latest connectors available, including the Spark-HPCC, and the upcoming HDFS connector plugin.	
3:15pm - 3:30pm	**Break - Poster Presentations**	
3:30pm - 5:00pm	**HPCC Systems Breakouts**	**Track 4**
3:30pm - 4:10pm	**Breakout Session Rotation 1**	

#HPCCSummit
Event Details: hpccsystems.com/hpccsummit2018
Livestream: http://bit.ly/2018hpccsummit

Chairperson: Flavio Villanustre	**Poster Presentation: How to Be Rich: A Study of Monsters and Mice of American Industry**	**Zhe Yu, NC State University**
	Documentation & Training: Optimizing Set-Similarity Join and Search with Different Prefix Schemes	**Fabian Fier, Humboldt University Berlin**
	Finding duplicate textual content is crucial for many applications, especially plagiarism detection. When dealing with millions of documents finding duplicate content becomes very time-consuming. Thus it needs scalable and efficient data structures and algorithms that solve this task in seconds rather than hours. In my talk, I present an optimization of a common filter-and-verification set-similarity join and search approach. Filter-and-verification means that we only consider such pairs of objects which share a common word or token in a prefix. Such pairs are potentially similar and are verified in a subsequent step. The candidate set is usually orders of magnitudes smaller than the cross product over an input set. We optimizied this approach by regarding overlaps larger than 1, which reduces the candidate set further and makes the verification faster. On the other hand this requires larger prefixes, which use more memory. Our experiments using HPCC Systems show that we can usually optimize the runtime by choosing an overlap different from the standard overlap 1.	
Chairperson: Roger Dev	**Poster Presentation: Equivalence Terms of Text Search Bundle**	**Farah Alshanik, Clemson University**
	Poster Presentation: Explore the Linguistics of Public Records on HPCC Systems	**Lili Xu, Clemson University**
	Machine Learning: Using HPCC Systems ML to Map Thousands of Public Records Data Descriptions to Standard Codes	**Lili Xu, Clemson University & Gus Reyna, LexisNexis Risk Solutions**
	There is a challenge of incorporating public records data into business processes given disparate descriptions across states for similar events, and then finding a standard that gives one consistent meaning for use. This session tells the story of how the HPCC Systems Machine Learning addressed the problem of mapping thousands of disparate public record data descriptions to a corresponding set of standard codes and the future direction for this approach.	
Chairperson: Richard Chapman	**Poster Presentation: MPI Proof of Concept**	**Saminda Wijeratne, Georgia Tech**
	System Tools: Automated Test Systems for QA in your HPCC Systems Environment	**Attila Vamos, LexisNexis Risk Solutions Suwanee**
	In the last 5 years, the Platform team progressed from an ad-hoc, mostly manual testing practice to different levels of automated test system environments. In this session, I will talk about the test collections: regression and performance suites, how to use the test engine in manual and in automatization, our examples of automated test harnesses: OBT and the Smoketest process, and how the automated systems report their results.	
Chairperson: Bob Foreman	**Poster Presentation: Measuring the Geo-Social Distribution of Opioid Prescriptions**	**Nicole Navarro, New College of Florida**
	User Interfaces: Visualizing your Data Natively on the HPCC Systems Platform with the "Visualizer Bundle"	**Gordon Smith, LexisNexis Risk Solutions Buckhead B**
	The Visualizer Bundle continues to improve with new features and enhancements enabling the user to create slick, sophisticated reports directly from ECL without needing additional plugins or third-party tools. Join me in this session as I look at how to leverage this bundle to visualize your data as well as a quick look under the covers to see how it integrates with ECL and ECL Watch.	

4:10pm - 4:20pm	**Room Change Break**	
4:20pm - 5:00pm	**Breakout Session Rotation 2**	
Chairperson: **Richard Taylor**	**Poster Presentation: Distributed Deep Learning on HPCC Systems** **Documentation & Training: Preparing an Open Source Documentation Repository for Translations**	**Robert Kennedy, Florida Atlantic University** **Jim DeFabia, LexisNexis Risk Solutions** **Augusta**
	Translating a manual once is not a terribly difficult task. However, our manuals are always evolving, so we needed a plan to update translations on a regular basis. This requires a process that is maintainable, repeatable, and robust. In this case study of our forays into documentation internationalization, you can learn from our successes and laugh at some of our missteps along the way.	
Chairperson: **Roger Dev**	**Poster Presentation: Dimensionality Reduction on Pbblas** **Poster Presentation: Cervical Cancer Risk Factors: Exploratory Analysis using HPCC Systems** **Machine Learning: Predicting College STEM Enrollment using HPCC Systems in Educational Research**	**Shah Muhummad Hamdi, Georgia State University** **Itauma Itauma, Keiser University** **Itauma Itauma, Keiser University** **Buckhead A**
	In this study, multiple regression analysis is used to determine if high school students' perception of their science self-efficacy, identity, and utility predict consideration to enroll in a college stem major. The HPCC Systems ML library will be used to build a multiple regression model utilizing secondary data analysis of the United States High School Longitudinal Study of 2009 (HSLS:09) dataset. The HSLS:09 is a national cohort study of over 23,000 ninth graders from 944 schools in 2009 through their secondary and post-secondary years including choices of college majors and careers. This study demonstrates the use of the HPCC Systems ML library for statistical modeling in education.	
Chairperson: **Kunal Aswani**	**Poster Presentation: HPCC Systems Robotics Sensor Interface** **System Tools: Visualizing HPCC Systems Log Data Using ELK**	**Aramis Tanelus, American Heritage School** **Rodrigo Pastrana & Miguel Vazquez, LexisNexis Risk Solutions** **Suwanee**
	Find out how to utilize ELK (a powerful log management stack comprised of Elastic Search, Logstash and Kibana) to visualize your HPCC Systems log data to help deliver actionable insights in real time, trend analytics, system monitoring and much more. In this session, we will walk through the log data extraction, visualization dashboard creation, and discuss related HPCC Systems and ELK insights.	
Chairperson: **Flavio Villanustre**	**Poster Presentation: The Future of Automotive Telemetry: Assessing Inherent Risk Implications and Cyber Security Vulnerabilities** **User Interfaces: Using the Open Source VS Code Editor with the HPCC Systems Platform**	**Matt Butler, Kennesaw State University** **Arjuna Chala, LexisNexis Risk Solutions** **Buckhead B**
	Are you a fan of the VS Code Editor for building and debugging your code? Join me as I demonstrate how this can be used as a modern, extensible alternative to the ECL IDE and why most developers choose the VS Code editor due to its benefits of being cross-platform, multi-language support, and open source.	
5:00pm	**Adjourn**	

#HPCCSummit
Event Details: hpccsystems.com/hpccsummit2018
Livestream: http://bit.ly/2018hpccsummit