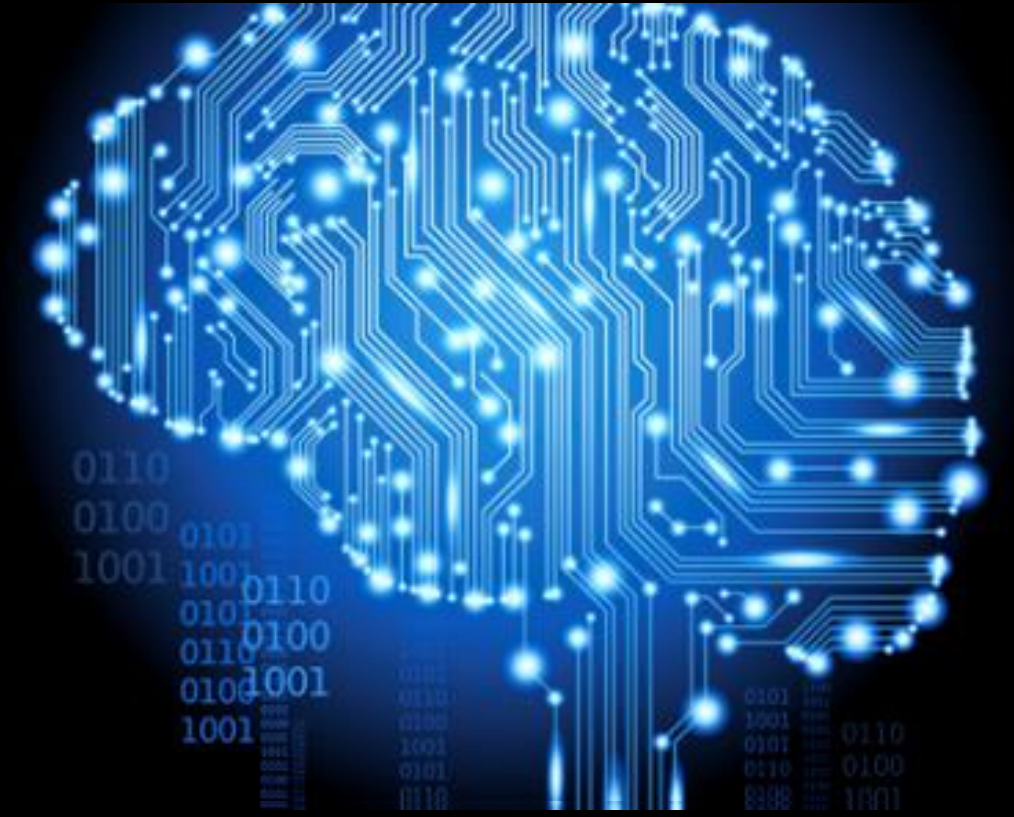
THE DOWNLOAD

TECH TALKS BY HPCC SYSTEMS



HPCC SYSTEMS®

The Download: Community Tech Talks Episode 4

April 20, 2017

RELX Group

Welcome!

- Please share: Let others know you are here with #HPCCTechTalks 
- Ask questions! We will answer as many questions as we can following each speaker.
- Look for polls at the bottom of your screen. Exit full-screen mode or refresh your screen if you don't see them.
- We welcome your feedback - please rate us before you leave today and visit our blog for information after the event.
- Want to be one of our featured speakers? Let us know! techtalks@hpccsystems.com

HPCC SYSTEMS®

Community Announcements

- Two days left for students to submit a proposal for the HPCC Systems Summer Internship Program!
  - Deadline extended to April 22
  - Details at https://hpccsystems.com/intern2017

- Please promote the call for Presentations and Poster Abstracts for the 2017 HPCC Systems Community Day!
  - DATE CHANGE - Community Day will be held in Atlanta on October 4, 2017
  - Poster Competition held on October 3.
  - Submission deadline on June 30
  - Reminder: Sponsorship opportunities available to allow our partners to have a bigger presence
  - Details at https://hpccsystems.com/hpccsummit2017

Dr. Flavio Villanustre
VP Technology
LexisNexis® Risk Solutions
Flavio.Villanustre@lexisnexis.com

HPCC SYSTEMS®

Today's Speakers

Gordon Smith

Enterprise/Lead Architect
LexisNexis Risk Solutions
gordon.smith@lexisnexisrisk.com

Gordon is an Enterprise/Lead Architect and manager of the HPCC Systems supercomputer clients. He is a member of the HPCC Core Platform team and a LexisNexis employee for over 18 years. Gordon is the principle developer for ECL related development and visualization tools, including the ECL IDE, ECL Plugin for Eclipse, ECL Watch, ECL Execution Graph Viewer and more recently the HPCC Visualization Framework.

Gordon is also involved in our HPCC Systems intern program and serves as a mentor on any projects relating to the HPCC Visualization Framework and/or a web based debugging front end for ECL.

John Holt

Enterprise/Lead Architect
LexisNexis Risk Solutions
John.d.holt@lexisnexisrisk.com

Dr. Holt is an Enterprise/ Lead Architect for LexisNexis Risk Solutions. Dr. Holt directs various projects such as the evolution of the Insurance Applications Systems to help assess risk and detect fraud which leverages the HPCC Systems platform.

Dr. Holt has been with LexisNexis for 36 years. Prior positions have included system architecture for the Risk Solutions Fabrication Systems, system architecture for the LexisNexis online system, project management, product management, and product development. Dr. Holt holds a PhD and an MS in Computer Science from Wright State University, an MBA from Wright State University, and a BS in Data processing from the University of Dayton.

HPCC SYSTEMS®

Today's Speakers

David de Hilster

Consulting Software Engineer

LexisNexis Risk Solutions

David.dehilster@lexisnexisrisk.com

David de Hilster is a consulting software engineer working on the development efforts of the ECL IDE component of the HPCC Systems platform. David has more than 20 years' experience in research, design, programming, and bringing innovative ideas to the market place. David has developed numerous software designs including a resume processor, an online card room, and a Visual Studio-like environment for creating analyzers that process human language. Known for rapid prototyping, enthusiasm, creativity, and ability to communicate technical ideas to non-technical clients.

Jessica Lorti

Director Marketing

LexisNexis Risk Solutions

Jessica.Lorti@lexisnexisrisk.com

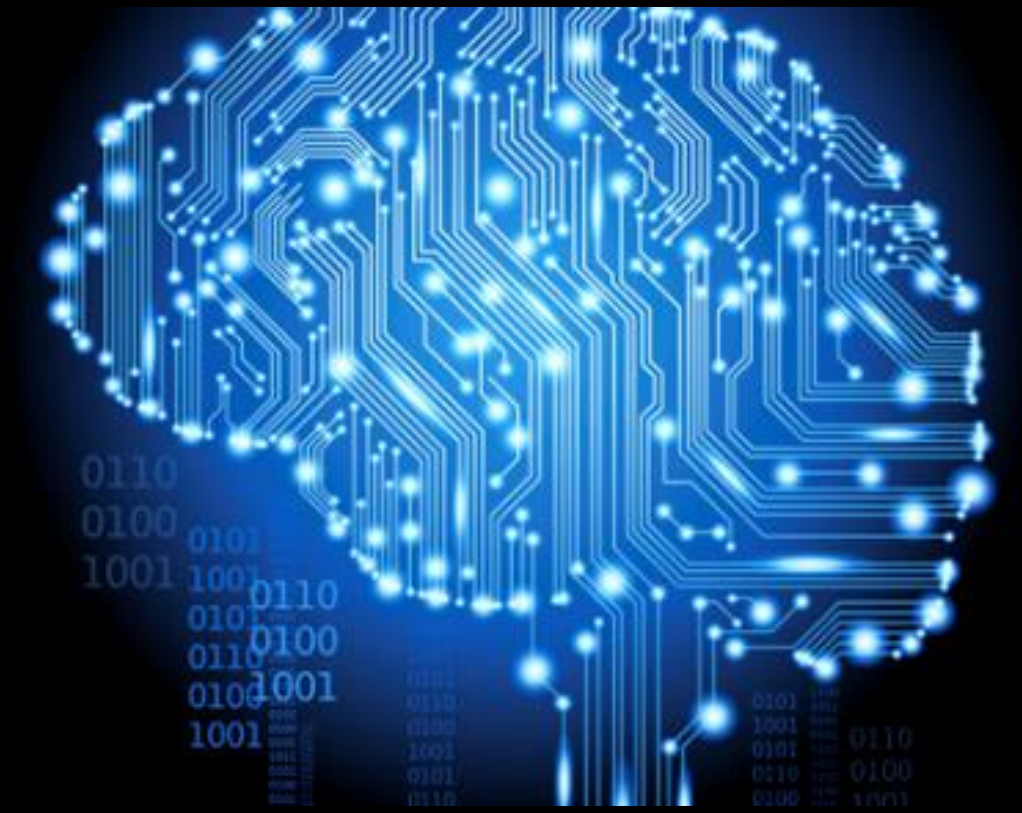
Jessica comes from an extensive background defining and implementing strategic programs across a variety of marketing disciplines for the technology, financial services, and energy industries. She has held senior marketing roles at GE, Intel, Compaq and Grant Thornton where she managed product marketing and brought new technologies to market, developed and launched social media and online marketing efforts, and developed new business models in conjunction with sales and key corporate partners.

Jessica holds a Bachelor of Science in International Economics from Texas Tech University and a Masters in International Management with a concentration in Marketing from Thunderbird, the Global School of International Management. She has also earned the LEAN Six Sigma Green Belt certification.

HPCC SYSTEMS®

THE DOWNLOAD

TECH TALKS BY HPCC SYSTEMS



HPCC SYSTEMS®

Visualizer – The ECL Bundle

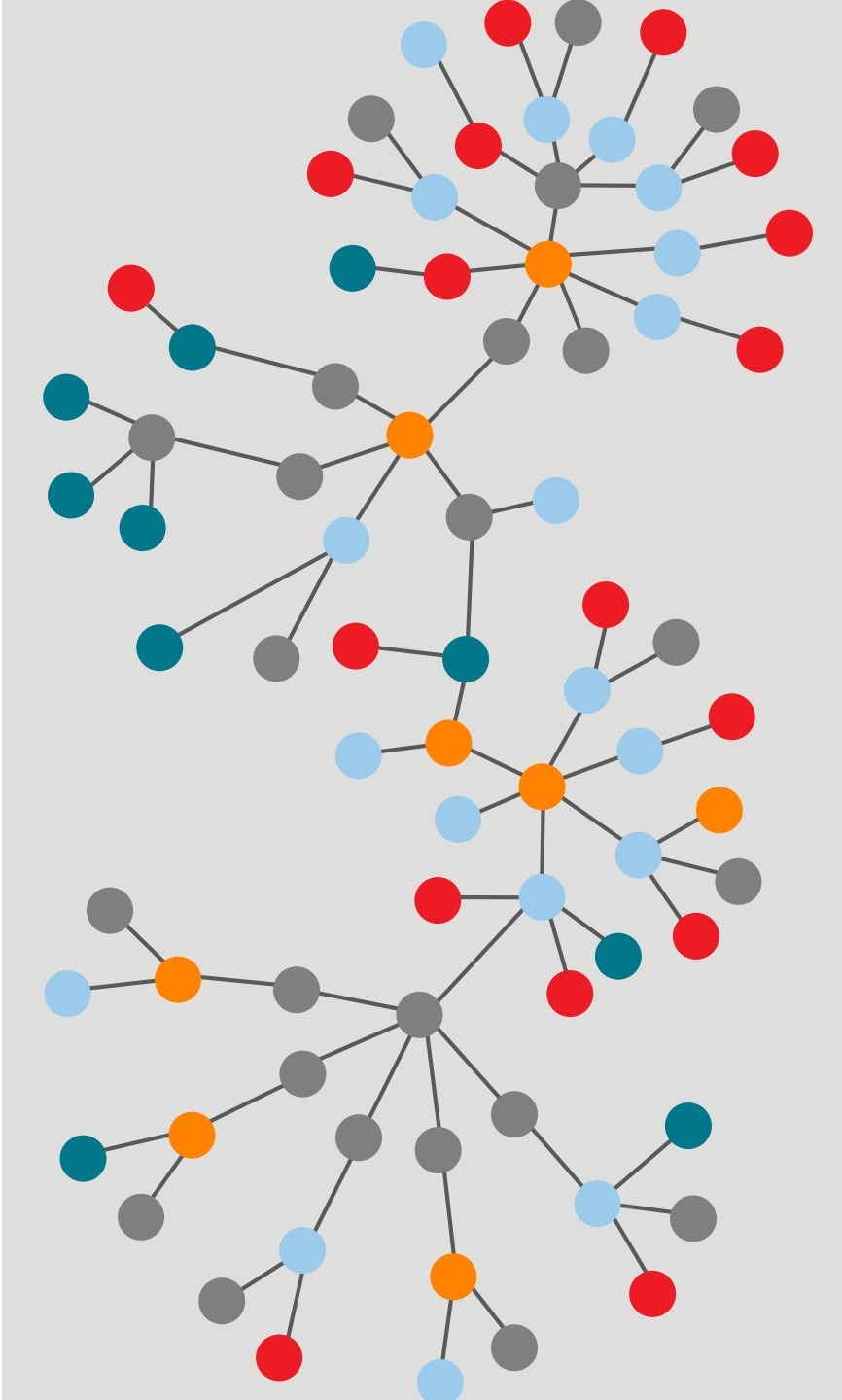
Gordon Smith
Enterprise/Lead Architect
LexisNexis Risk Solutions

LexisNexis® RISK SOLUTIONS

Quick poll: Do you consider visualizations to be an important analytics aid during the R&D of your big data?

See poll on bottom of presentation screen

HPCC SYSTEMS®



Getting Started

- Download
  https://github.com/hpcc-systems/Visualizer/archive/master.zip
- Unzip to “Visualizer” folder
  …\Downloads\Visualizer-master.zip -> …\Downloads\Visualizer
- Install (command prompt)

```
set PATH=%PATH%;"c:\Program Files (x86)\HPCCSystems\6.2.0\clienttools\bin"
ecl bundle install %USERPROFILE%\Downloads\Visualizer
```

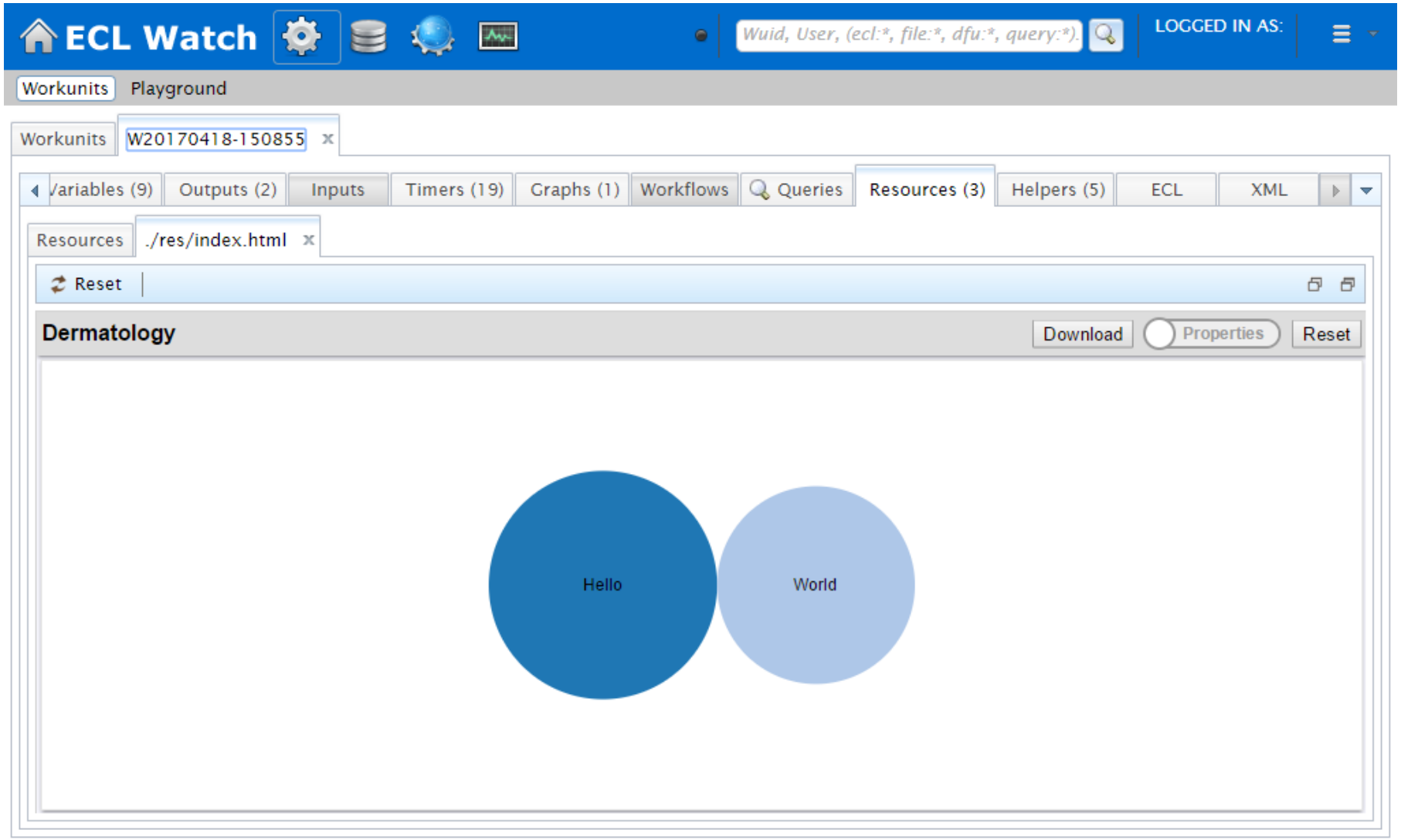
- Success:

```
Installing bundle Visualizer version 1.0.0
Visualizer    1.0.0      ECL Visualization Bundle
Installation complete
```

Hello World

```
IMPORT Visualizer;

ds := DATASET([     {'Hello', 20},
                    {'World', 15}],
                    {STRING label, INTEGER4 weight});

OUTPUT(ds, NAMED('HelloWorldViz'));

Visualizer.TwoD.Bubble('bubble’,, 'HelloWorldViz');
```

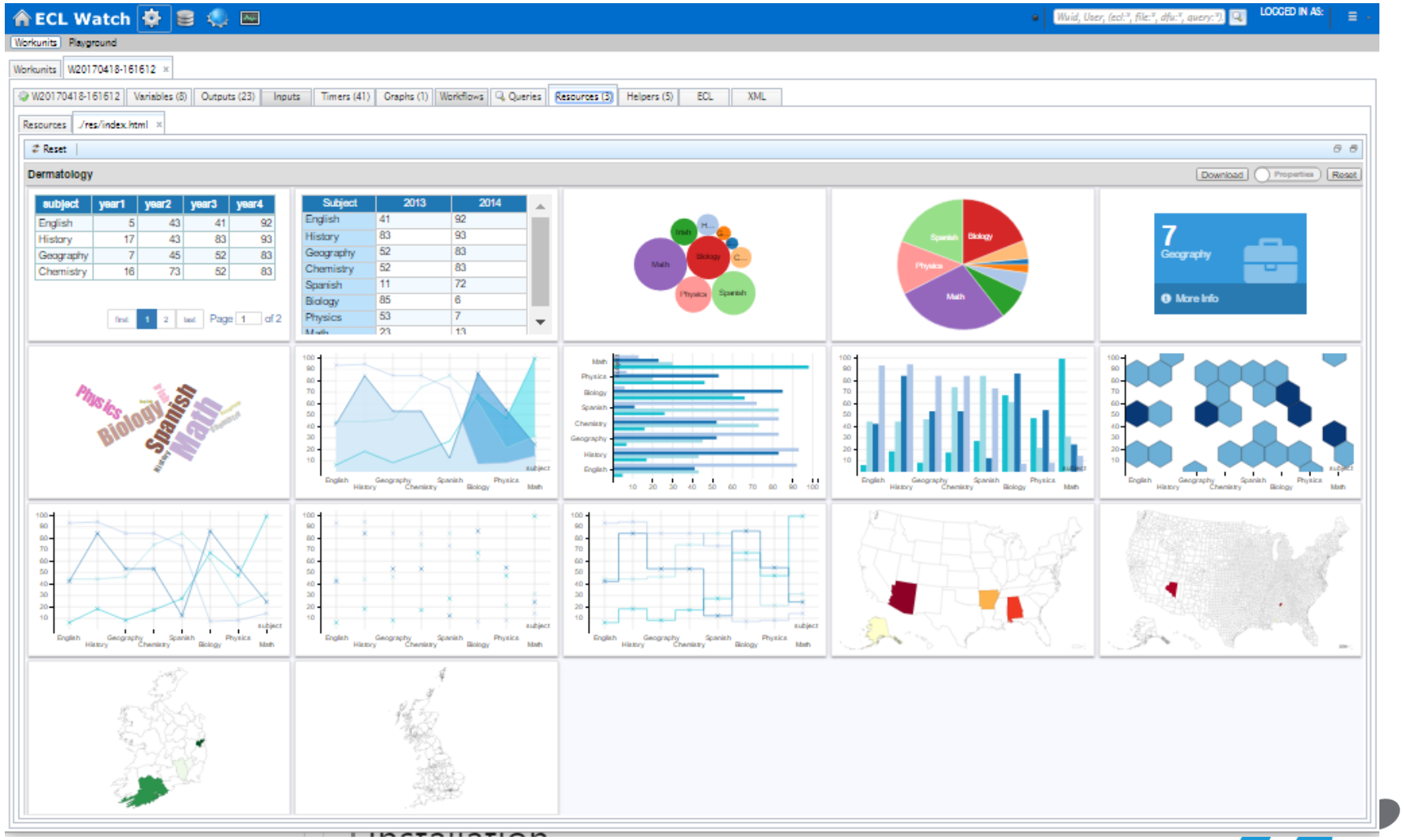
Hello World



HPCC SYSTEMS®

Self Test / Quick Demo / Read the Source Luke

```
IMPORT Visualizer;
Visualizer.main;
```

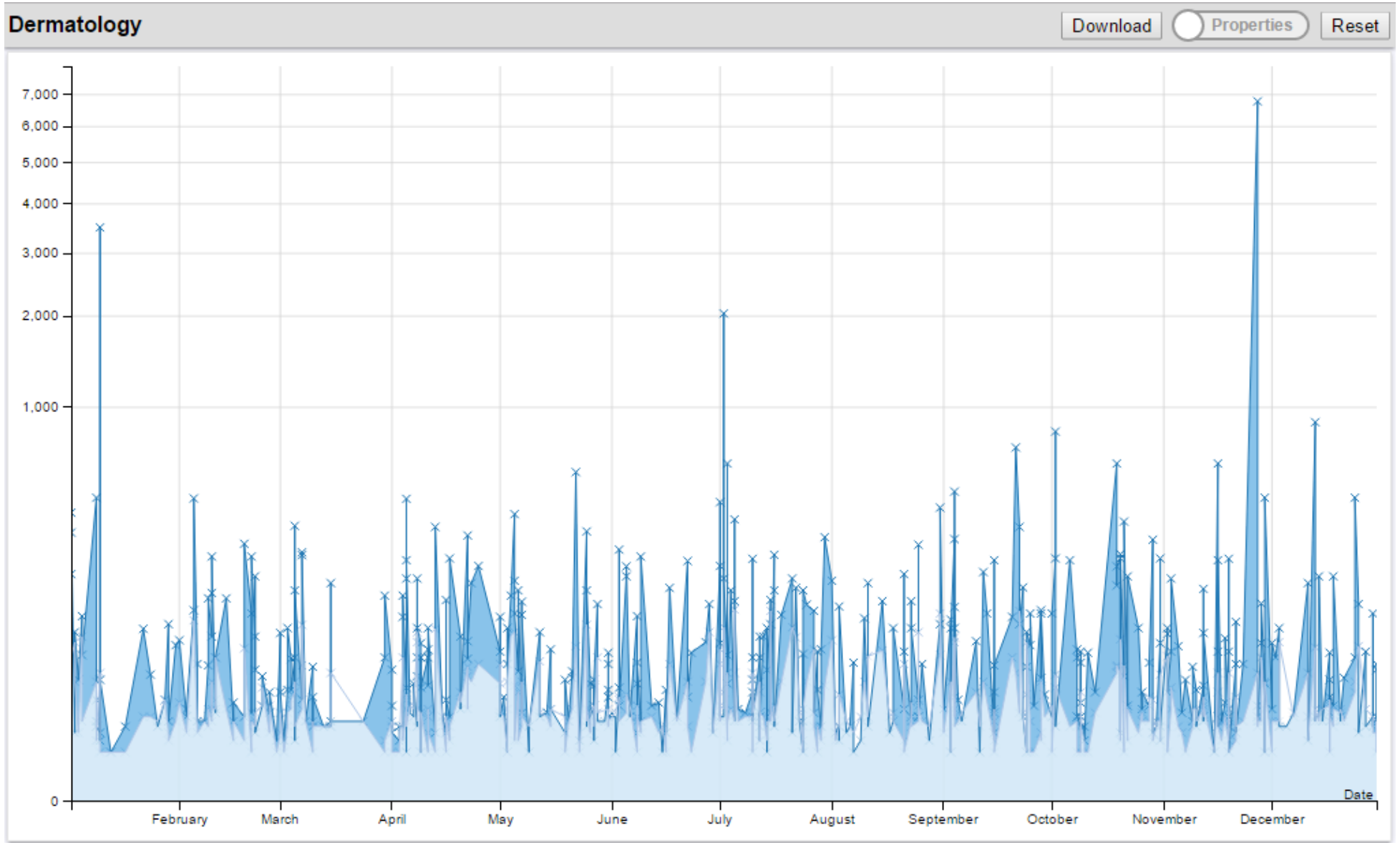


HPCC SYSTEMS®

Inline Dermatology

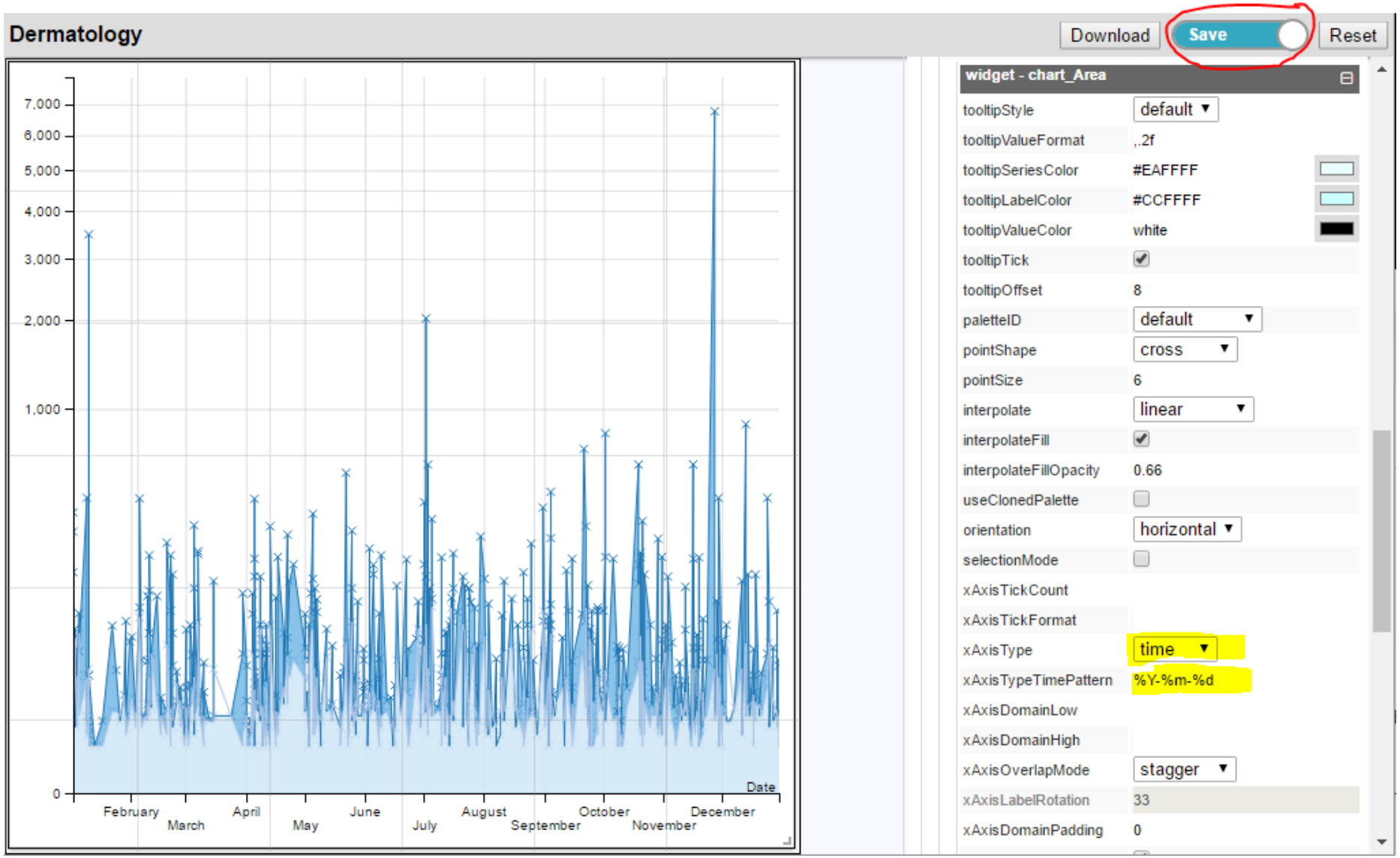
```
IMPORT SampleData.Sales;
IMPORT Visualizer;

OUTPUT(CHOOSEN(SORT(Sales.CleanDataset(Region='West'), Fixed_Order_Date), ALL), NAMED('Sales'));

// Declare some "dermatology" properties
properties := DATASET([  {'xAxisType', 'time'},
                         {'xAxisTypeTimePattern', '%Y-%m-%d'},
                         {'yAxisType', 'pow’},
                         {'yAxisTypePowExponent', 0.3}],
                         Visualizer.KeyValueDef);

Visualizer.MultiD.area('myChart',, 'Sales',,, properties );
```

Inline Dermatology



HPCC SYSTEMS®

Inline Dermatology

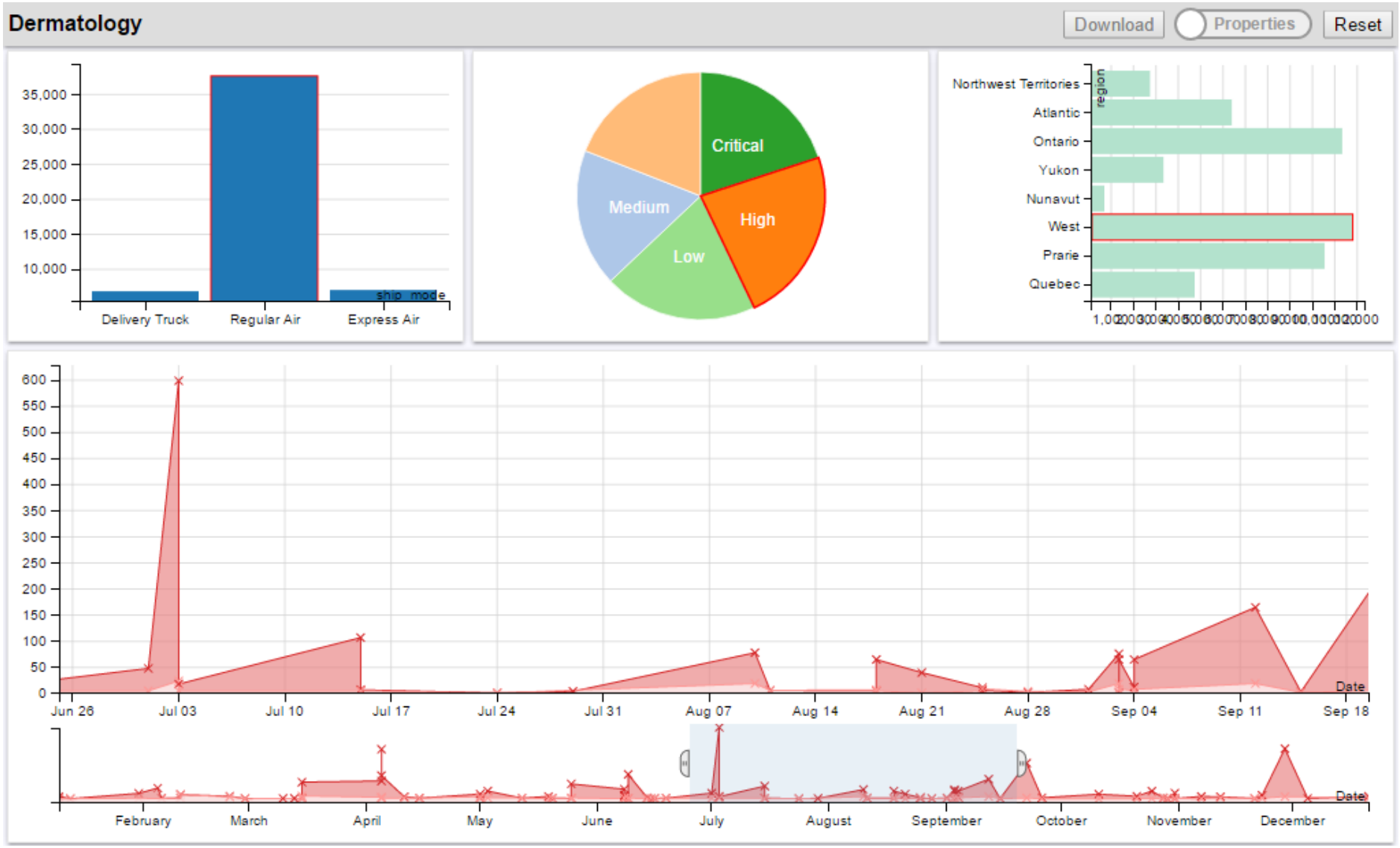


HPCC SYSTEMS®

Interactive Dashboard

```
// Aggregate by Ship_Mode ---
OUTPUT(TABLE(Sales.CleanDataset, {Ship_Mode, UNSIGNED INTEGER4 Sum_Order_Quantity := SUM(GROUP, Order_Quantity)}…);
Visualizer.MultiD.Column('myColumnChart',, 'Ship_Mode',,,, Visualizer.KeyValueDef));

// Aggregate by Order_Priority ---
OUTPUT(TABLE(Sales.CleanDataset, {Order_Priority, UNSIGNED INTEGER4 SumOrderQuantity := SUM(GROUP, …);
Visualizer.TwoD.Pie('myPieChart',, 'Order_Priority');

// Aggregate by Region ---
OUTPUT(TABLE(Sales.CleanDataset, {Region, UNSIGNED INTEGER4 SumOrderQuantity := SUM(GROUP, Order_Quantity)}, …);
Visualizer.MultiD.Bar('myBarChart',, 'Region');

// All data filtered by previous visualizations ---
OUTPUT(CHOOSEN(SORT(Sales.CleanDataset, Fixed_Order_Date), ALL), NAMED('Sales'));

filter := DATASET([ {'myColumnChart', [{'Ship_Mode', 'Ship_Mode'}]},
{'myPieChart', [{'Order_Priority', 'Order_Priority'}]},
{'myBarChart', [{'Region', 'Region'}]}], Visualizer.FiltersDef);

Visualizer.MultiD.Area('myLine',, 'Sales',, filter);
```

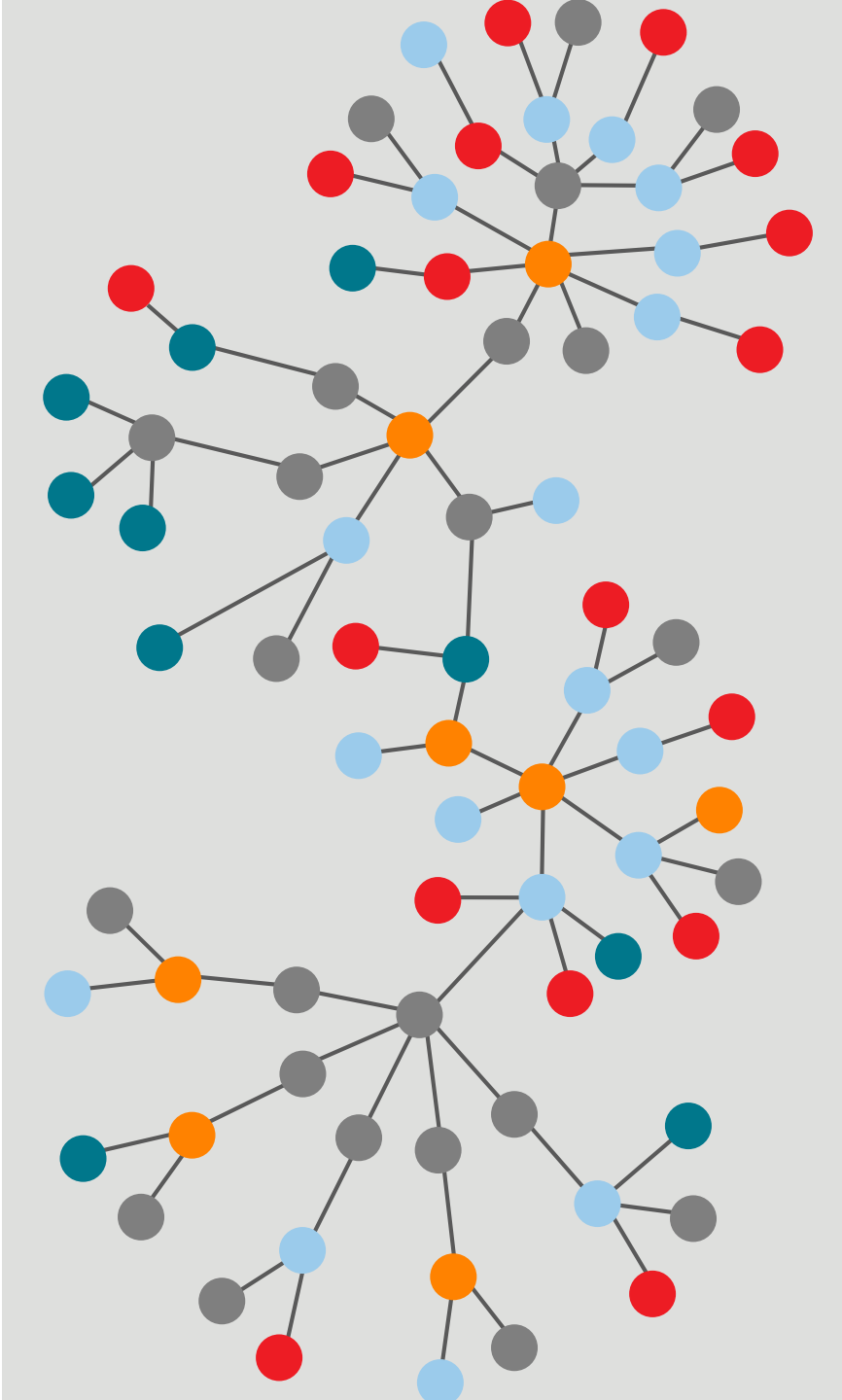
Interactive Dashboard



HPCC SYSTEMS®

Quick poll: Do you consider dashboards to be an important asset as part of the final product?

See poll on bottom of presentation screen



HPCC SYSTEMS®

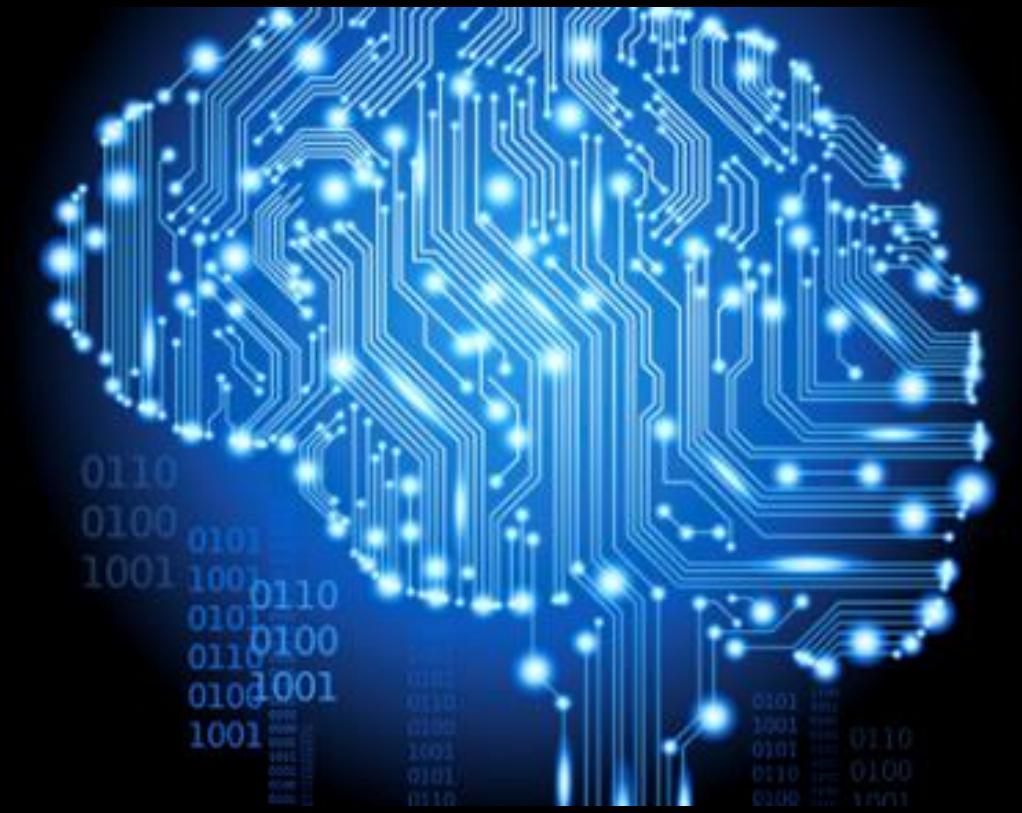
Questions?

Gordon Smith
Enterprise/Lead Architect
LexisNexis Risk Solutions
gordon.smith@lexisnexisrisk.com

HPCC SYSTEMS®

THE DOWNLOAD

TECH TALKS BY HPCC SYSTEMS



An Update on the Machine Learning Bundles

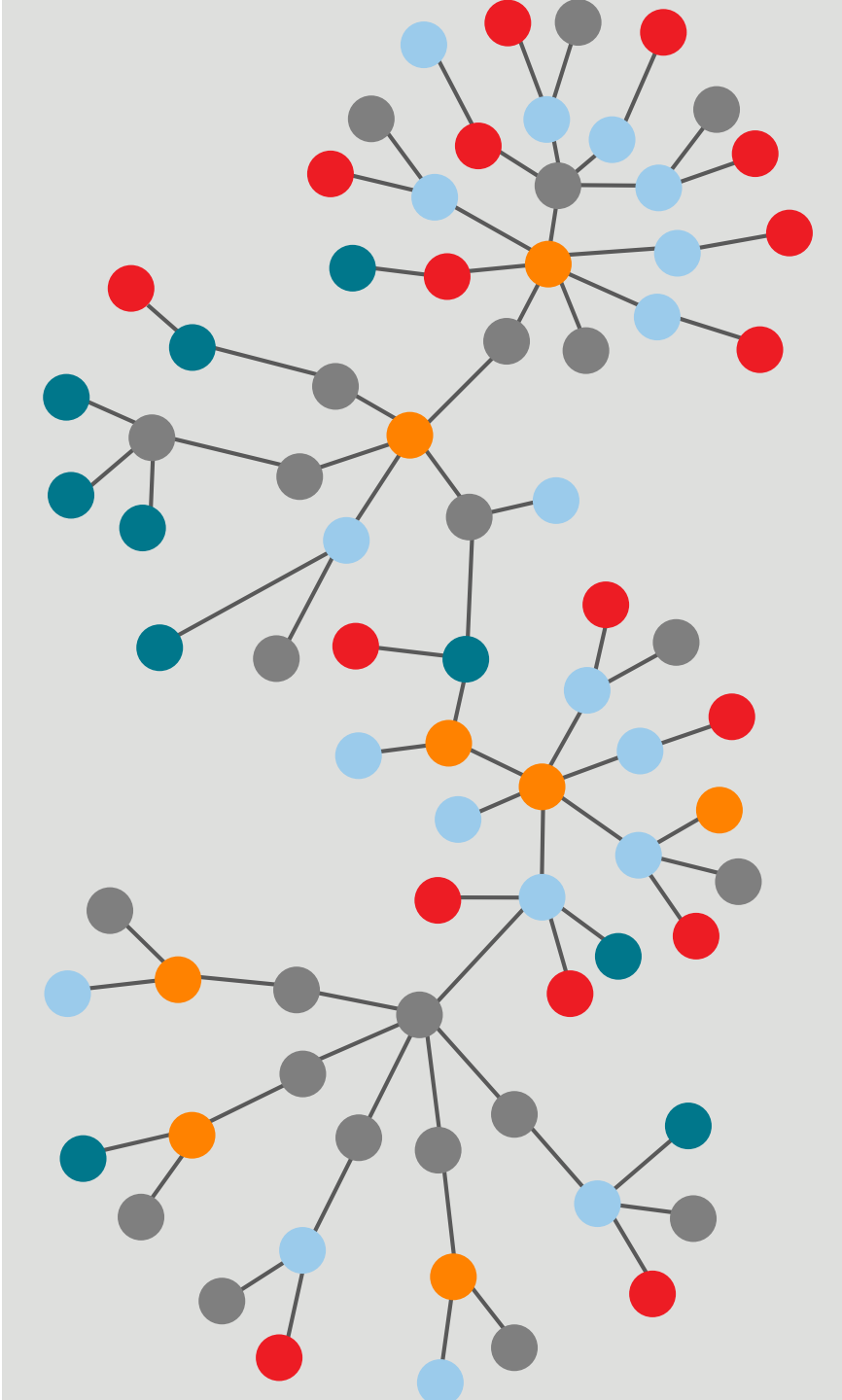
HPCC SYSTEMS®

John Holt
Enterprise/Lead Architect
LexisNexis Risk Solutions

LexisNexis® RISK SOLUTIONS

Quick poll: Which prediction machine learning algorithms do you use most?

See poll on bottom of presentation screen



HPCC SYSTEMS®

Topics

- Short review of the restructure
- Prerequisites
- Machine Learning bundles for 6.4
- Validation testing
- Machine Learning bundles for 7.0
- Closing poll followed by Q & A

Review of Restructure

- Myriad interface
  - Many problems treated in a similar manner
  - Aligned with SIMD nature of THOR (Single Instruction Multiple Data)
- Feature and Capability oriented bundles instead of one big collection
- Publish validation results and validation based upon known implementation
- Specific performance profiles
  - Large problem profile => Requires more than a single node
  - Myriad problem profile => Each problem can typically be run on a single node
- The ecl-bundles repository is a central bundles list

Prerequisites

- Platform version 6.2 is required
  - These ML bundles need Basic Linear Algebra System (BLAS) support
  - BLAS support was added to the ECL Standard Library with the 6.2 version of the platfrom
- Platform version 6.4 is desired
  - Compile times will be longer than usual on platform version earlier than 6.4
- Logistic Regression and Linear Regression bundles require ML Core and PBblas bundles

Machine Learning Bundles for the 6.4 Platform

- Logistic Regression
  - Binomial response variable
  - Uses Myriad to handle multiple response variables in parallel
  - Predicts a Yes/No (Binomial) response value based upon values of one or more measurement or cardinal variables
- Multiple Linear Regression
  - Ordinary Least Squares Linear Regression
  - Multiple Regression Supported (multiple independent variables)
  - Multivariate Regression (multiple dependent variables)
  - Myriad Support (multiple separate regressions in one operation)
  - Rich set of analytic functions (R-squared, ANOVA, AIC, confidence intervals, etc.)

Validation Testing

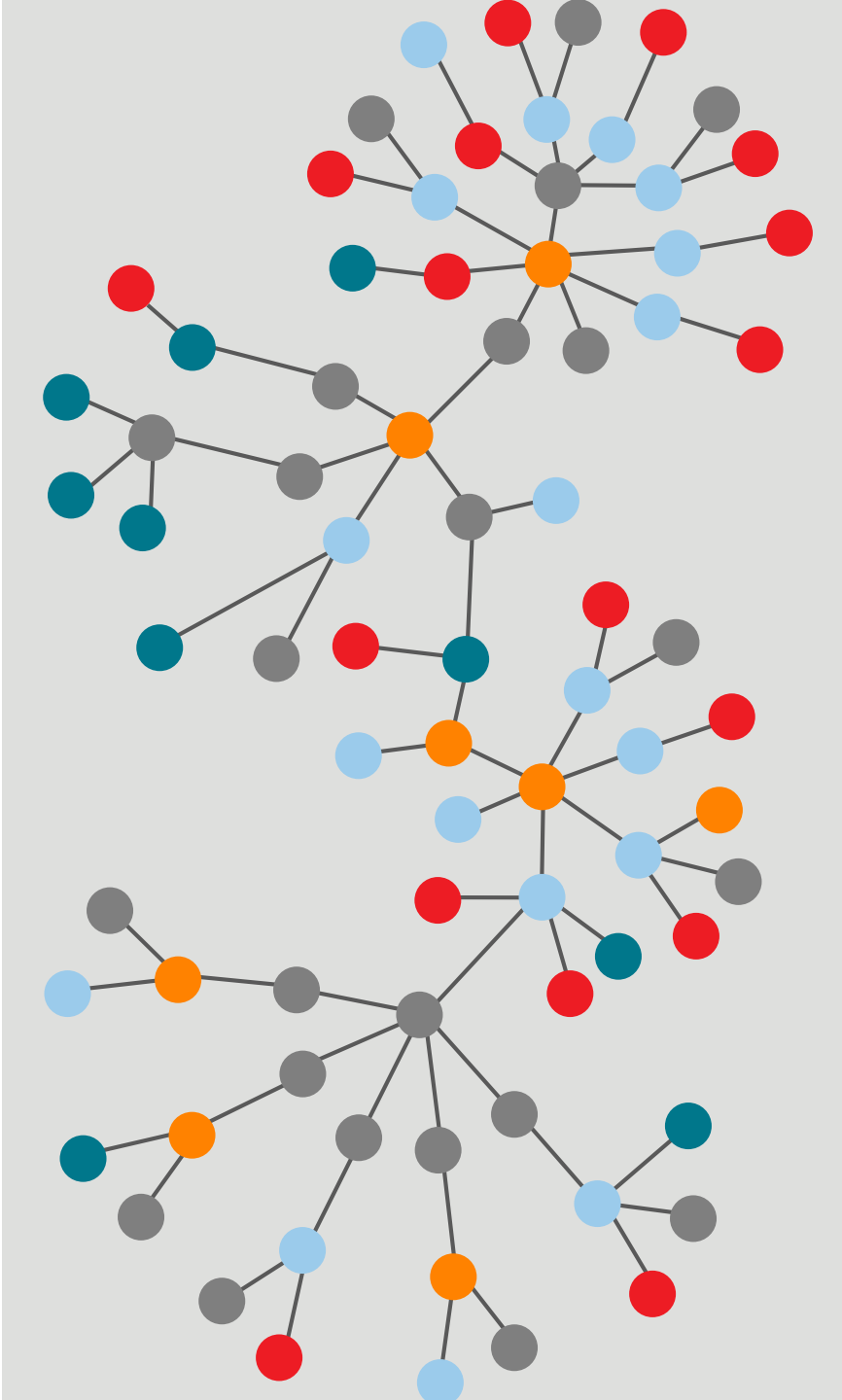
- Logistic Regression
  - Validated (matched) against Python Stats Models Logit package
  - Similar results to the GLM package in R
  - Differences due to variations in approach such as regularization
  - Differences appear as small differences in the coefficients
- Multiple Linear Regression
  - All functions validated against Python Stats Models

Machine Learning Bundles for 7.0 Platform

- SVM (may make 6.4)
  - Leverages Myriad by generating a grid search for hyper parameters
- Stepwise Logistic Regression and Logistic Regression for the multinomial case
- Stepwise Multiple Linear Regression
- Descriptive Statistics

Quick poll: When it comes to predictive analytics, the statement most applicable to me is:

See poll on bottom of presentation screen



HPCC SYSTEMS®

Questions?

John Holt

Enterprise/Lead Architect,

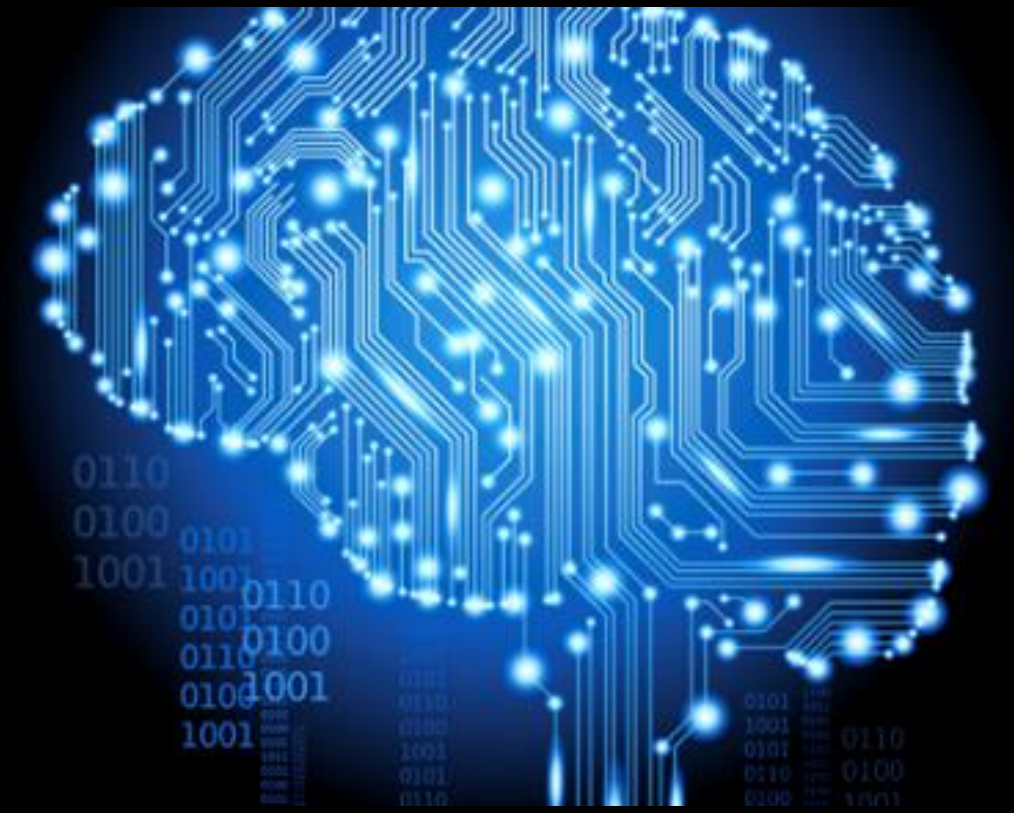
LexisNexis® Risk Solutions

John.d.holt@lexisnexisrisk.com

HPCC SYSTEMS®

THE DOWNLOAD

TECH TALKS BY HPCC SYSTEMS



HPCC SYSTEMS®

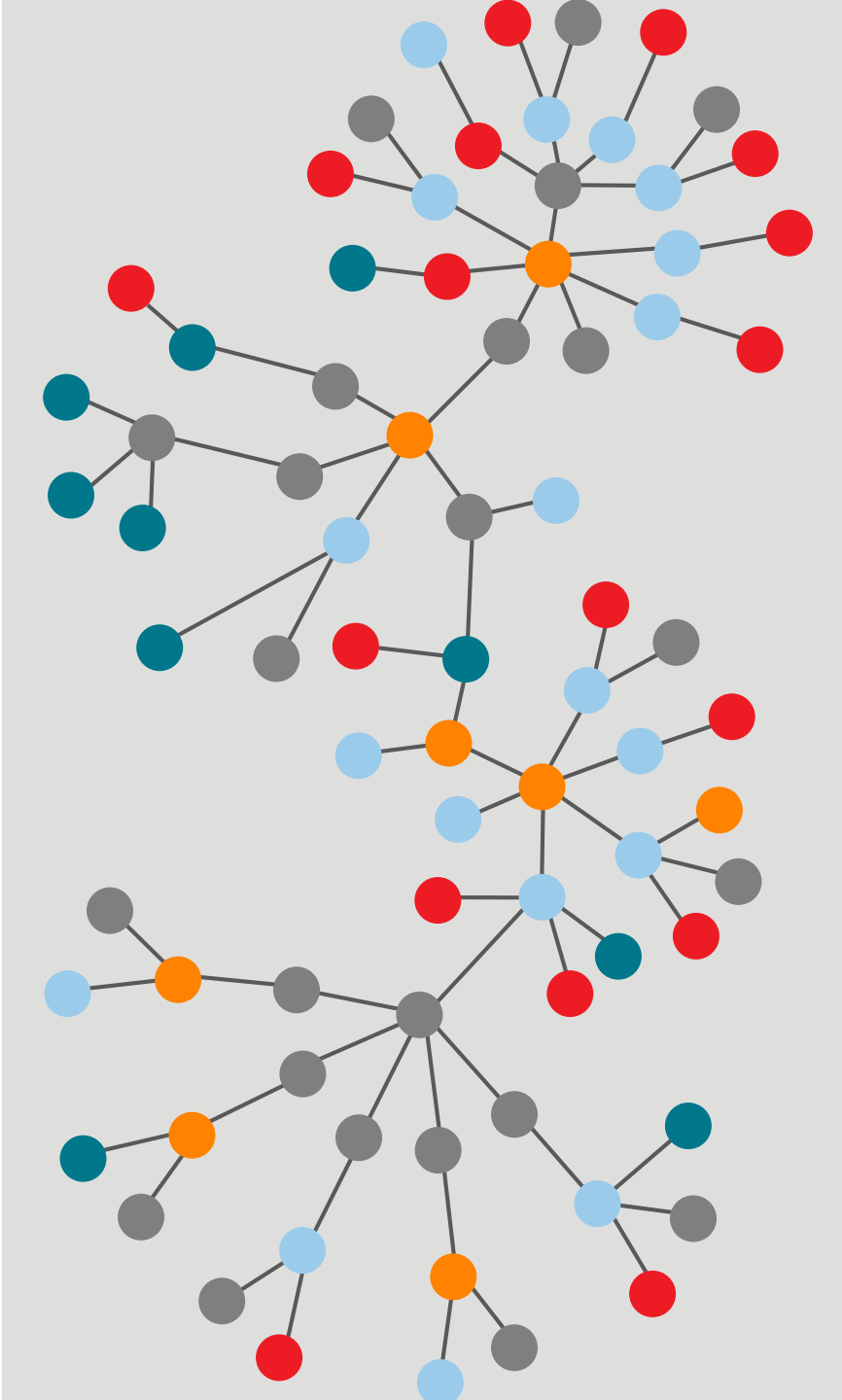
The ECL IDE Goes Multi-Language – Computer Languages that Is!

David de Hilster
Consulting Software Engineer
LexisNexis® Risk Solutions

LexisNexis® RISK SOLUTIONS

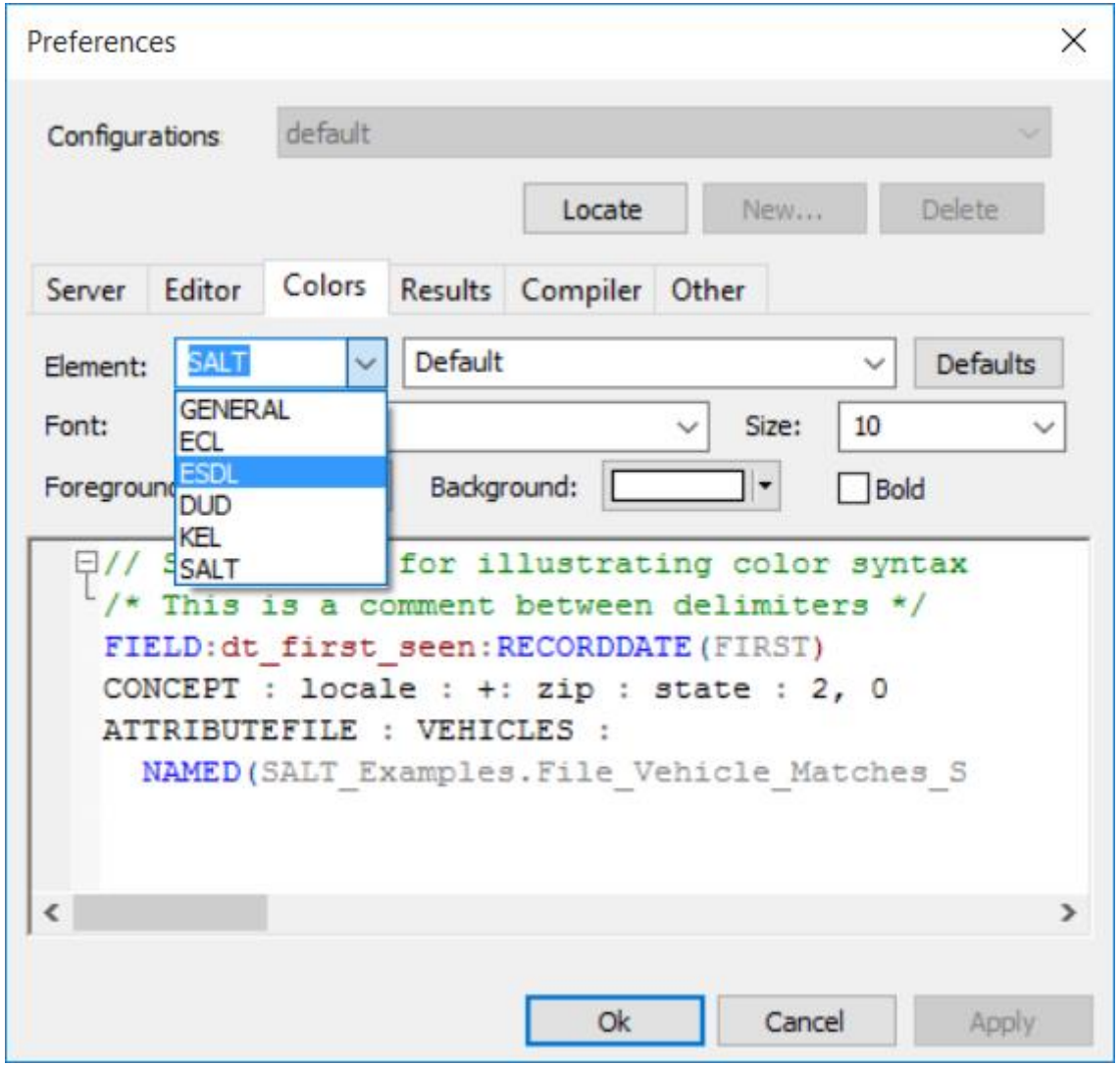
Quick poll: Which Interface do you use for ECL?

See poll on bottom of presentation screen



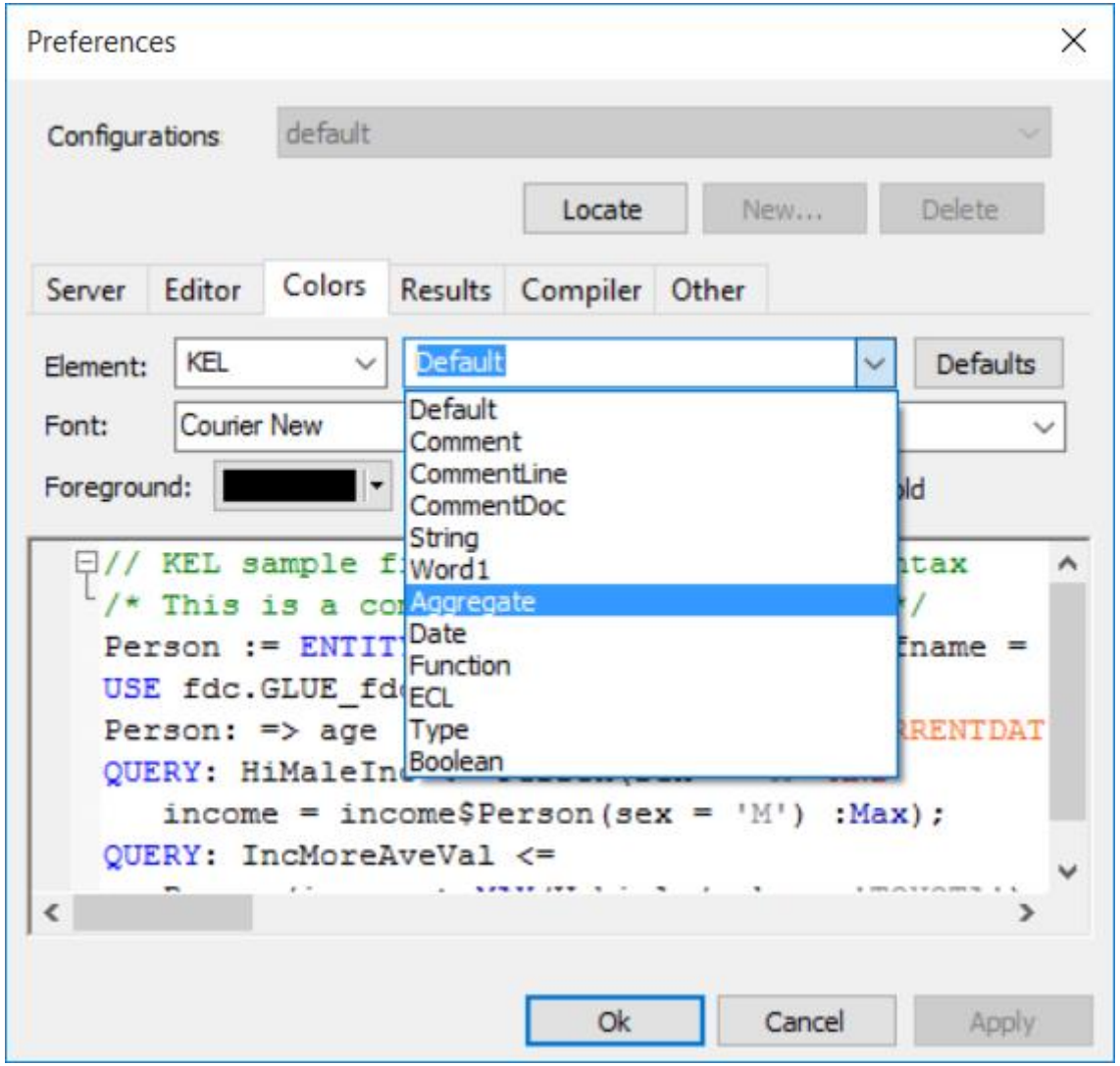
HPCC SYSTEMS®

Colorized Languages



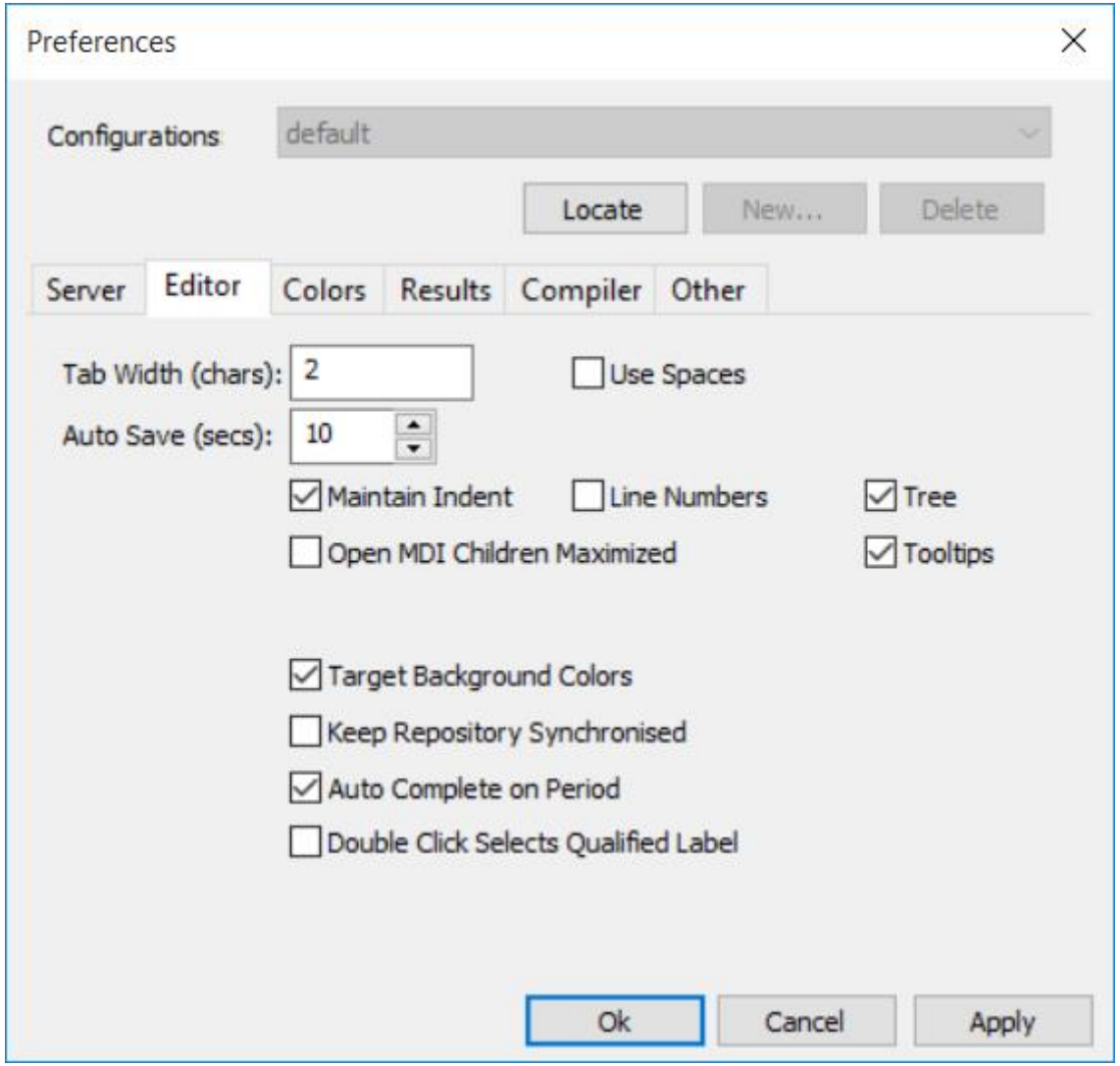
HPCC SYSTEMS®

Language Specific Element Colors



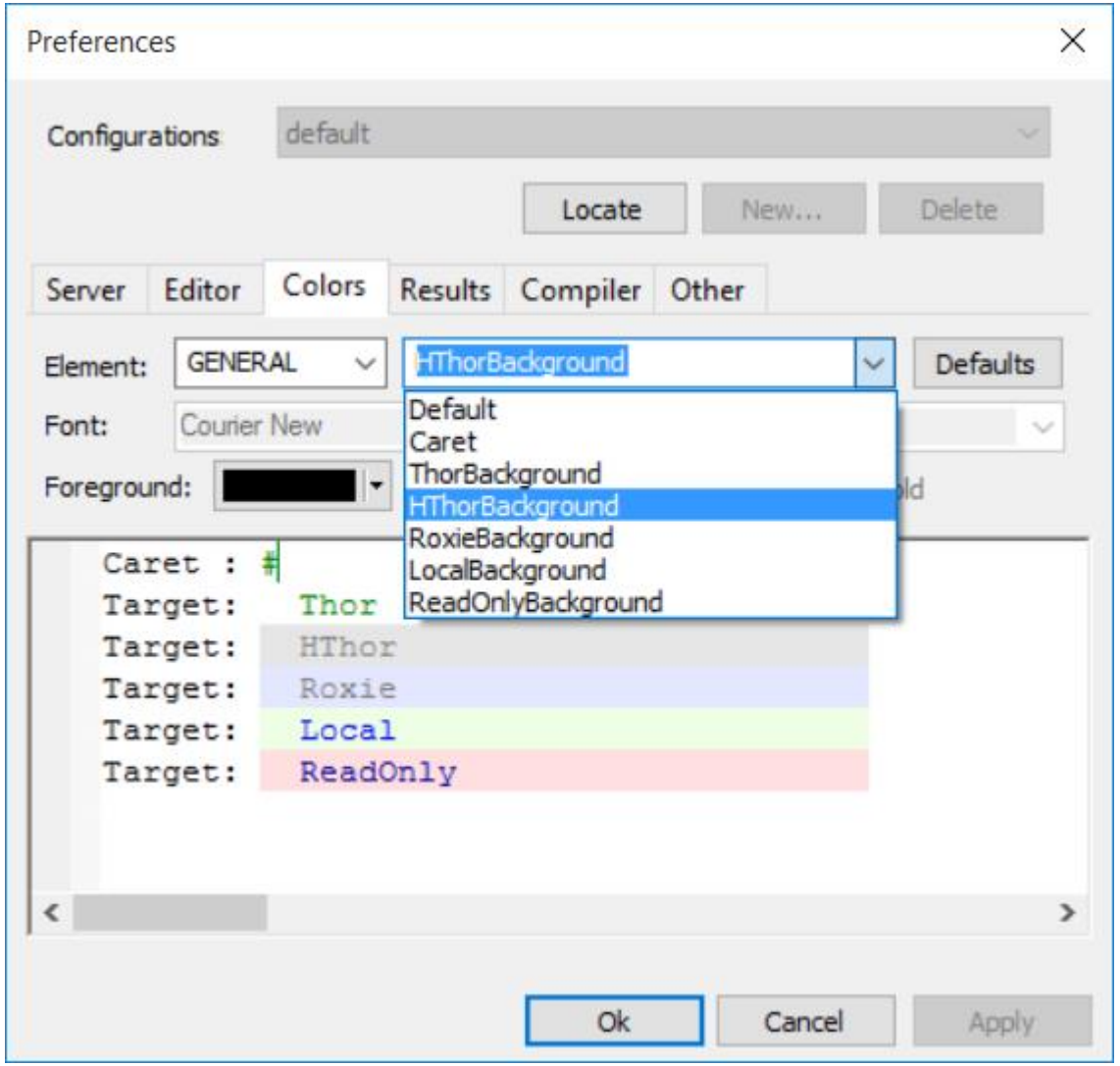
HPCC SYSTEMS®

Target Background Colors



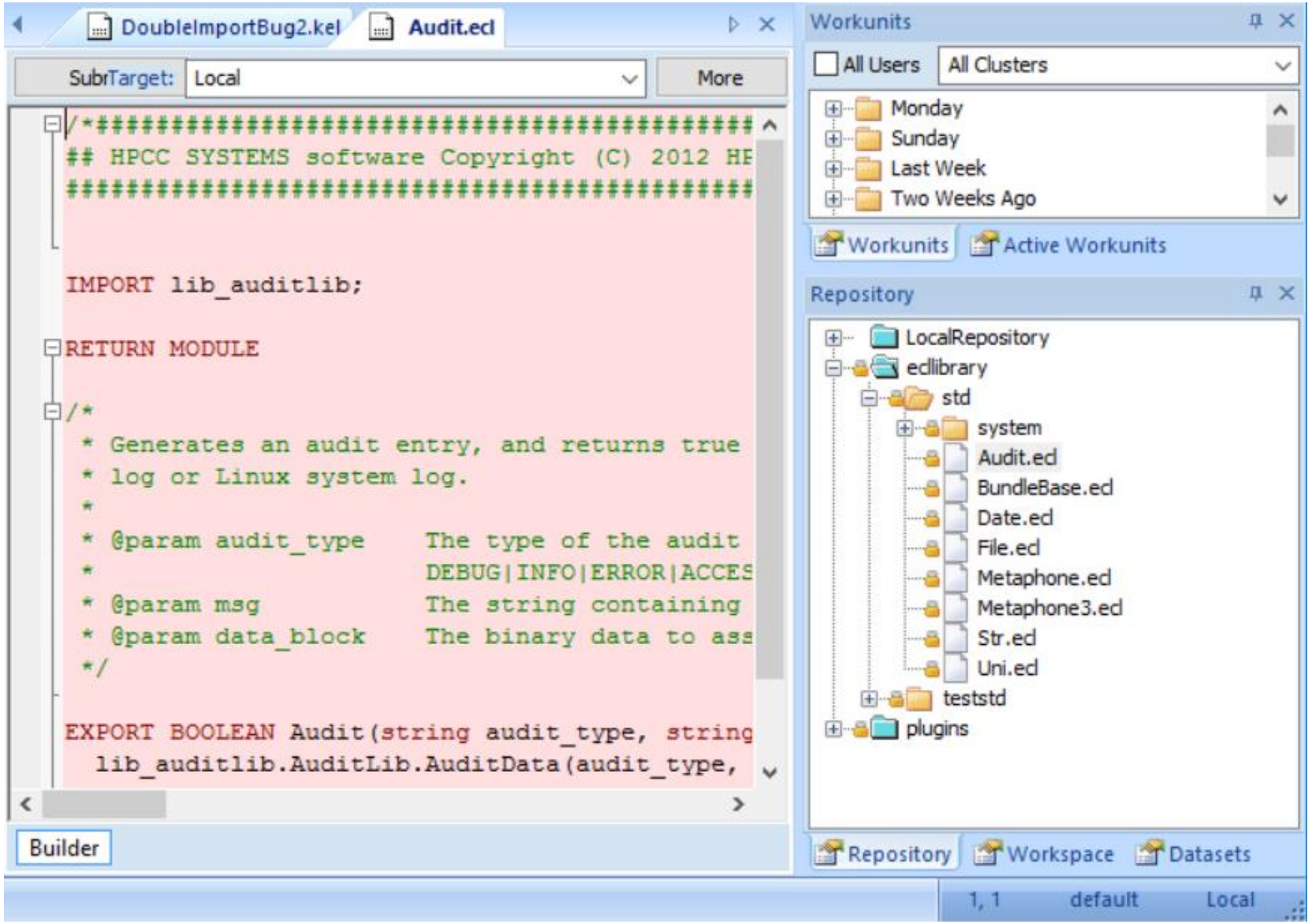
HPCC SYSTEMS®

General Colorization (Non-Language Specific)



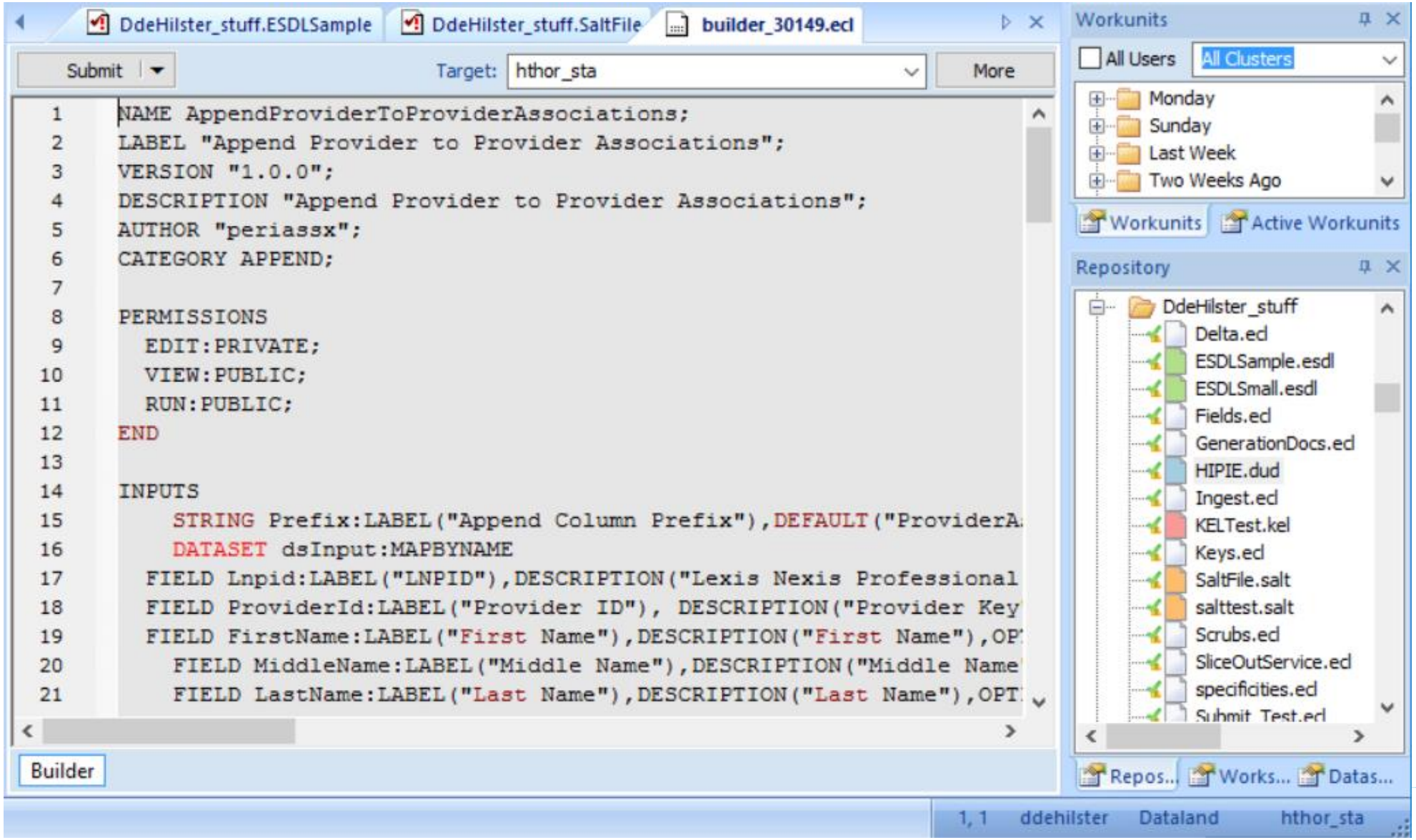
HPCC SYSTEMS®

Read Only Background



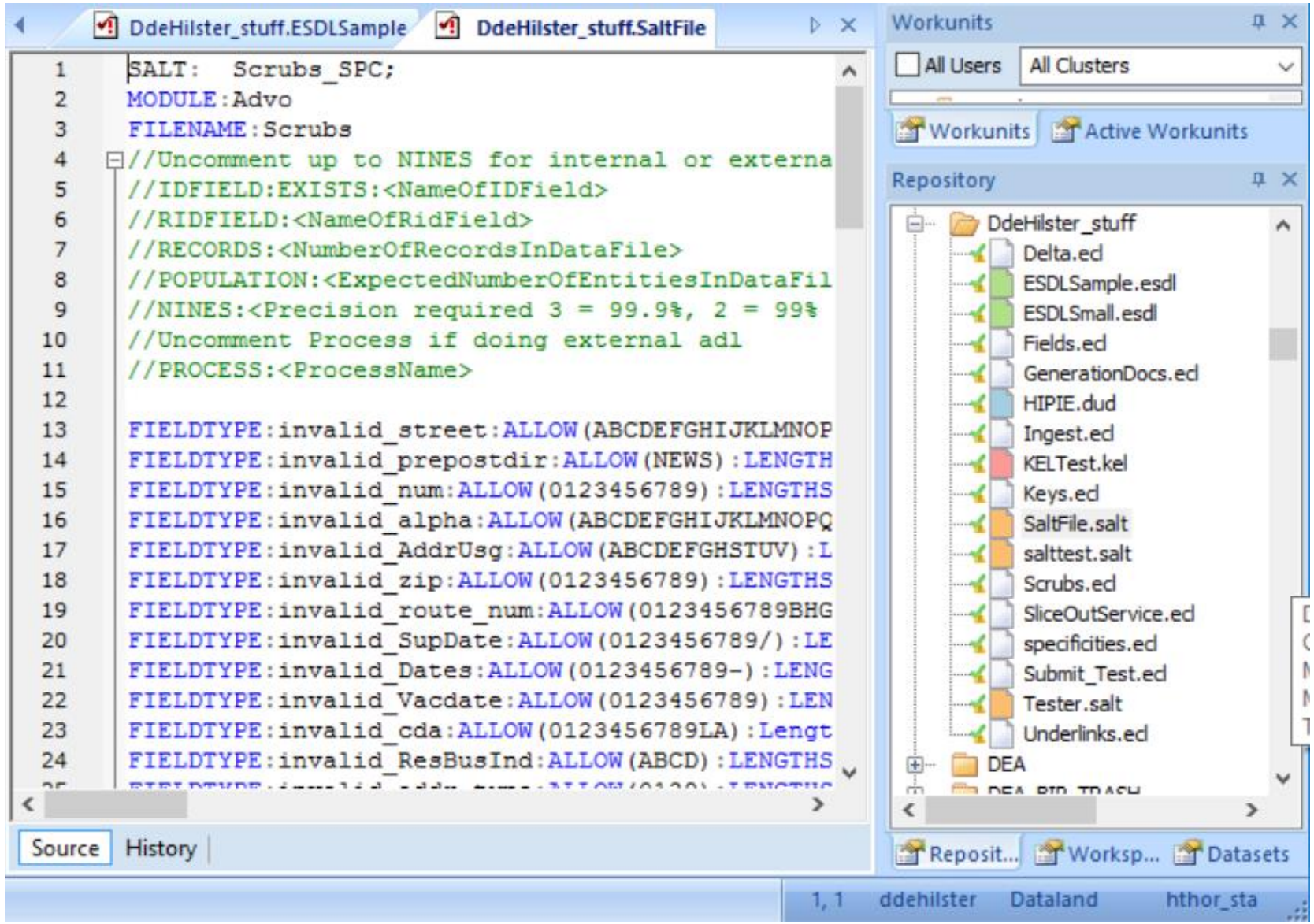
HPCC SYSTEMS®

hthor Background



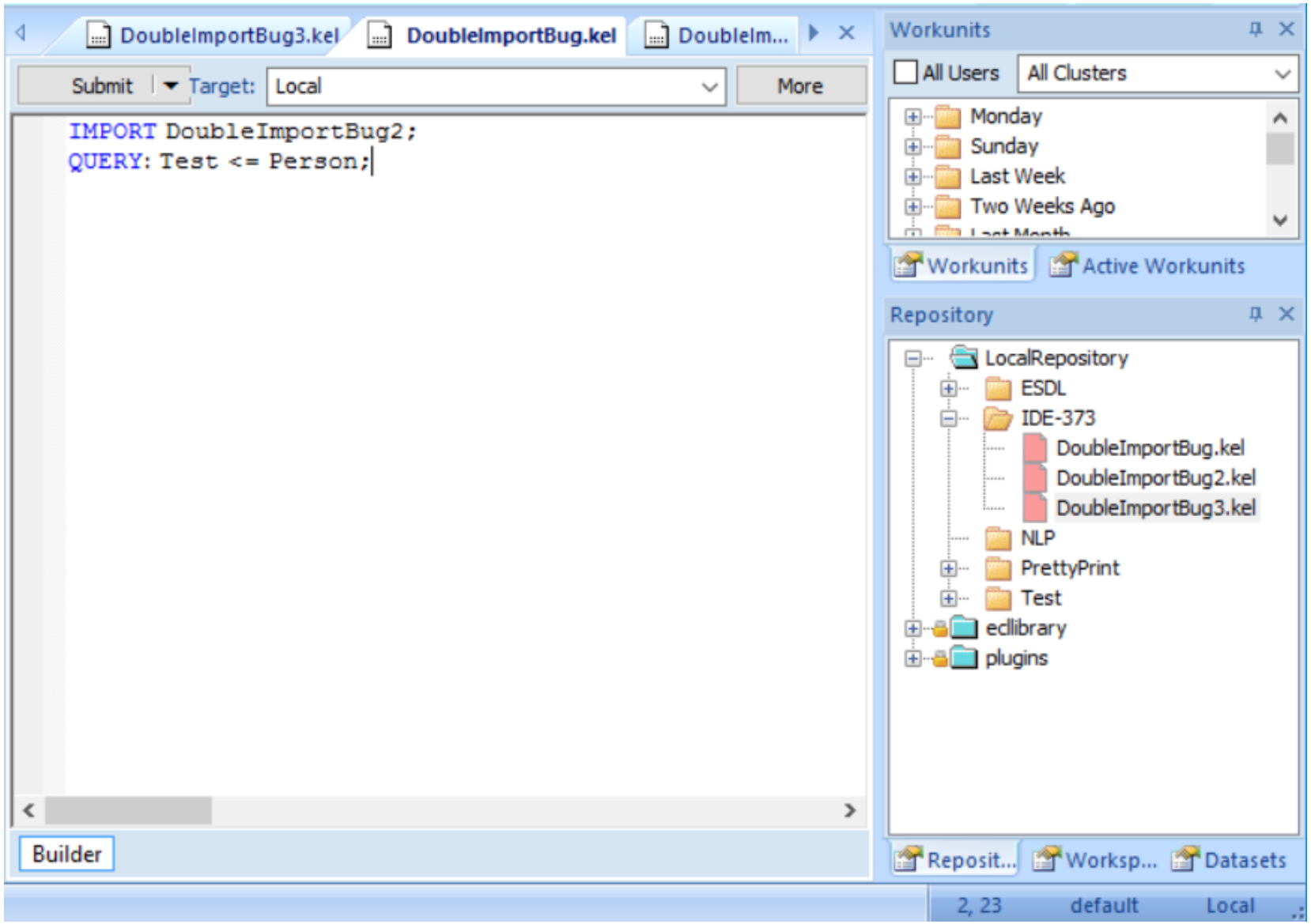
HPCC SYSTEMS®

File Color Coding



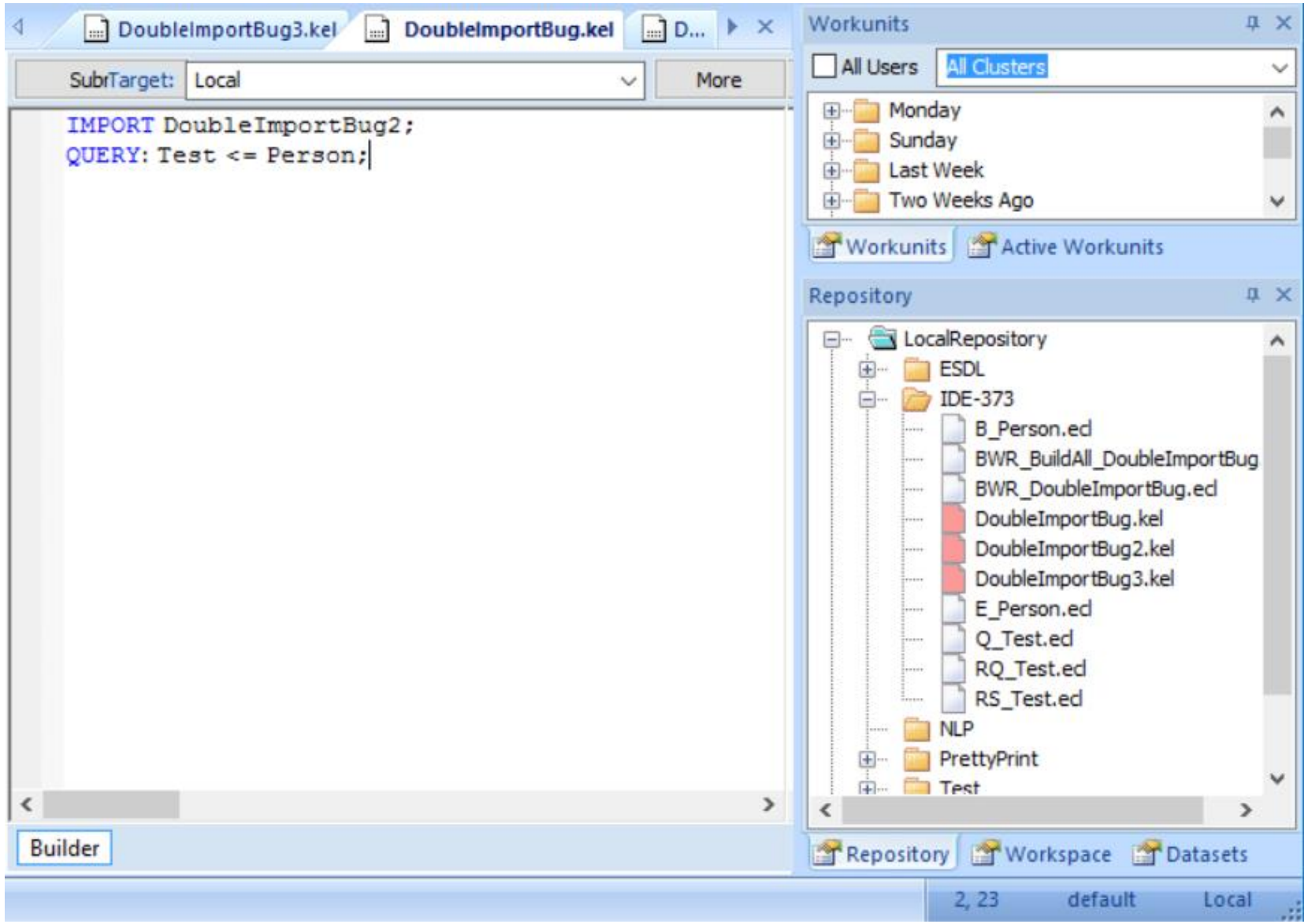
HPCC SYSTEMS®

Generating ECL for KEL



HPCC SYSTEMS®

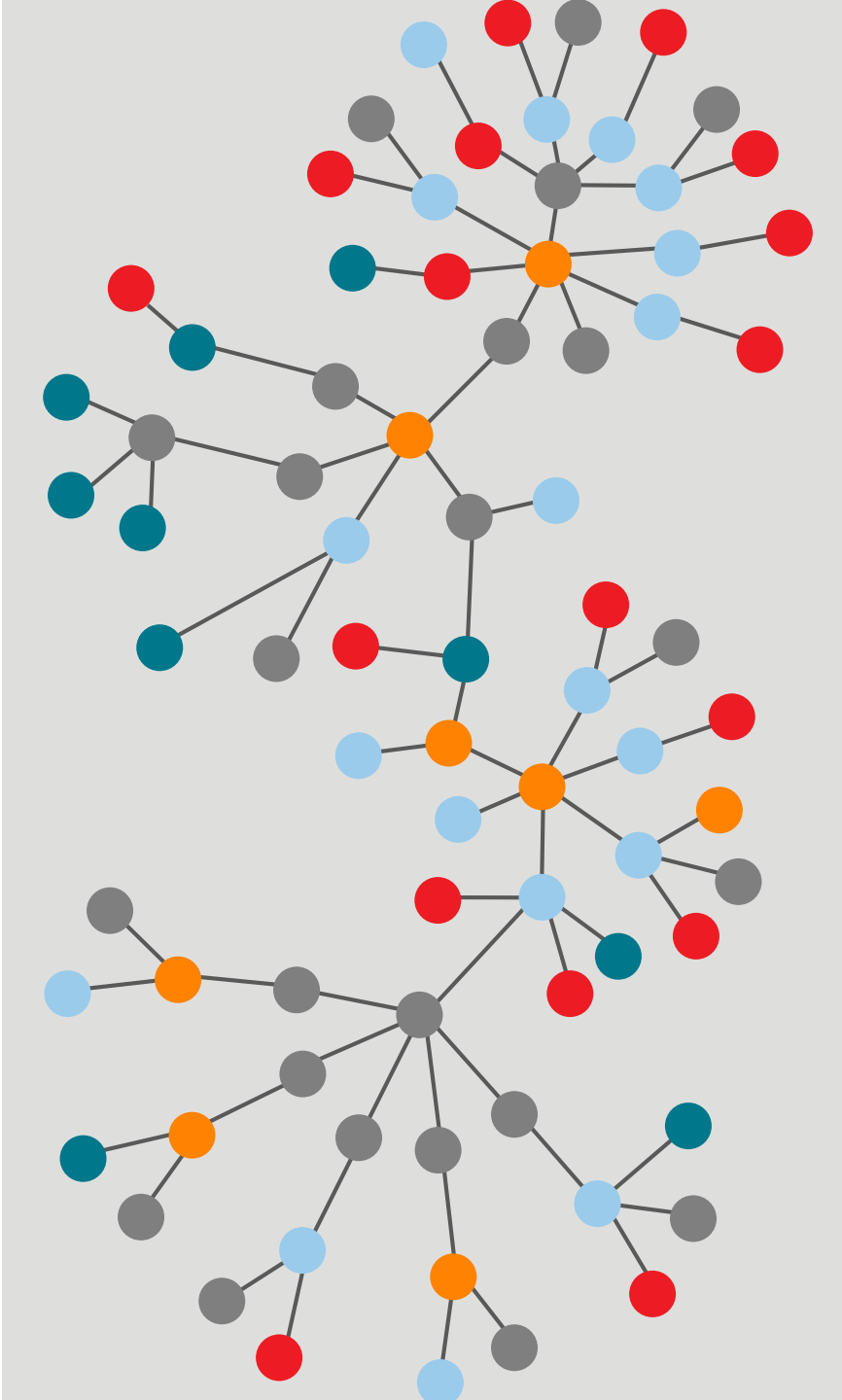
RESULT: Generating ECL for KEL



HPCC SYSTEMS®

Quick poll: What language other than ECL have you used with HPCC Systems? (choose the one you use most)

See poll on bottom of presentation screen



HPCC SYSTEMS®

Questions?

David de Hilster

Consulting Software Engineer

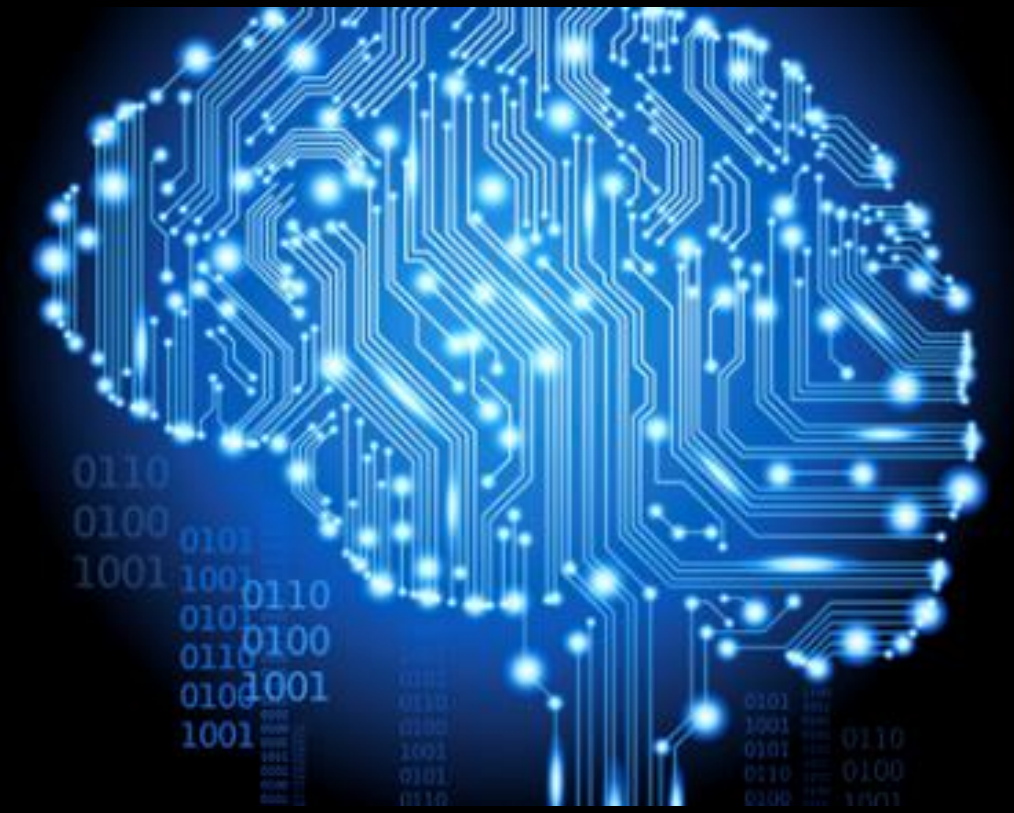
LexisNexis® Risk Solutions

david.dehilster@lexisnexisrisk.com

HPCC SYSTEMS®

THE DOWNLOAD

TECH TALKS BY HPCC SYSTEMS



HPCC SYSTEMS®

HPCC Systems - New Website Preview

Jessica Lorti
Director Marketing
LexisNexis Risk Solutions

LexisNexis® RISK SOLUTIONS

Questions?

Jessica Lorti
Director Marketing
LexisNexis® Risk Solutions
jessica.lorti@lexisnexisrisk.com

HPCC SYSTEMS®

Submit a Talk for an Upcoming Episode!

- Have a new success story to share?
- Want to pitch a new use case?
- Have a new HPCC Systems application you want to demo?
- Want to share some helpful ECL tips and sample code?
- Have a new suggestion for the roadmap?
- Be a featured speaker for an upcoming episode! Email your idea to Techtalks@hpccsystems.com

Save the Date for our next Tech Talk on May 25!

Visit The Download Tech Talks wiki for more information:
https://wiki.hpccsystems.com/display/hpcc/HPCC+Systems+Tech+Talks

HPCC SYSTEMS®

Thank You!

A copy of this presentation will be made available soon on our blog:
hpccsystems.com/blog