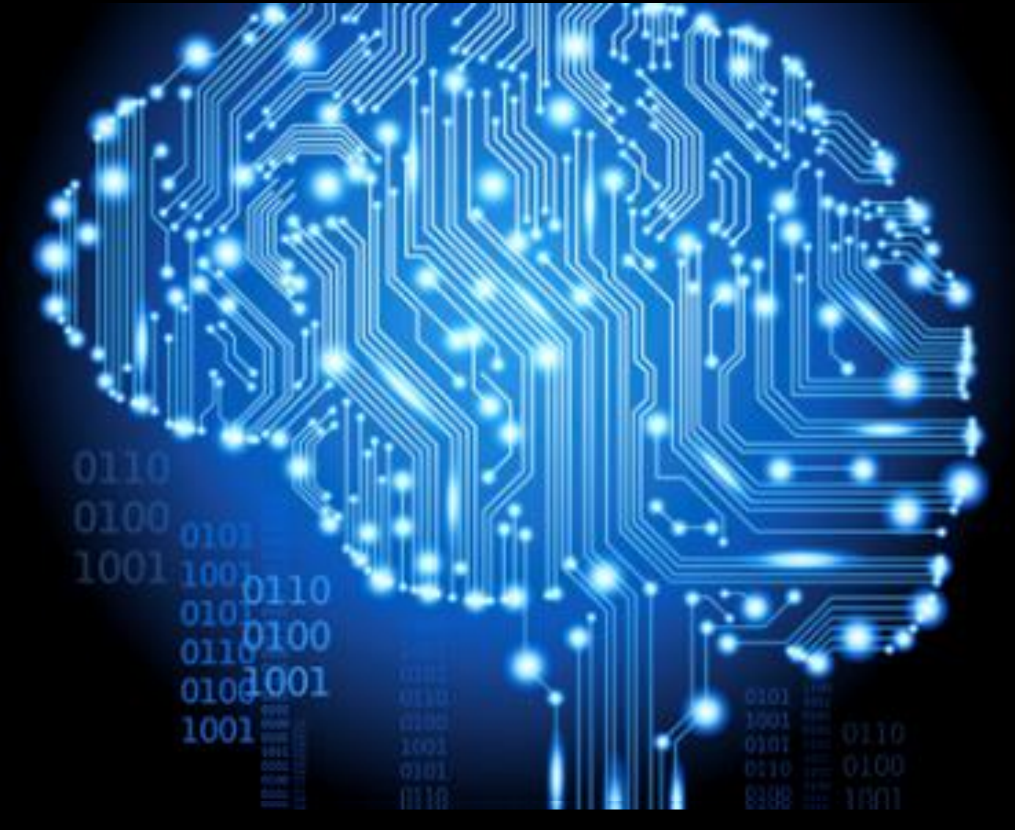
THE DOWNLOAD

TECH TALKS BY HPCC SYSTEMS



HPCC SYSTEMS®

The Download: Community Tech Talks Episode 7

September 14, 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

Community announcements

- HPCC Systems Platform updates
  - 6.4.0-1 gold and 6.4.2-rc2 are the latest builds available
  - WsSQL 6.4.0 and Wsclient 1.2 now released
  - New PBblas blog for matrix operations on HPCC Systems
  - **What's Coming in 7.0!**
    - Improved stats and improved metadata for smart editing
    - Faster compilation and fast syntax checking.
    - ECL IDE, Graphviewer and new ECL watch UI features
    - Remote projection/filtering of code
    - Spark integration
    - More machine learning bundles
    - Improvements in spraying area
    - Session management in ECL Watch
    - Improved Configuration Manager

Dr. Flavio Villanustre
VP Technology
RELX Distinguished Technologist
LexisNexis® Risk Solutions
Flavio.Villanustre@lexisnexisrisk.com

HPCC SYSTEMS®

Community announcements

- Reminder: Still time to register!
- 2017 HPCC Systems Community Day
  - October 3-4, 2017 in Atlanta, GA
    - Hands-on ECL Training Workshop on October 3
    - Community Day Sessions on October 4
    - Registration is open to the public to attend
  - Details at https://hpccsystems.com/hpccsummit2017
  - Thank you to our Sponsors!

Platinum Sponsor

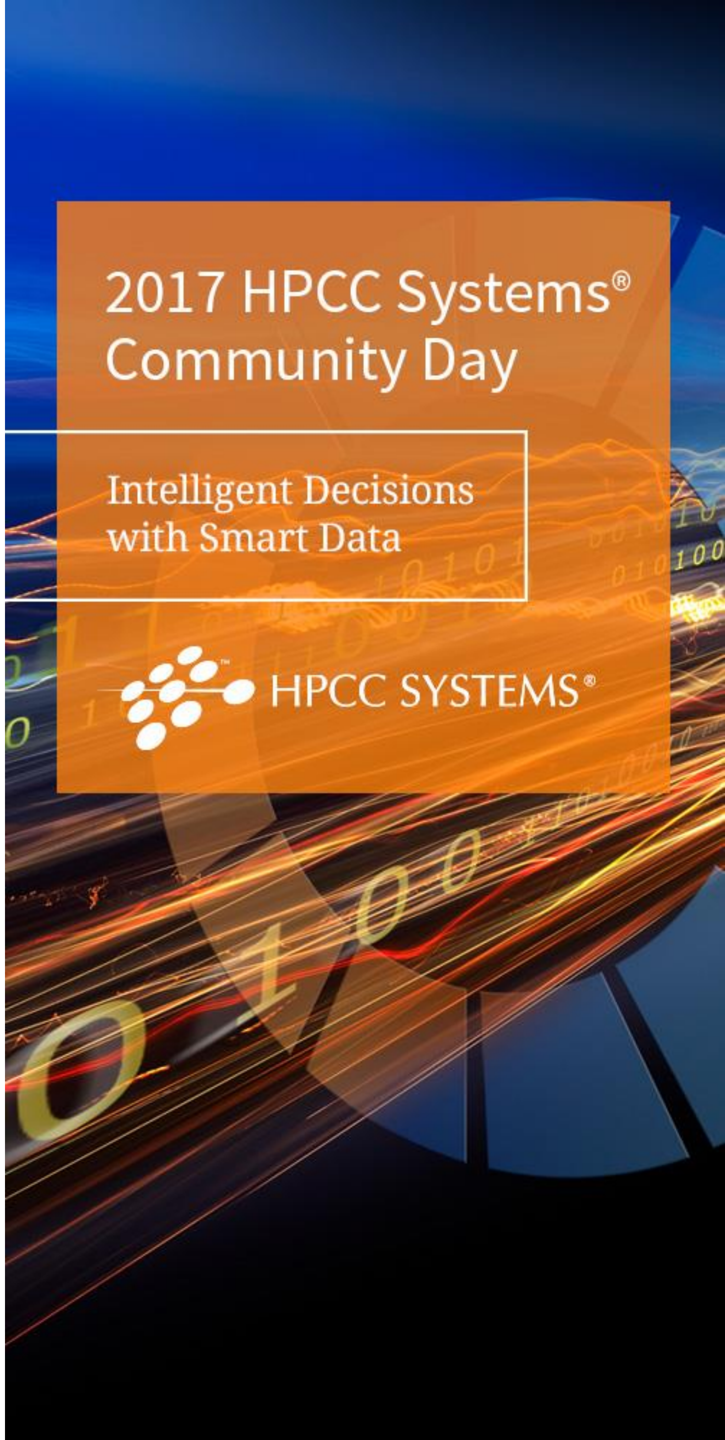
Cognizant®

Gold Sponsor

Infosys®

Staffing Sponsor

Datum SOFTWARE


2017 HPCC Systems® Community Day
Intelligent Decisions with Smart Data
HPCC SYSTEMS®

Community Day pre-event workshop

Mastering Your Big Data with ECL

This class is for attendees who want to understand the HPCC Systems platform and learn ECL to build powerful data queries. Anyone who needs a basic familiarity and learn best practices with ECL should attend. The one day class will take the student through the entire ETL cycle from Spray (Extract) to Transform (THOR) and finally to Load (ROXIE).

Topics include:

- **Part 1: Data Extraction and Transformation**
  - Quick overview of THOR cluster, and the parallel distributed data processing concept, setting up a cluster, ECL Watch overview, spraying data, ECL IDE, ECL language essentials, and more…
- **Part 2: Prepare the Data Search Engine**
  - Defining and building an INDEX, getting single and batch results, data indexing, filtering and normalization, searching, and more…
- **Part 3: Write and Publish ROXIE query**
  - Call Search, Implicit function, publish in ECL Watch, test in WS-ECL, and more…

What:
Mastering Your Big Data w/ ECL

When:
Tuesday October 3, 9am – 4pm

Where:
Ritz Carlton Buckhead, Atlanta, Ga

Register:
hpccsummit2017.eventbrite.com

Community Day agenda

Wednesday, October 4, 2017

The agenda will run from 8:30am – 5:00pm ET. We have a fantastic line-up of speakers featuring industry experts, academia and thought leaders. Check it out!

Time	Topic
7:00am – 8:30am	Registration and Breakfast
8:30am – 9:15am	Welcome and Sponsor Keynotes from Cognizant and Infosys
9:15am – 10:30am	Track 1: HPCC Systems in Industry: Real World Use Cases Featuring DataSeers, 3LOQ, Couchbase,
10:30am – 10:45am	Break - Poster Presentations, Robotics Display & Exhibits
10:45am - 12:00pm	Track 2: HPCC Systems in Academia: Beyond the Classroom Featuring Humboldt University Berlin and North Carolina State University
12:00pm - 12:45pm	Lunch - Poster Presentations and Robotics Display
12:45 – 1:00pm	Community Awards Ceremony
1:00pm – 2:00pm	Panel Discussion: Integrated Scientific Discovery
2:00pm - 3:15pm	Track 3: HPCC Systems in the Limelight: Success Across RELX Group Featuring LexisNexis Risk Solutions, Reed Business Information and Reed Exhibitions
3:15pm - 3:30pm	Break - Poster Presentations, Robotics Display & Exhibits
3:30pm - 4:50pm	Track 4: HPCC Systems Roadmap Tech Talks Featuring topics on the Platform Roadmap, Visualization, Machine Learning and Architecture Improvements
4:50pm - 5:00pm	Closing Words & Adjourn

Register today at hpccsummit2017.eventbrite.com

Today's speakers

Xiaoming Wang (Ming)

Consultant Software Engineer

LexisNexis Risk Solutions

Xiaoming.Wang@lexisnexisrisk.com

Xiaoming Wang (Ming), joined LexisNexis in 2013 on the HPCC Systems core platform team. His main responsibilities include working on the HPCC Systems Platform products builds, deployment and configuration tools and deployment solutions including AWS AMI/Instant Cloud, Juju Charm, and HPCC Systems, and more.

Bob Foreman

Senior Software Engineer

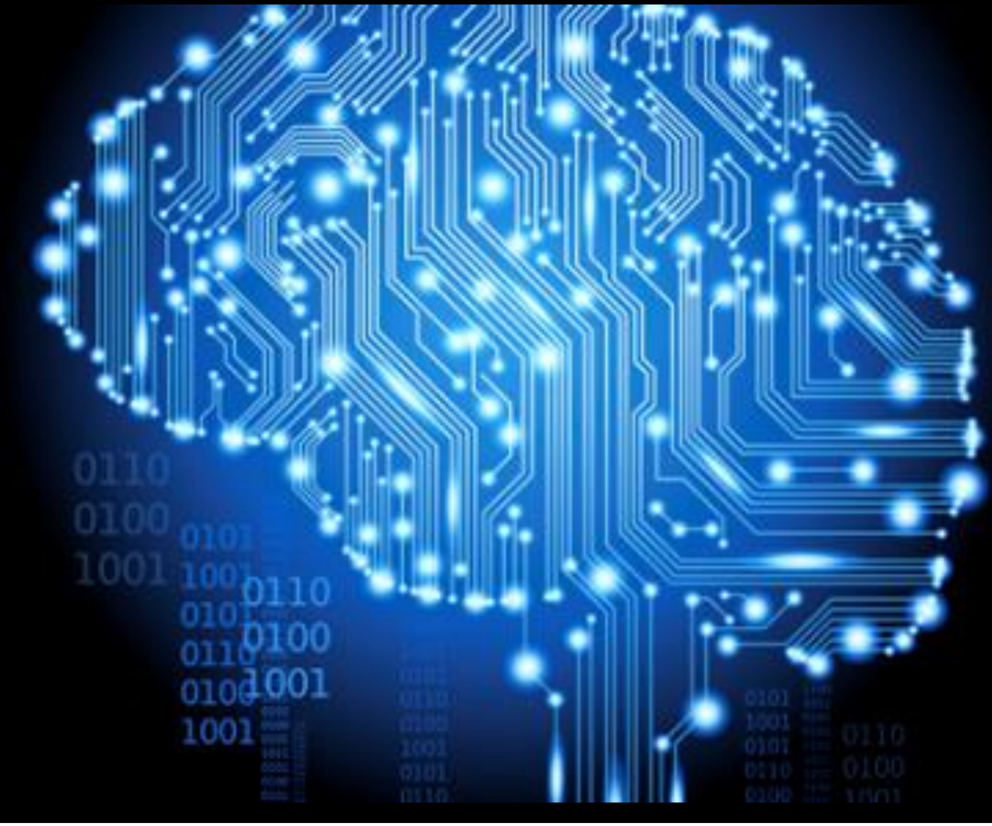
LexisNexis Risk Solutions

Robert.Foreman@lexisnexisrisk.com

Bob Foreman has worked with the HPCC Systems technology platform and the ECL programming language for over 5 years, and has been a technical trainer for over 25 years. He is the developer and designer of the HPCC Systems Online Training Courses, and is the Senior Instructor for all classroom and Webex/Lync based training.

THE DOWNLOAD

TECH TALKS BY HPCC SYSTEMS



Initial HPCC Systems integration with Jupyter Notebook

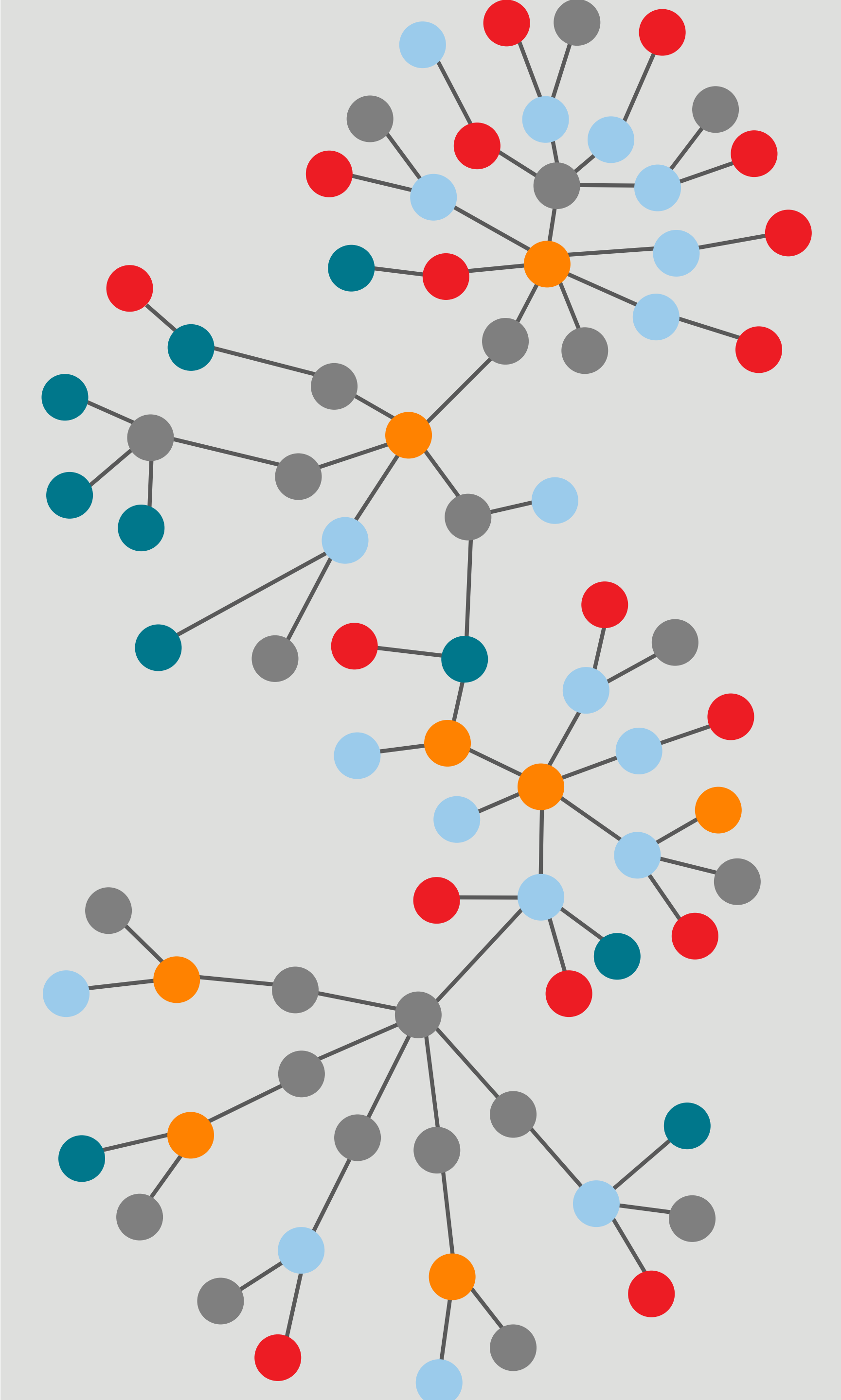
HPCC SYSTEMS®

Xiaoming Wang (Ming)
Consulting Software Engineer
LexisNexis Risk Solutions

LexisNexis®
RISK SOLUTIONS

Quick poll:
Have you used Jupyter?

See poll on bottom of presentation screen



HPCC SYSTEMS®

Introduction

The Jupyter Notebook is an open-source web application that allows you to create and share documents that contain live code, equations, visualizations and explanatory text. Uses include: data cleaning and transformation, numerical simulation, statistical modeling, machine learning and much more.

Main features:

- In-browser editing for code
- The ability to execute code from the browser
- Displaying the result of computation using rich media representations, such as HTML, LaTeX, PNG, SVG, etc
- In-browser editing for rich text using the Markdown markup language
- Download as various document format such as HTML, Markdown, PDF, etc.

Integration with HPCC Systems ECL

- Jupyter Notebook kernel provides programming language support in Jupyter. IPython is the default kernel. Other kernels:

  https://github.com/jupyter/jupyter/wiki/Jupyter-kernels

- New kernel “HPCC ECL – TSECL” to support ECL. “TSECL” refers to ECL kernel implemented with TypeScript (a superset of JavaScript which primarily provides optional static typing, classes and interfaces).

- TSECL implementation references ITypeScript kernel and includes:
  1. jpecl-kernel modify from jp-kernel
  2. Utilize hpcc-js/comms package developed by Gordon Smith. It communicates with esp to submit ecl code and get back workunit result.

- Github:

  tsecl: https://github.com/xwang2713/itsecl

  Jpecl-kernel: https://github.com/xwang2713/jpecl-kernel

Current TSECL features and limitations

- Features
  - Configure ESP connection parameters, such as ip, port, cluster type, etc and test the connection
  - Run ECL code
  - Display ESP configuration
  - Run JavaScript/TypeScript code

- Limitations
  - Only first workunit displayed
  - No ECL syntax check
  - ……

Installation

- Virtualbox image:

  http://10.240.32.242/data3/jupyter/jupyter-hpcc-vm-0.1.1.ova

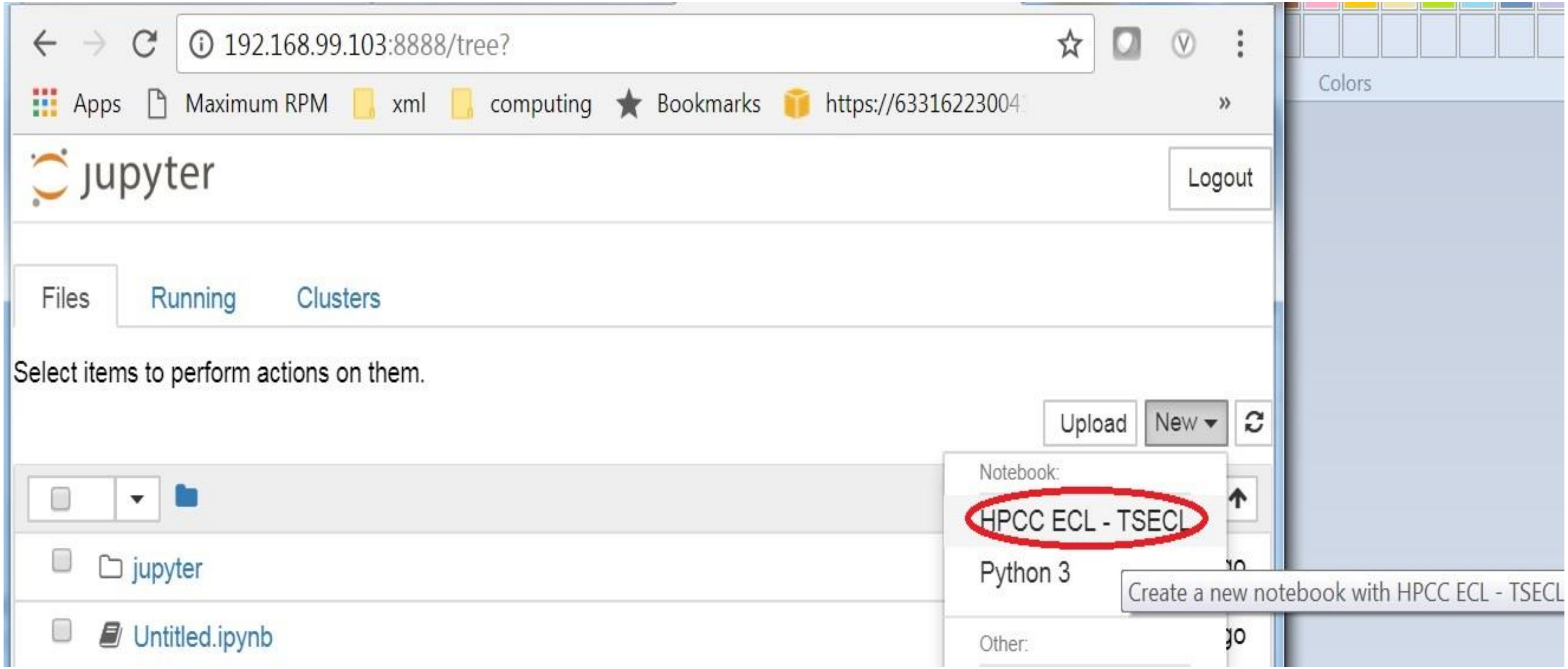
- Docker image:

  hpccsystems/jupyter:itsecl-0.1.1

  https://hub.docker.com/r/hpccsystems/jupyter/

- Manually install prerequisites and TSECL following instruction from github: https://github.com/xwang2713/itsecl/blob/master/README.md

Demo – Select TSECL kernel



192.168.99.103:8888/tree?

Apps Maximum RPM xml computing Bookmarks https://63316223004

jupyter

Logout

Files Running Clusters

Select items to perform actions on them.

Upload New

Notebook:

HPCC ECL - TSECL

Python 3

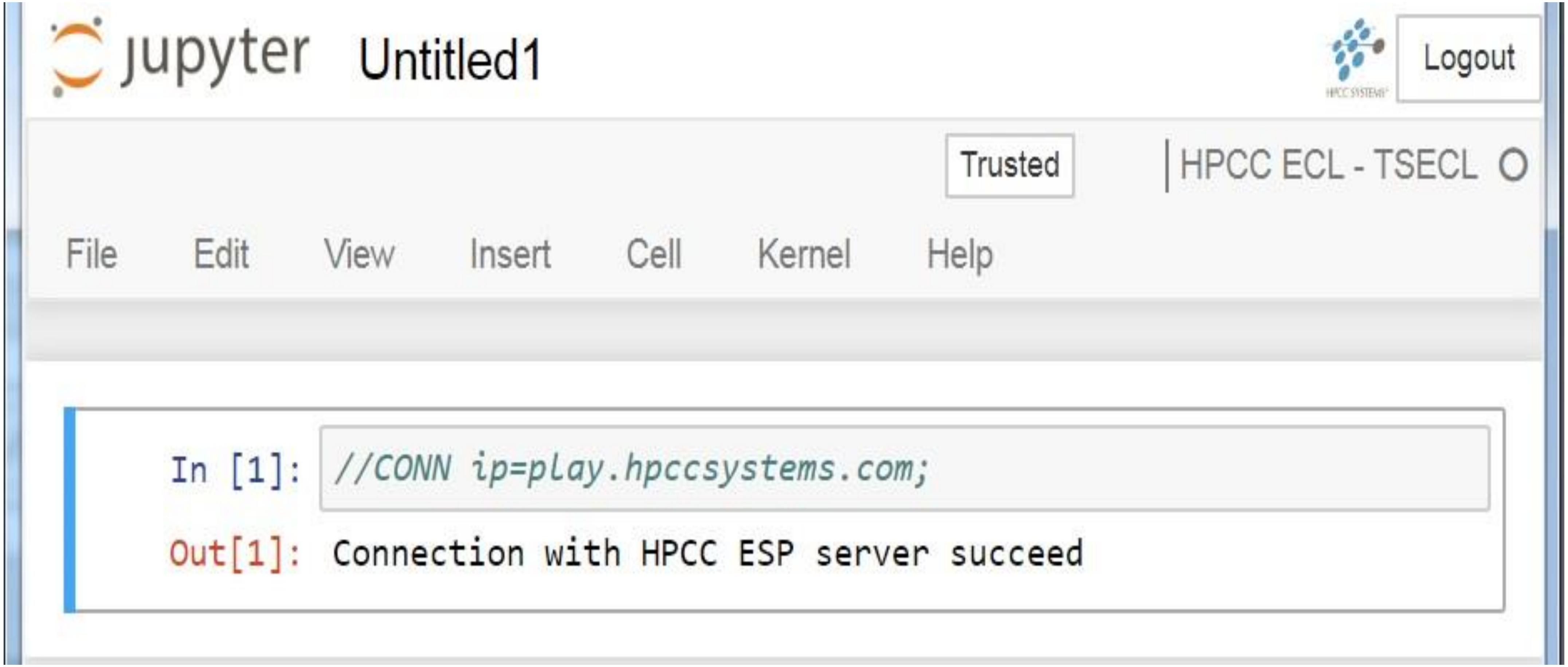
Create a new notebook with HPCC ECL - TSECL

Other:

jupyter

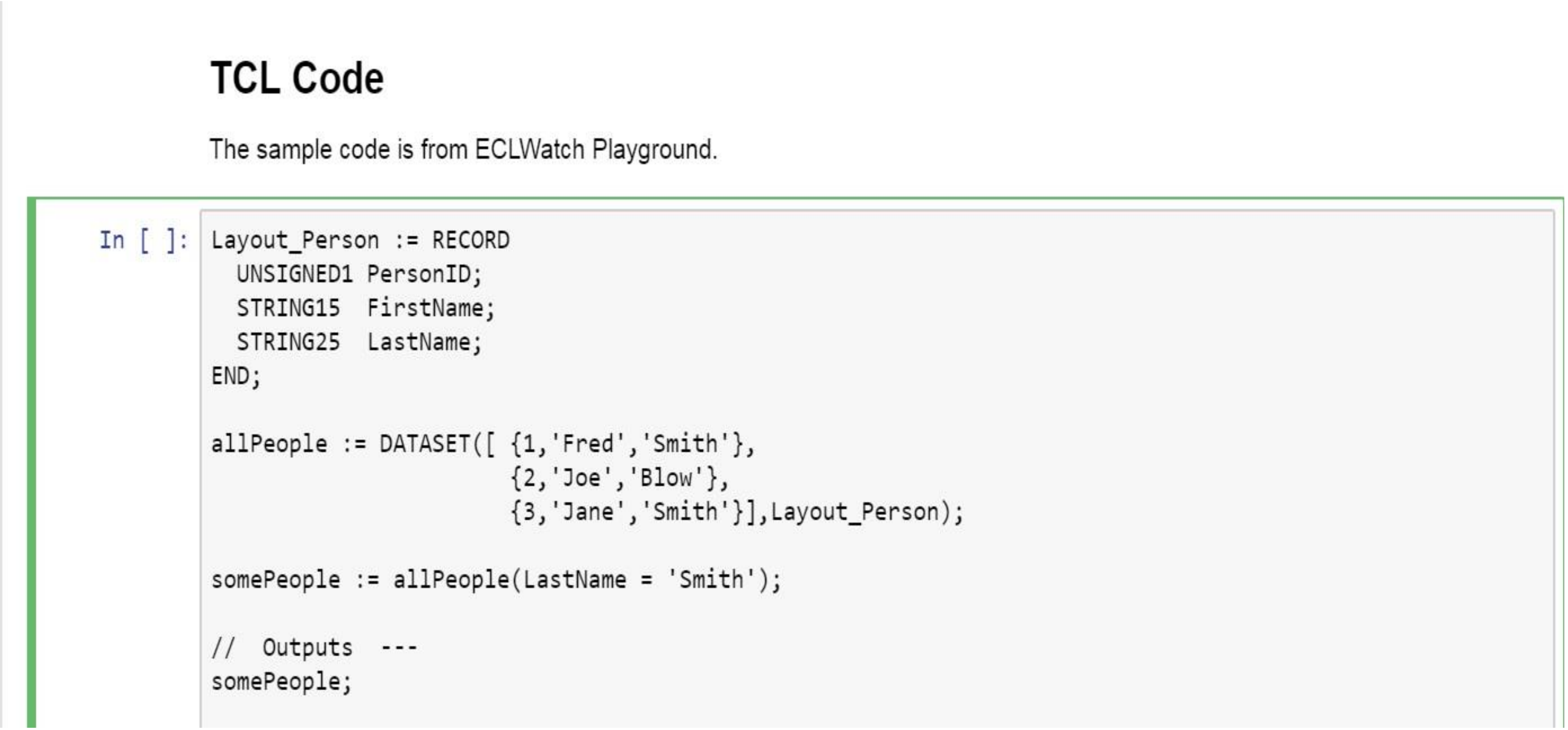
Untitled.ipynb

Demo - Test ESP connection



```
In [1]: //CONN ip=play.hpccsystems.com;
Out[1]: Connection with HPCC ESP server succeed
```

Demo – Write ECL code



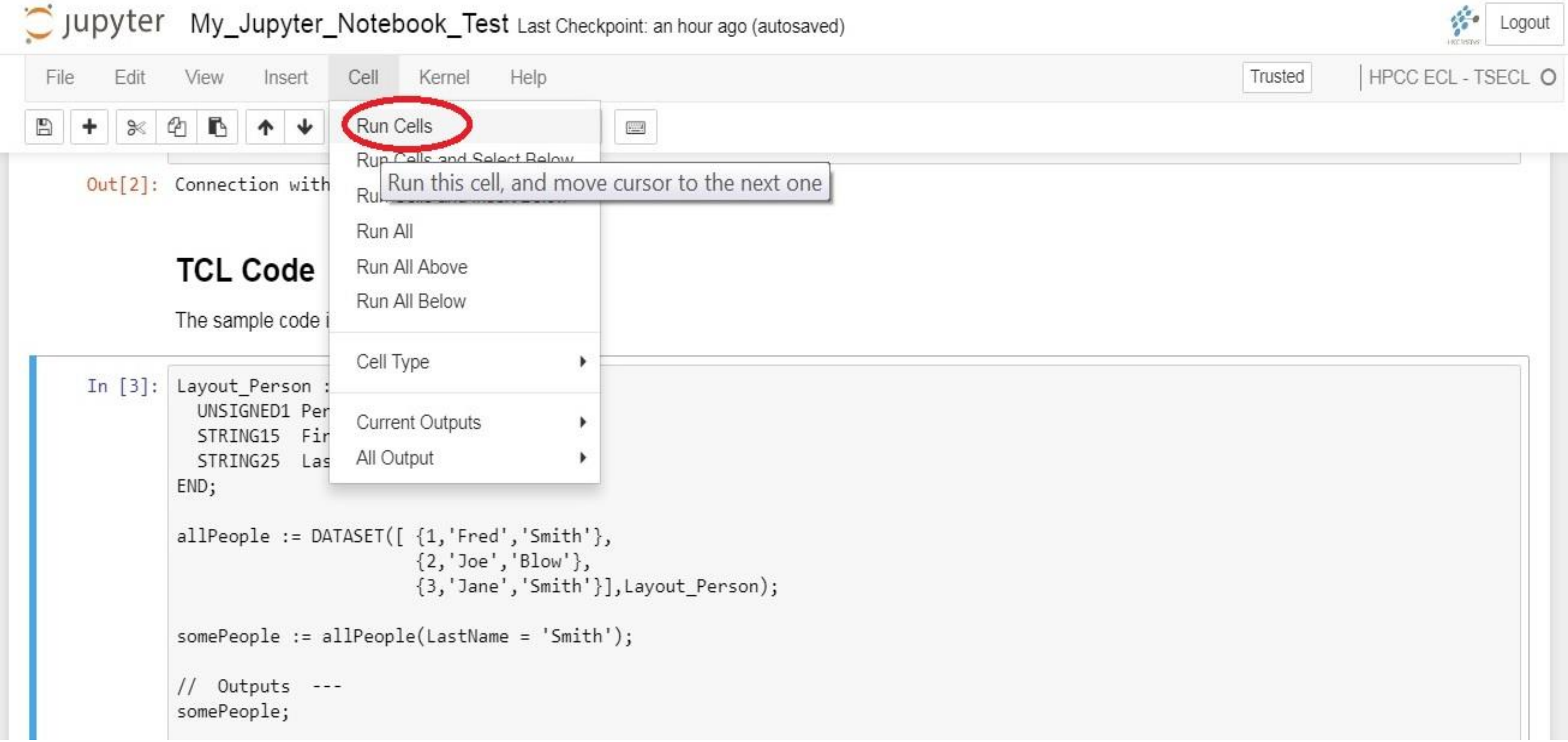
TCL Code

The sample code is from ECLWatch Playground.

In [ ]:

```
Layout_Person := RECORD
  UNSIGNED1 PersonID;
  STRING15  FirstName;
  STRING25  LastName;
END;

allPeople := DATASET([ {1,'Fred','Smith'},
                       {2,'Joe','Blow'},
                       {3,'Jane','Smith'}],Layout_Person);

somePeople := allPeople(LastName = 'Smith');

//  Outputs  ---
somePeople;
```

Demo – Run ECL code


jupyter
My_Jupyter_Notebook_Test Last Checkpoint: an hour ago (autosaved)
Logout
File
Edit
View
Insert
Cell
Kernel
Help
Trusted
HPCC ECL - TSECL
Run Cells
Run Cells and Select Below
Run Cells and Insert Below
Run this cell, and move cursor to the next one
Run All
Run All Above
Run All Below
Cell Type
Current Outputs
All Output

```
Out[2]: Connection with
```

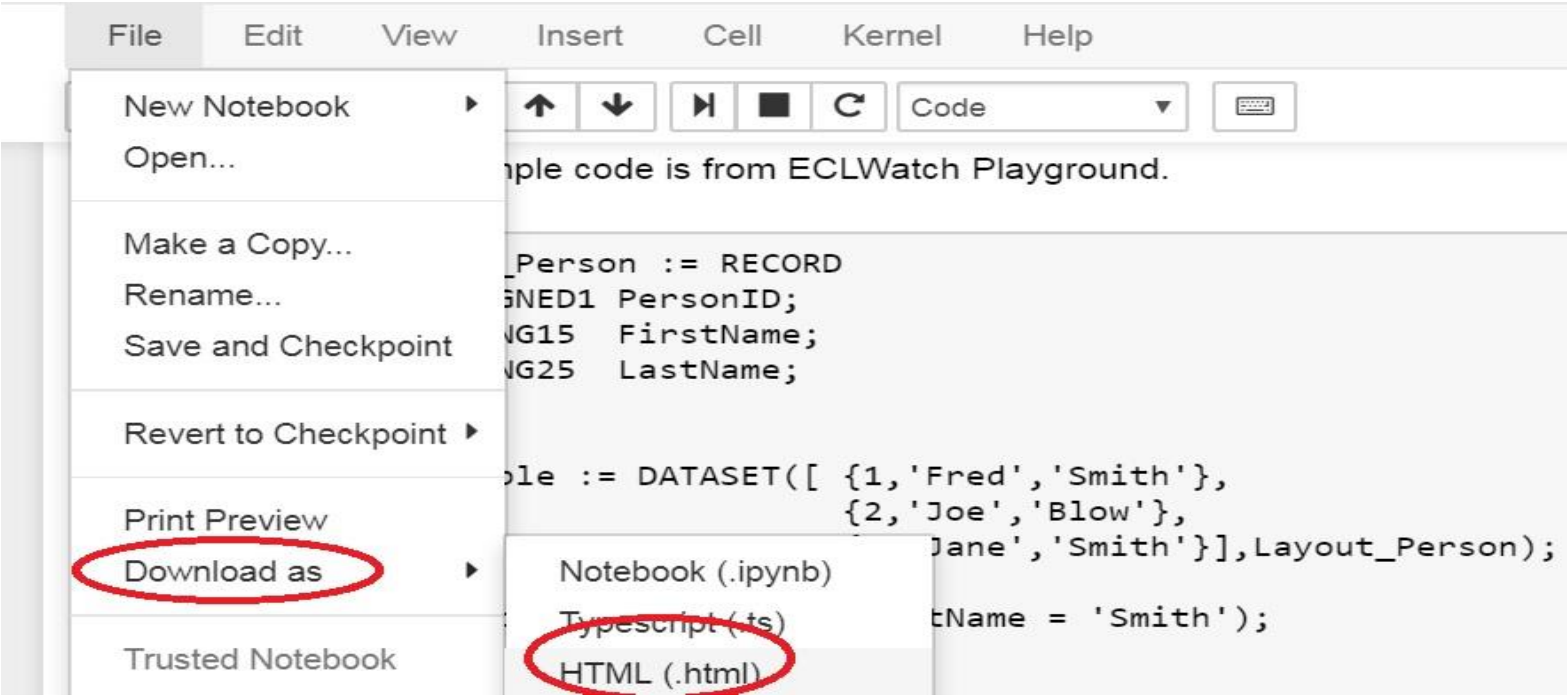
TCL Code

The sample code i

```
In [3]: Layout_Person :
          UNSIGNED1 Per
          STRING15  Fir
          STRING25  Las
        END;

        allPeople := DATASET([ {1,'Fred','Smith'},
                               {2,'Joe','Blow'},
                               {3,'Jane','Smith'}],Layout_Person);

        somePeople := allPeople(LastName = 'Smith');

        //  Outputs  ---
        somePeople;
```

HPCC SYSTEMS®

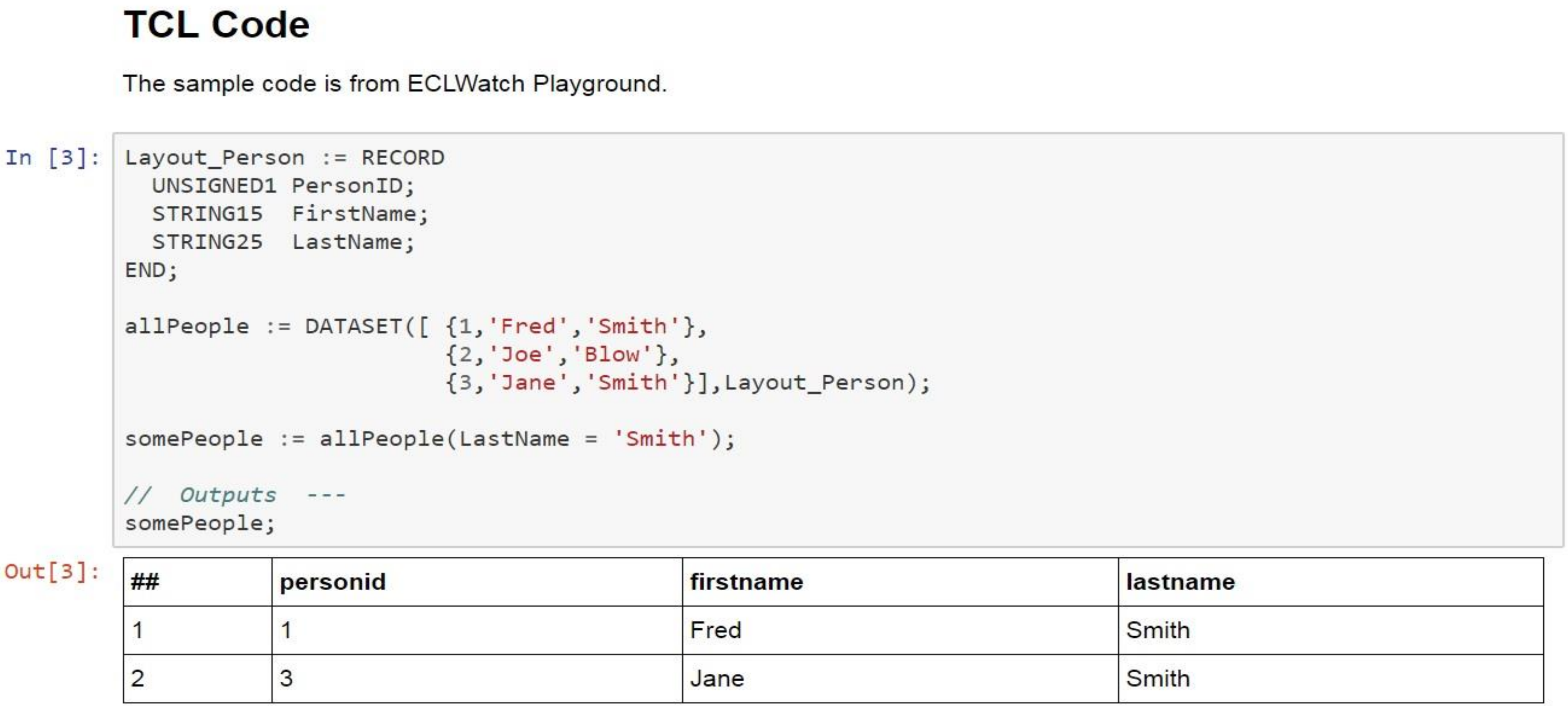
Demo – Output to HTML



ple code is from ECLWatch Playground.

```
_Person := RECORD
GNED1 PersonID;
NG15  FirstName;
NG25  LastName;


ple := DATASET([ {1,'Fred','Smith'},
                 {2,'Joe','Blow'},
                    Jane','Smith'}],Layout_Person);

                    tName = 'Smith');
```

Demo – HTML output



TCL Code

The sample code is from ECLWatch Playground.

In [3]:

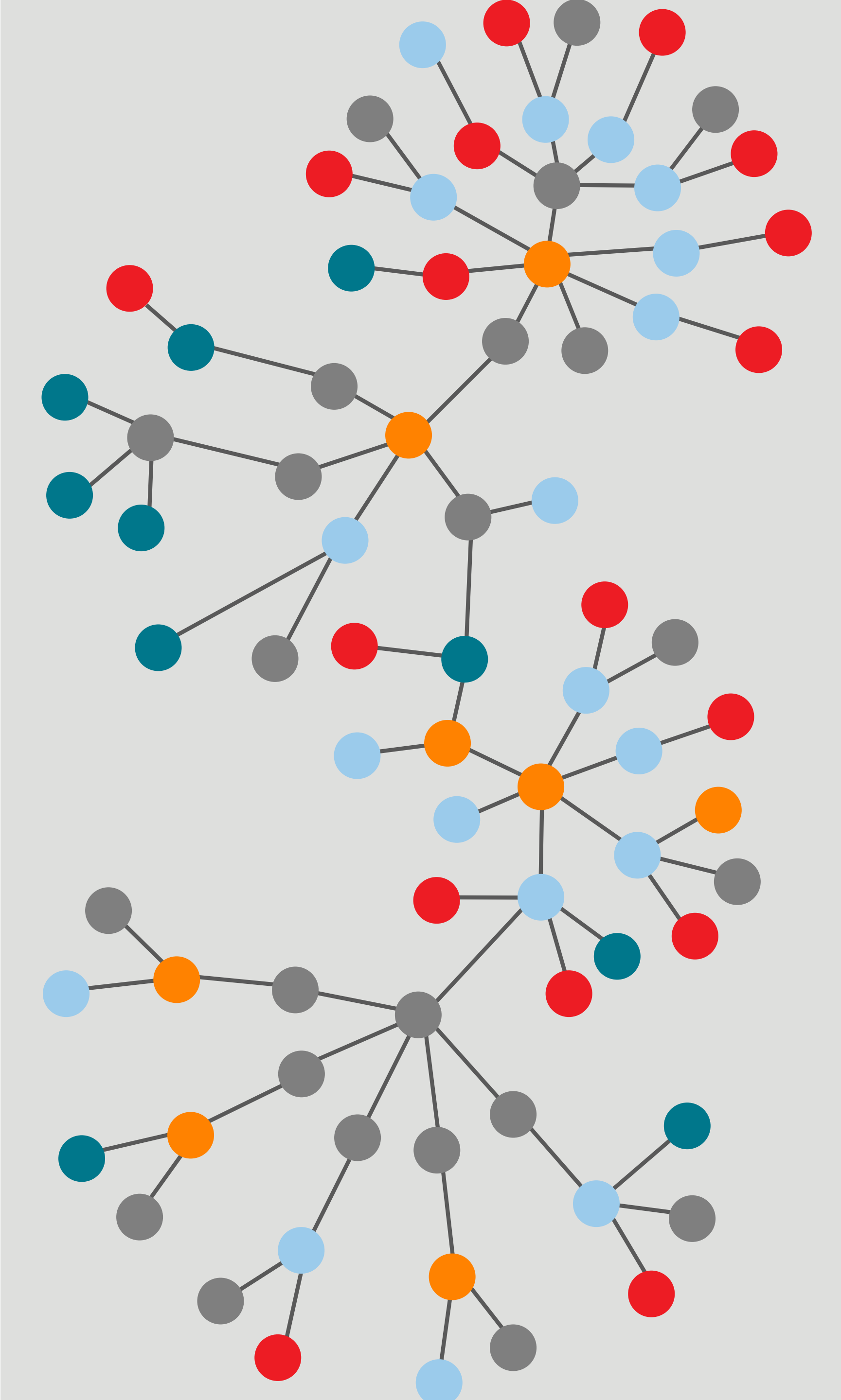
```
Layout_Person := RECORD
  UNSIGNED1 PersonID;
  STRING15  FirstName;
  STRING25  LastName;
END;

allPeople := DATASET([ {1,'Fred','Smith'},
                       {2,'Joe','Blow'},
                       {3,'Jane','Smith'}],Layout_Person);

somePeople := allPeople(LastName = 'Smith');

//  Outputs  ---
somePeople;
```

Out[3]:

##	personid	firstname	lastname
1	1	Fred	Smith
2	3	Jane	Smith

Quick poll: Could you leverage this new Jupyter kernel in your work?

See poll on bottom of presentation screen



HPCC SYSTEMS®

Questions?

Xiaoming Wang (Ming)
Consultant Software Engineer
LexisNexis Risk Solutions
Xiaoming.Wang@lexisnexisrisk.com

THE DOWNLOAD

TECH TALKS BY HPCC SYSTEMS

HPCC SYSTEMS®

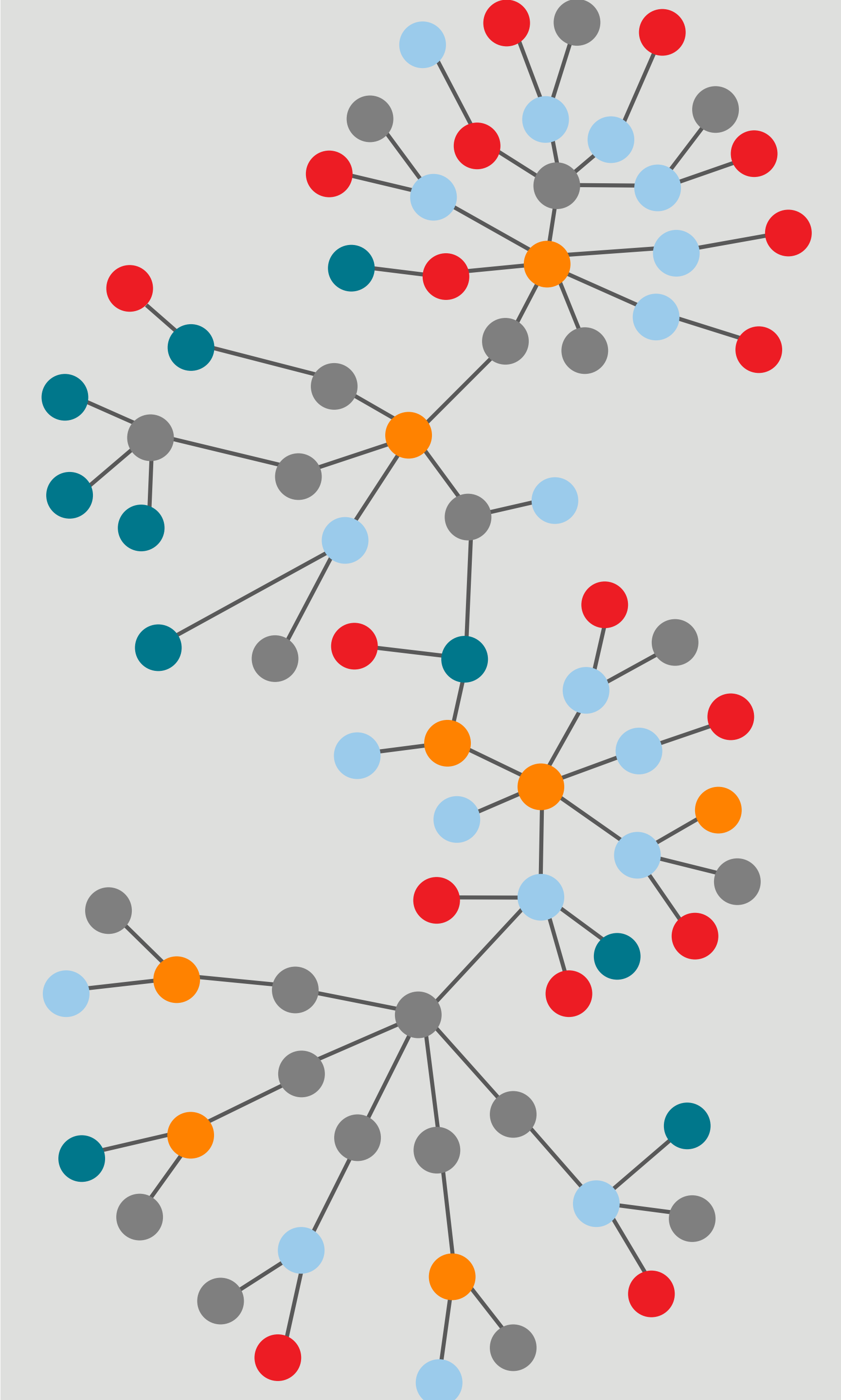
ECL Tips: PROCESS and AGGREGATE functions
Upgrading your ITERATE and ROLLUP

Bob Foreman
Senior Software Engineer
LexisNexis Risk Solutions

LexisNexis®
RISK SOLUTIONS

Quick poll: Are you currently using PROCESS or AGGREGATE in your operations code?

See poll on bottom of presentation screen



HPCC SYSTEMS®

PROCESS

PROCESS(*ds*, *row*, *dstransform*, *rowtransform* **[**,LOCAL**])**

- ✓ *ds* – The set of records to process.
- ✓ *row* – The initial RIGHT record to process.
- ✓ *dstransform* – The TRANSFORM function to call for each record in the *ds*.
- ✓ *rowtransform* – The TRANSFORM function to call to produce the next RIGHT record to process.

The **PROCESS** function operates in a similar manner to ITERATE in that it processes through all records in the *ds* one pair of records at a time, performing the *dstransform* function on each pair of records in turn. The first record in the *ds* is passed to the *dstransform* as the first left record, paired with the *row* as the right record. The *rowtransform* is used to construct the right record for the next pair. If either the *dstransform* or the *rowtransform* contains a SKIP, then no record is produced by the *dstransform* for the skipped record.

The *dstransform* and *rowtransform* functions both must take LEFT and RIGHT records of the same format as the *ds*. The format of the resulting record set for both must be the same as the input *ds*. Optionally, the *dstransform* may take a third parameter: an integer COUNTER specifying the number of times the transform has been called for the *ds* or the current group in the *ds*.

PROCESS example:

```
rec := {STRING2 Letter};
ds   := DATASET([{'A'},{'B'},{'C'},{'D'},{'E'}],rec);

rec XF1(rec L,rec R) := TRANSFORM
  SELF.Letter := L.Letter[1] + R.Letter[1];
END;

rec XF2(rec L,rec R) := TRANSFORM
  SELF.Letter := R.Letter[1] + L.Letter[1];
END;

p := PROCESS(ds, ROW({'Z'},rec), XF1(LEFT,RIGHT), XF2(LEFT,RIGHT));
```

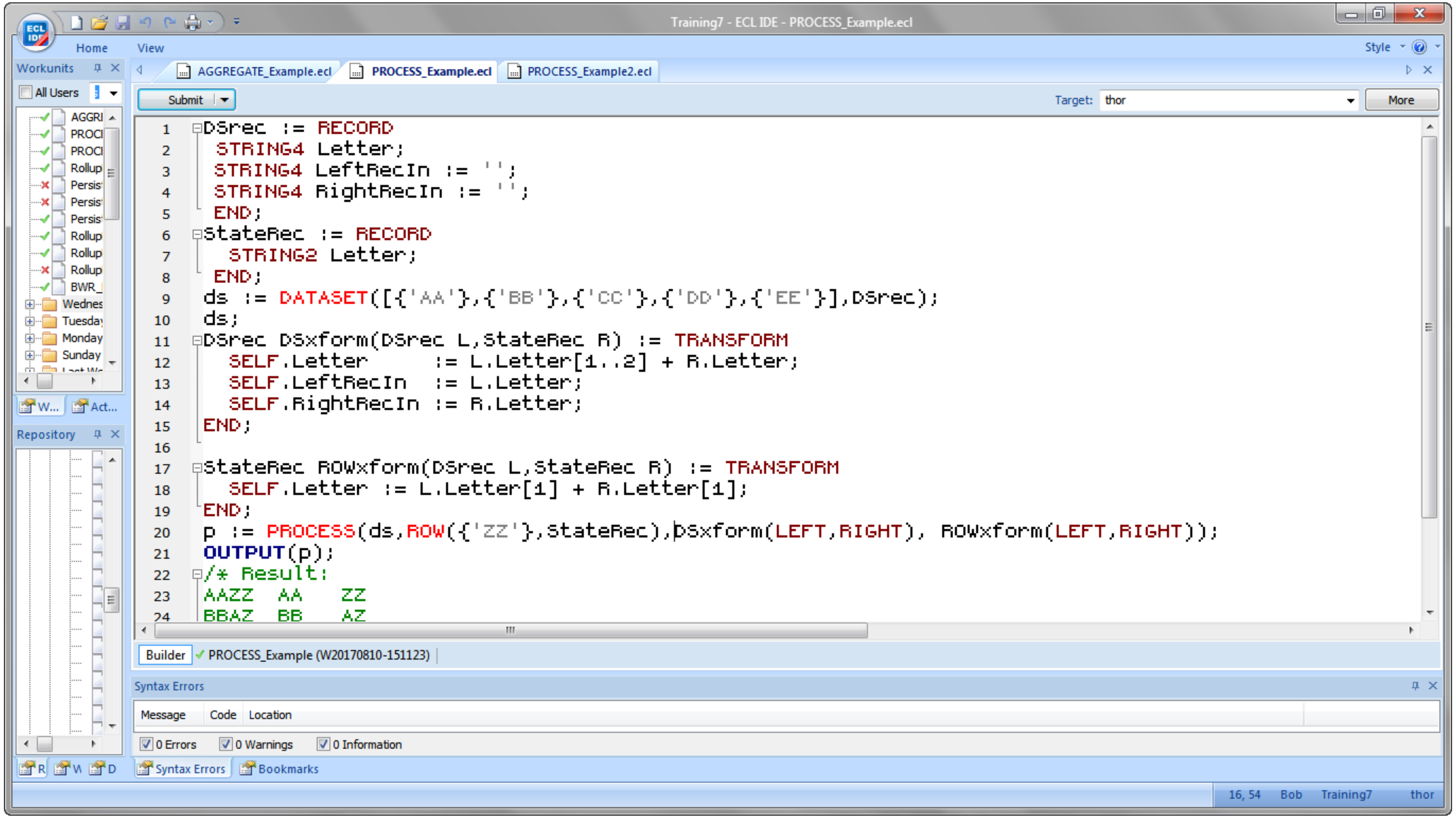
AGGREGATE

AGGREGATE(*recordset, resultrec,maintransform* **[** *, mergetransform* **(RIGHT1,RIGHT2)] [,***groupingfields* **]**

- ✓ *Recordset* The set of records to process.
- ✓ *Resultrec* The RECORD structure of the result record set.
- ✓ *Maintransform* The TRANSFORM function to call for each matching pair of records in the recordset. This is implicitly a local operation on each node.
- ✓ *mergetransform* Optional. The TRANSFORM function to call to globally merge the result records from the maintransform. If omitted, the compiler will attempt to deduce the merge from the maintransform.
- ✓ *groupingfields* Optional. A comma-delimited list of fields in the recordset to group by. Each field must be prefaced with the keyword LEFT. If omitted, then all records match.

The **AGGREGATE** function is similar to ROLLUP except its output format does not need to match the input format. It also has similarity to TABLE in that the *groupingfields* (if present) determine the matching records such that you will get one result for each unique value of the *groupingfields*. The input *recordset* does not need to have been sorted by the *groupingfields*.

More examples:

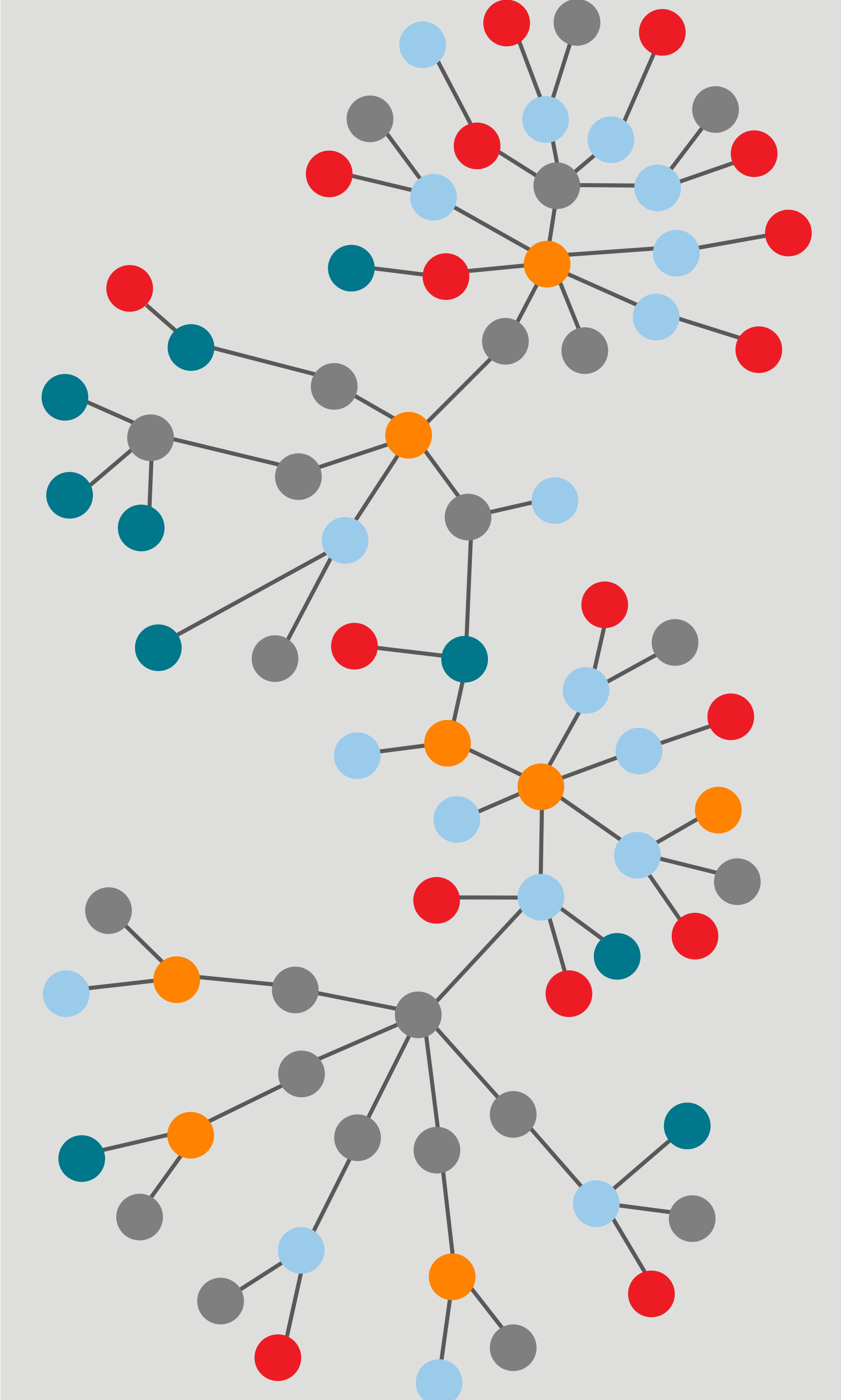


```
DSrec := RECORD
 STRING4 Letter;
 STRING4 LeftRecIn := '';
 STRING4 RightRecIn := '';
 END;
StateRec := RECORD
  STRING2 Letter;
 END;
ds := DATASET([{'AA'},{'BB'},{'CC'},{'DD'},{'EE'}],DSrec);
ds;
DSrec DSxform(DSrec L,StateRec R) := TRANSFORM
  SELF.Letter     := L.Letter[1..2] + R.Letter;
  SELF.LeftRecIn  := L.Letter;
  SELF.RightRecIn := R.Letter;
END;

StateRec ROWxform(DSrec L,StateRec R) := TRANSFORM
  SELF.Letter := L.Letter[1] + R.Letter[1];
END;
p := PROCESS(ds,ROW({'ZZ'},StateRec),DSxform(LEFT,RIGHT), ROWxform(LEFT,RIGHT));
OUTPUT(p);
/* Result:
AAZZ  AA   ZZ
BBAZ  BB   AZ
```

✓ PROCESS_Example (W20170810-151123)

Quick poll:

How many of you outside of LexisNexis or other RELX Group companies are planning to attend our October HPCC Systems Community Day event?



HPCC SYSTEMS®

Questions?

Bob Foreman

Senior Software Engineer

LexisNexis Risk Solutions

Robert.Foreman@lexisnexisrisk.com

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

Join us October 4 when we take Episode 8 on the road for Community Day!

- Visit the Community Day event page for more information.

Thank You!

HPCC SYSTEMS®

RELX Group

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