



Challenge Overview



FEBRUARY 2023

Bob Foreman
Software Engineer Lead
LexisNexis Risk Solutions

The Challenge!

Find Your Paradise!

Have you ever thought about building an application that can help people find places to live that maximize their quality of life and happiness?

The goal of this challenge is to analyze numerous datasets across different categories and make correlations with the data, using the HPCC Systems platform.

After analyzing, the participants will be asked to design an interface to query this data and assign it a scoring system, then deliver it to the user via ROXIE and or Visualization and show the user where they should most likely want to live. Users should be given choices in an easy-to-use form that when submitted will generate a unique set of scores based on locations (Example: By State).



The Challenge!

Find Your Paradise!

Every person's definition of what make them happy can vary depending on several factors. To this challenge, we have narrowed these factors to four (4) categories:

- Crime
- Environment (Weather)
- Health
- Education

Crime – A dataset by US State with crimes between 2018-2021 is provided. Data points can include the number of crimes by State and Violent Crimes by State.

Climate – A dataset of Storm Data in the US over the last 10 years will be provided. Data Points can include number of storms by State, also injuries and fatalities due to a weather incident.

Health – A mortality rate by State from 1980 to 2020 dataset will be provided. Data points can include total mortality and/or average mortality by year and State.

Education – A dataset of the number of public and private schools by State will be provided. Data points include total school by State and percentage of Private Schools available to the Total Schools by State. Also, enrollment and student-to-teacher ratio can be a factor.

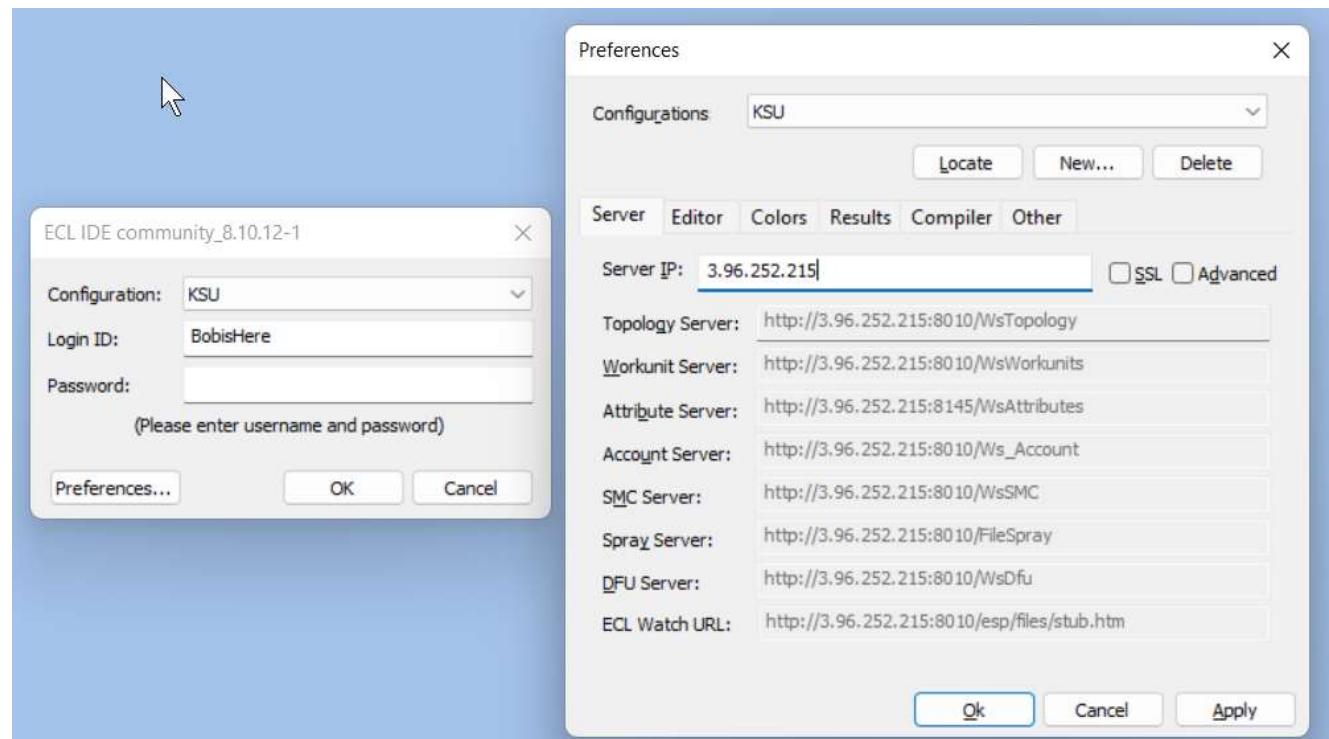
The Playing Field!

HPCC Cluster ECL Watch:

<http://3.96.252.215:8010>

```
{
  "name": "KSU",
  "type": "ecl",
  "request": "launch",
  "protocol": "http",
  "serverAddress": "3.96.252.215",
  "port": 8010,
  "path": "",
  "targetCluster": "thor",
  "rejectUnauthorized": true,
  "resultLimit": 100,
  "timeoutSecs": 60,
  "user": "YourNameHere",
  "password": ""
},
```

{ } launch.json X



The Repo!

<https://github.com/hpccsystems-solutions-lab/FindYourParadise>

The screenshot shows the GitHub repository page for 'FindYourParadise'. The repository is public and has 1 branch and 0 tags. The 'Code' tab is selected. A recent commit by user 'bobf2000' is shown, which updated the README.md file. Below the commit list, the README.md file is displayed, containing the following content:

```
FindYourParadise
An HPCC Systems Hackathon Challenge!
Challenge
Find Your Paradise by weighing selected "quality of life" factors
```



The Resources!

KSU Hackathon HPCC Systems Wiki Page:

<https://wiki.hpccsystems.com/display/hpcc/Kennesaw+State+University+Hackathon+2023>

Learn ECL Academy

<https://hpccsystems-solutions-lab.github.io>

ECL Training containing six short videos

https://www.youtube.com/watch?time_continue=192&v=Lk78BCCtM-0

ECL Documentation

http://cdn.hpccsystems.com/releases/CE-Candidate-8.10.12/docs/EN_US/ECLLanguageReference_EN_US-8.10.12-1.pdf

Visualization Document

https://cdn.hpccsystems.com/releases/CE-Candidate-8.10.12/docs/EN_US/VisualizingECL_EN_US-8.10.12-1.pdf

Standard Library

https://cdn.hpccsystems.com/releases/CE-Candidate-8.10.12/docs/EN_US/ECLStandardLibraryReference_EN_US-8.10.12-1.pdf

Machine Learning

<https://hpccsystems.com/download/free-modules/machine-learning-library>





Education

Education

Public Schools:

This dataset was downloaded from:

https://hifld-geoplatform.opendata.arcgis.com/datasets/87376bdb0cb3490cbda39935626f6604_0

Provided by the Homeland Infrastructure Foundation-Level Data (HIFLD) without a license and for Public Use.

More information about the data can be found here:

<https://hifld-geoplatform.opendata.arcgis.com/datasets/geoplatform::public-schools/about>

Private Schools:

This dataset, taken from the US Department of Homeland Security, contains information on all private schools with attributes regarding their geographical distribution.

License: Public Domain

https://hifld-geoplatform.opendata.arcgis.com/datasets/0dfe37d2a68545a699b999804354dacf_0

Education

Step 1: Combine Public and Private and mark each record accordingly.

BWR_BuildSchools

Step 2: Create new All_Schools File

File_AllSchools

BWR_AllSchoolsData

Step 3: Analyze All Schools

BWR_AnalyzeSchools

Step 4: Build Education Rankings by State

BWR_BuildEducationScores





Crime

Crime Data (see File_Crimes.ecl)

Estimated Crimes 1979-2020:

This dataset was downloaded from the Crime Data Explorer:

<https://cde.ucr.cjis.gov/LATEST/webapp/#/pages/downloads#nibrs-downloads>

It contains estimated data at the state and national level and was derived from the Summary Reporting System (SRS). This dataset reflects the estimates the FBI has traditionally included in its annual publications. See the FBI's estimated crime totals for the nation and all 50 states, 1979 to 2020.

[Summary Reporting System \(SRS\)](#)

This dataset contains estimated data at the state and national level and was derived from the Summary Reporting System (SRS). These data reflect the estimates the FBI has traditionally included in its annual publications. Download this dataset to see the FBI's estimated crime totals for the nation and all 50 states, 1979 to current year available. Due to the full transition to NIBRS, Summary Reporting System data will end with the 2020 data year.

[DOWNLOAD](#)

Data type	Estimated crimes
Years	1979-2020
Last modified	September 10, 2021
File type	CSV
File size	200.8 KB

What's in the download

File	Description
estimated_crimes.csv	Information about estimated crimes

Crime

Step 1: Aggregate Crime By Year and generate ratios by population.

BWR_AnalyzeCrime

Step 2: Generate scoring for Paradise Service

BWR_BuildCrimeScores

Generate Rankings

Example shows Violent Crime and Property Crimes





Health (Mortality)

Health (Mortality) - File_Mortality

Mortality By County (1980-2014):

Originally from the Institute for Health Metrics and Evaluation

<https://www.kaggle.com/datasets/IHME/us-countylevel-mortality>

Grouped by County FIPS, Category - 5-year increments or all years.

Mortality By State, Race, and Gender (1990-2019):

Source: Institute for Health Metrics and Evaluation

<https://ghdx.healthdata.org/record/ihme-data/united-states-life-expectancy-by-state-white-black-hispanic-race-ethnicity-1990-2019>

Estimates were produced for mortality rates, life expectancy, and population at the state level in the United States, and by racial/ethnic group, for each year between 1990-2019. These estimates were produced using population and deaths data from the National Center for Health Statistics.

Health

Step 1: Add State information to County Mortality Dataset.

Step 2: Grab the State aggregates for scoring

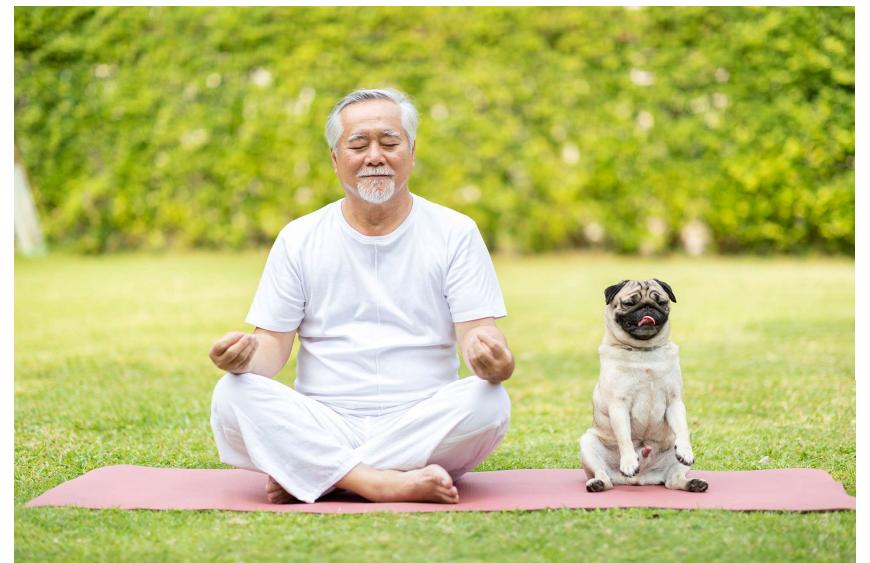
Step 3: CrossTab Mortality aggregates by State - and get averages!

File_StateFIPS - Lookup table used in JOIN

BWR_AnalyzeHealth

Step 4: Build Mortality Scores for Paradise Ranking

BWR_BuildHealthScore





Weather

Weather

Storm Events (2018-2022):

The Storm Events Database contains the records used to create the official NOAA Storm Data publication, documenting:

- The occurrence of storms and other significant weather phenomena having sufficient intensity to cause loss of life, injuries, significant property damage, and/or disruption to commerce;
- Rare, unusual, weather phenomena that generate media attention, such as snow flurries in South Florida or the San Diego coastal area; and
- Other significant meteorological events, such as record maximum or minimum temperatures or precipitation that occur in connection with another event.

The database currently contains data from January 1950 to October 2022, as entered by NOAA's National Weather Service (NWS).

<https://www.ncdc.noaa.gov/stormevents/>



Weather

Step 1: Combine Yearly Weather Datasets into a single Superfile.

File_Weather

Step 2: Sequence Records and add State Code

Step 3: Cross-Tab by State, Weather Event Count

Step 4: Create a Severity Code for Scoring

Step 5: Combing Events, Injuries and Fatalities for Scoring

File_StateFIPS

BWR_AnalyzeWeather

Step 6: Build Weather Score Ranking by State

BWR_BuildWeatherScores





Composite Scores

Building a Composite Dataset

Step 1: Combine Scores from all Categories into a single recordset.

Step 2: Output to a file for indexing.

Step 3: Build the INDEX

File_Composite

BWR_BuildCompositeScores

Step 4: Build the INDEX

BWR_BuildIndex





Data Delivery (ROXIE or Visualization)

Creating an HTHOR/ROXIE query

Step 1: Determine from your Index what are your Paradise parameters?

iParadise

Step 2: Write the Query! Calculate Paradise Scores based on User input.

FindURParadiseSvc

Step 3: Test with HTHOR first. Set Target, Compile, Publish and Test

Step 4: Publish to ROXIE. Set Target, Compile, Publish and Test!



Alternate Delivery: Visualization

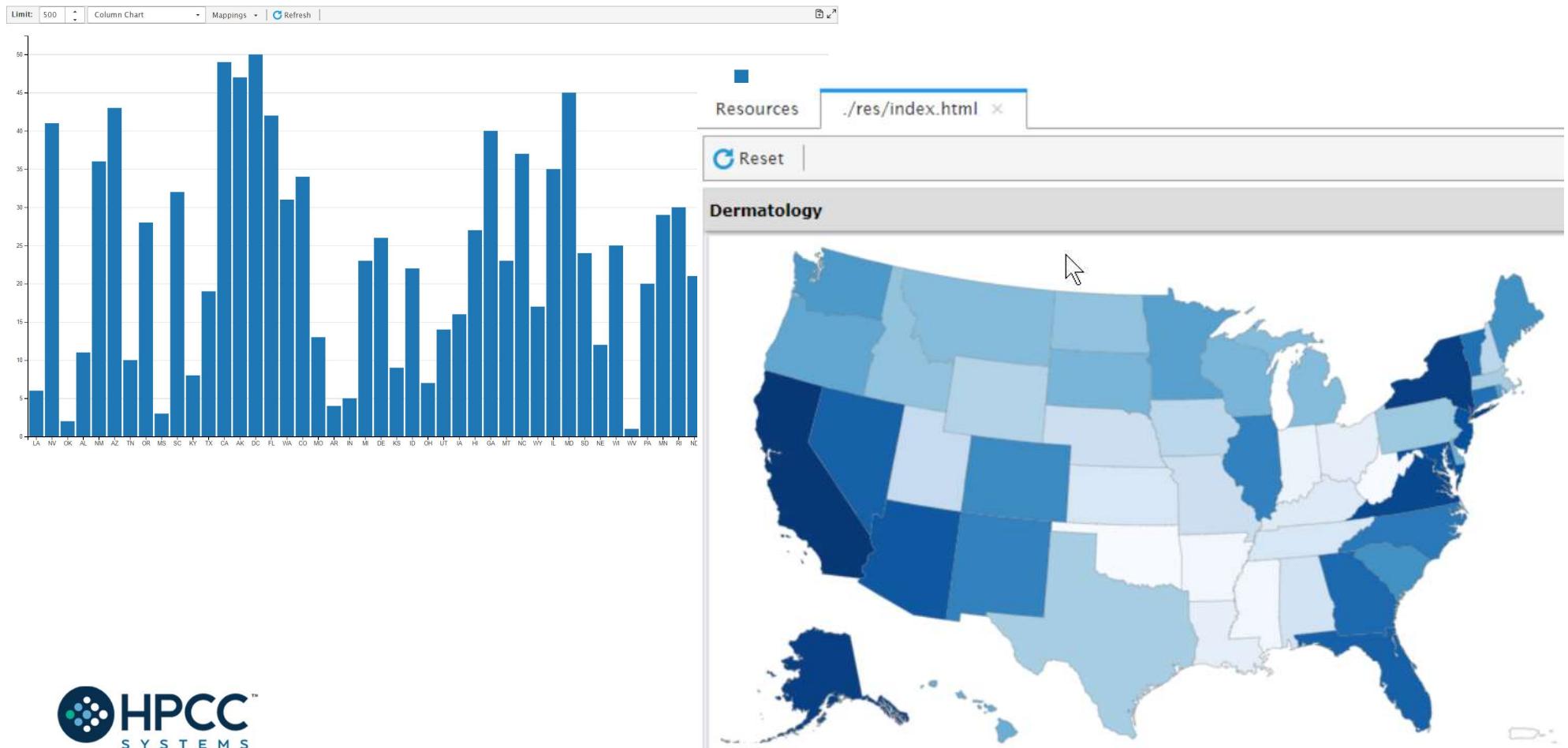
HPCC Systems provides built-in Visualization of your output data in a variety of charts and graphs. You can visualize your data in three ways:

- Using the Chart Tool in the ECL Playground.
- Accessing the Visualize tab in all ECL workunits
- Using the Resources tab in conjunction with the ECL Visualizer bundle.

Installing:

```
ecl bundle install https://github.com/hpcc-systems/Visualizer.git
```

Visualization Examples:



Final Thoughts

- ✓ Since your solution is the key part to this challenge you can use `#OPTION('obfuscateOutput' , TRUE)`; at the start of your code to hide it from being viewed on ECL Watchpage. If you decide to use `#OPTION` make sure to remove it from the WUID that you shared with the judges. When `obfuscateOutput` set to true, details are removed from the generated workunit, including ECL code, estimates of record size, and number of records.
- ✓ If you want to write the result to a file, make sure the file name starts with your team's name for uniqueness purpose.
- ✓ Make sure the query names are unique and easy to identify. Do not use generic names like test, mentors, or roxie. We suggest adding your team's name as well. General names will result in other teams overwriting your files, queries, and results
- ✓ We encourage team play so teams that help answer questions in Slack will be considered favorably in judging.
- ✓ Direct emails and direct messages to judges asking for support will be *ignored* and it won't work in team's favor
- ✓ We also encourage students to leverage our community forum and/or StackOverflow for ECL coding related questions. Please make sure to tag your questions with `hpcc-ecl`.

Get in Touch

Robert.Foreman@lexisnexisrisk.com



