Unified Learning Environment for Analytics & Data

Data Week – Day 1
May 24, 2021
Andrew Weaver, Mounir Benjelloun, Jason Wang
Vision for ULEAD

UNIFIED ENVIRONMENT
Single location for data, analytics, and compute

SCALABLE ENVIRONMENT
Supports growing business needs

SECURE ENVIRONMENT
Analysis inside UCLA Health security perimeter

FASTER DATA APPROVAL
Reduce time for data release approvals

Users

Researchers
Data Scientists
Bio Statisticians
Analysts
Administrators

OHIA | Office of Health Informatics & Analytics
https://ulead.mednet.ucla.edu
ULEAD High-Level Architecture
Functionality of ULEAD Pilot

The ULEAD Pilot phase will allow the following activities:

**CTSI**
1. Ability to provide data to researchers in ULEAD (flat files)

**Researchers**
1. Ability to copy data provided by CTSI into a User Folder within ULEAD
2. Ability to bring other data into ULEAD
3. Ability to connect to approved data sources
4. Ability to blend data from CTSI with other data sources. E.g. Cancer Registry, images, etc.
5. Ability to load data into a database
6. Ability to analyze and visualize data
7. Limited ability to develop, train and test ML models
8. Ability to summarize results and prepare reports/graphs
9. Ability to share summarized data with the UCLA Health research community, within ULEAD

---

**ULEAD Specifications**

The ULEAD pilot platform will offer each user the features listed below, and will support 250 concurrent users.

- Virtual Desktop: Compute: 4 CPUs (10~ GHz), RAM: 32 GB. Can be increased on a case-to-case basis
- Database: MS Access
- Software: Python, R, SQL Server, SAS
  - For licensed software such as SAS and MATLAB, users must bring their own license
- Personal Storage: User Folder within ULEAD. 50GB+ storage Can be increased on a case-to-case basis
- Collaboration: Team Folders within ULEAD
- Data Transfer In & Out: Data Transfer Folders within ULEAD
- Internet access: Websites approved by OHIA
- Interfaces: Allowed with approved databases and applications
Accessing Data in ULEAD

Data Transfer Inbound - CTSI

Data Transfer Folders – OUTSIDE ULEAD

CTSI Folders

Data Transfer Folder

– OUTSIDE ULEAD

Data Transfer Folder – OUTSIDE ULEAD

Collaboration within ULEAD

Data Transfer Outbound - User

Data Transfer Folders – OUTSIDE ULEAD

CTSI Folders

Data Transfer Folder

– OUTSIDE ULEAD

Data Transfer Folder – OUTSIDE ULEAD

Outbound
PRODUCT ROADMAP

v1
- Adobe Acrobat
- Anaconda
- Azure DevOps
- Jupyter Notebooks
- MS Office (Word, Excel, Powerpoint, Access)
- PyCharm
- R Studio
- MATLAB
- SAS
- SSMS SQL Developer
- Visual Studio

v2
- REDCap
- Tableau
- Power BI
- Oracle SQL Developer
- Databricks
- GitHub
- STATA
- Image processing software (Gimp, Osiris)

Lifecycle Mgmt
- Ongoing Backlog by User Request

Analytical & Visualization Tools

Data
- Text
- Images
- Waveforms
- Genomics

Compute
- CPUs
- GPUs
- High Performance Compute
Challenges with Licensed Applications

- Previous generations (Legacy xDR Desktop, Tableau Desktop, On Premises VDI):
  - Does not support individual licensed applications
  - Only supported non-persistent desktops
    - Affects support for licensed applications
    - Application compatibility issues
  - Does not support publishing applications to individual users or groups of users.

- ULEAD supports:
  - Licensed applications for individuals or groups of users
  - Provides the ability to publish applications via Citrix and make them available to our users
  - Applications can be installed and stay persistent per user by using UPL (User Personalization Layer)
Individually Licensed Applications Provisioning

• Determine if application is compatible with desktop (Virtual Desktop) or server OS (Virtual Applications)

• Evaluate application license requirements
  o Configure firewall to communicate with license servers as needed

• Install and test application

• Configure storage connections from the application

• Procure and install license as needed
  o Bring your own license
  o Volume licensing

• Publish application to users workspace

• Test and validate application functionality and performance

• Recycle the environment as needed
Deliver Virtual Applications per user

- If application is compatible with server OS (Virtual apps), we can publish applications to users, as needed.
- When applications are no longer needed, access is removed and granted to a different user maintaining license requirements
- Preferred method
Provision applications per users as needed

- Users login to Citrix Workspace and access applications based on entitlements
Deliver Virtual Desktops with apps installed (UPL)

- With applications that are compatible with desktop OS, Citrix User Personalization layer is used to install applications per user.
## Applications Available per User

### User A

<table>
<thead>
<tr>
<th>Application</th>
<th>Version</th>
<th>Date</th>
<th>Size (MB) 1</th>
<th>Size (MB) 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Google Chrome</td>
<td>2.1.520 (x64)</td>
<td>2/11/2021</td>
<td>69.9 MB</td>
<td>88.0 MB</td>
</tr>
<tr>
<td>IIS 10.0 Express</td>
<td></td>
<td>1/29/2021</td>
<td>40.3 MB</td>
<td>10.0 MB</td>
</tr>
<tr>
<td>Microsoft .NET Core SDK 2.1.520 (x64)</td>
<td></td>
<td>1/28/2021</td>
<td>477 MB</td>
<td>2.1 MB</td>
</tr>
<tr>
<td>Microsoft .NET Framework 4.5 Multi-Targeting Pack</td>
<td></td>
<td>1/29/2021</td>
<td>41.8 MB</td>
<td>4.5 MB</td>
</tr>
<tr>
<td>Microsoft .NET Framework 4.5.1 Multi-Targeting Pack</td>
<td></td>
<td>1/29/2021</td>
<td>49.3 MB</td>
<td>4.5 MB</td>
</tr>
<tr>
<td>Microsoft .NET Framework 4.5.1 Multi-Targeting Pack...</td>
<td></td>
<td>1/29/2021</td>
<td>74.5 MB</td>
<td>4.5 MB</td>
</tr>
</tbody>
</table>

### User B

<table>
<thead>
<tr>
<th>Application</th>
<th>Version</th>
<th>Date</th>
<th>Size (MB) 1</th>
<th>Size (MB) 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Google Chrome</td>
<td></td>
<td>2/11/2021</td>
<td>69.9 MB</td>
<td>88.0 MB</td>
</tr>
<tr>
<td>IIS 10.0 Express</td>
<td></td>
<td>1/29/2021</td>
<td>40.3 MB</td>
<td>10.0 MB</td>
</tr>
<tr>
<td>Oracle</td>
<td></td>
<td>4/6/2021</td>
<td>91.9 MB</td>
<td>6.0 MB</td>
</tr>
<tr>
<td>Oracle</td>
<td></td>
<td>4/6/2021</td>
<td>87.3 MB</td>
<td>6.0 MB</td>
</tr>
<tr>
<td>MATLAB R2020b</td>
<td></td>
<td>4/1/2021</td>
<td>129 KB</td>
<td>6.2 KB</td>
</tr>
<tr>
<td>MathWorks</td>
<td></td>
<td>5/13/2021</td>
<td>9.9</td>
<td></td>
</tr>
<tr>
<td>Microsoft .NET Core SDK 2.1.522 (x64)</td>
<td></td>
<td>3/24/2021</td>
<td>477 MB</td>
<td>2.1 MB</td>
</tr>
</tbody>
</table>
ULEAD Demo

Jason Wang
ULEAD DEMO

Vu Vu
Manager, Advanced Analytics, OHIA

Jason Wang, MS. PhD
Researcher
Advanced Analytics, OHIA
Dept. of Medicine, David Geffen School of Medicine
A use case for this demo

Excess Patient Visits for Cough and Pulmonary Disease at a Large US Health System in the Months Prior to the COVID-19 Pandemic: Time-Series Analysis
Joann G Elmore 1; Pin-Chieh Wang 1; Kathleen F Kerr 2; David L Schriger 3; Douglas E Morrison 4; Ron Brookmeyer 4; Michael A Pfeffer 1; Thomas H Payne 5; Judith S Currier 1

More than 30+ media from regional and international

UCLA Newsroom Article:
"COVID-19 may have been in LA as early as December, UCLA-led study suggests" (UCLA Newsroom, 9/16/20).
http://newsroom.ucla.edu/releases/covid-may-have-been-in-la-as-early-as-december-2020

Major US Papers and Media (~10):
"Coronavirus Today: Was the virus in L.A. last year?" (LA Times Newsletter, 9/10/20).
http://www.latimes.com/science/newsletter/2020-09-10/ucla-hospital-december-coronavirus-
today

"High numbers of Los Angeles patients complained about coughs as early as December, study says" (Washington Post, 9/10/20).

Video:
Joann Elmore interviewed. "Researchers at UCLA say COVID-19 may have been in Los Angeles since December" (9/30/20).
https://www.foxla.com/video/840508

Spanish media:
"UCLA: COVID-19 podría haber estado circulando en Los Ángeles en diciembre" (Telemundo, 9/30/20).
http://www.telemundo52.com/noticias/local/ucla-covid-19-puede-haber-estado-circulando-
en-los-angeles-en-diciembre/2114513/

International:
"More evidence coronavirus had spread to the West BEFORE China came clean: LA doctors find evidence it may have been circulating in US MONTHS earlier than first reported case" (Daily Mail, 9/19/20).
http://www.dailymail.co.uk/health/article-8716753/COVID-19-LA-CHRISTMAS-study-
另行一句.html

Additional Media, TV and Radio (~20):
"UCLA Study: COVID-19 May Have Been Spreading In LA Back In December" (CBS 2/KCAL 9 Los Angeles, 9/30/20).

"Coronavirus may have been in LA as early as December, UCLA study suggests (ABC 7 Los Angeles, 9/12/20).
https://abc7.com/coronavirus-los-angeles-study-ucla/6418477/
Study background

- Clinicians have noted an unusual number of patients with respiratory complaints at the end of 2019 and early 2020, well before COVID-19 was officially categorized by the World Health Organization (WHO) as a pandemic.

- The excess can represent undetected and early COVID-19 cases prior to established clinical awareness and testing capabilities for the virus.

- Limited data available to accurately assessing regional activity of COVID pandemic, It is unclear whether such anecdotal reports are correct or the result of hindsight bias.
Can we use EHR data to evaluate whether there was an excess of patients presenting with symptoms and diseases suggestive of COVID-19 in the months prior to the first known cases in the US health system in March 2020?

Specifically, can we use symptoms suggestive of COVID-19 as proxy measures to better understand the COVID pandemic?

We can use words found in chief complaint fields in EHR with symptoms suggestive of COVID-19 using time-series methods can address whether there was an excess number of patients presenting for complaints of cough, or hospitalizations for respiratory ailments.

Specifically, can we use symptoms suggestive of COVID-19 as proxy measures to better understand the COVID pandemic?
UCLA Research Project Process Flow

**Researcher + OHIA**
- Brainstorming
- Forming study hypothesis
- Visit for Cough as proxy measure

**IRB**
- Submit IRB application

**CTSI**
- Refine study design
- Data preparation

**Researcher + CTSI + OHIA**
- Refine data specification
- Obtaining data
- Work with OHIA Advanced Analytics team to develop predictive model
- OHIA provide development platform (ULEAD)
**Study design for this demo**

**Study hypothesis:** Excess patient visits for cough in the Months Prior to the COVID-19 Pandemic

**Outcome measures:** visit for cough

**Data source and time frame:** Emergency department data from 2014/July to 2020/Feb

**Statistical analysis:** Using Time-Series Analysis (ARIMA model) to analyze and forecast time series data.

**Method:** Examine the time when the actual value starts to deviate from the predicted data.
ARIMA Model

1. Use SAS Seasonal ARIMA model to take into account seasonal pattern in this data.
2. Use 4 years data as a train data to develop a model (between 2014 – 2018/Nov)
3. Use one year data to validate model (between 2018/Dec – 2019/Nov)
4. Compare actual data with predicted interval from 2019/Dec to 2020/Feb for hypothesis testing.
Options to do Time-Series Analysis in ULEAD

- **SAS modules/procedures**: ARIMA, model, forecast, autoreg, X11, X12, Spectra, statespace, varmax, time series forecasting system, Call R program directly into SAS
- **R packages**: forecast and tseries
- **Python** (Scipy environment: pandas, statsmodels, scikit-learn)
Steps for this demo in ULEAD

1. Verify data from CTSI in File explore & Excel
2. Visualizing time-series data using R-Studio
3. ARIMA modeling using SAS
4. Auto.ARIMA using R in Jupyter Notebook
5. Generate reports

Statistical analysis/ machine learning/ predictive model development
Summary

With the ULEAD go live:

• Fully secure environment
• High performance machine
• Easily to access data
• Capable of using multiple programs
• Ability to generate high quality reports