



Innovation in
Medical
Evidence
Development and
Surveillance

IMEDS Research Lab

Reagan-Udall Foundation for the FDA

Mark Khayter

July 17, 2014

All You Ever Wanted to Know About the IMEDS Research Lab

- National resource
- Incubator for methods and outcomes research
- Use of large healthcare datasets
- Rich resource for innovation
- Community of people
- Learning forum
- Secure, centralized infrastructure with support
- Tools to decrease redundancy of work – availability of results datasets
- Create once, use many times

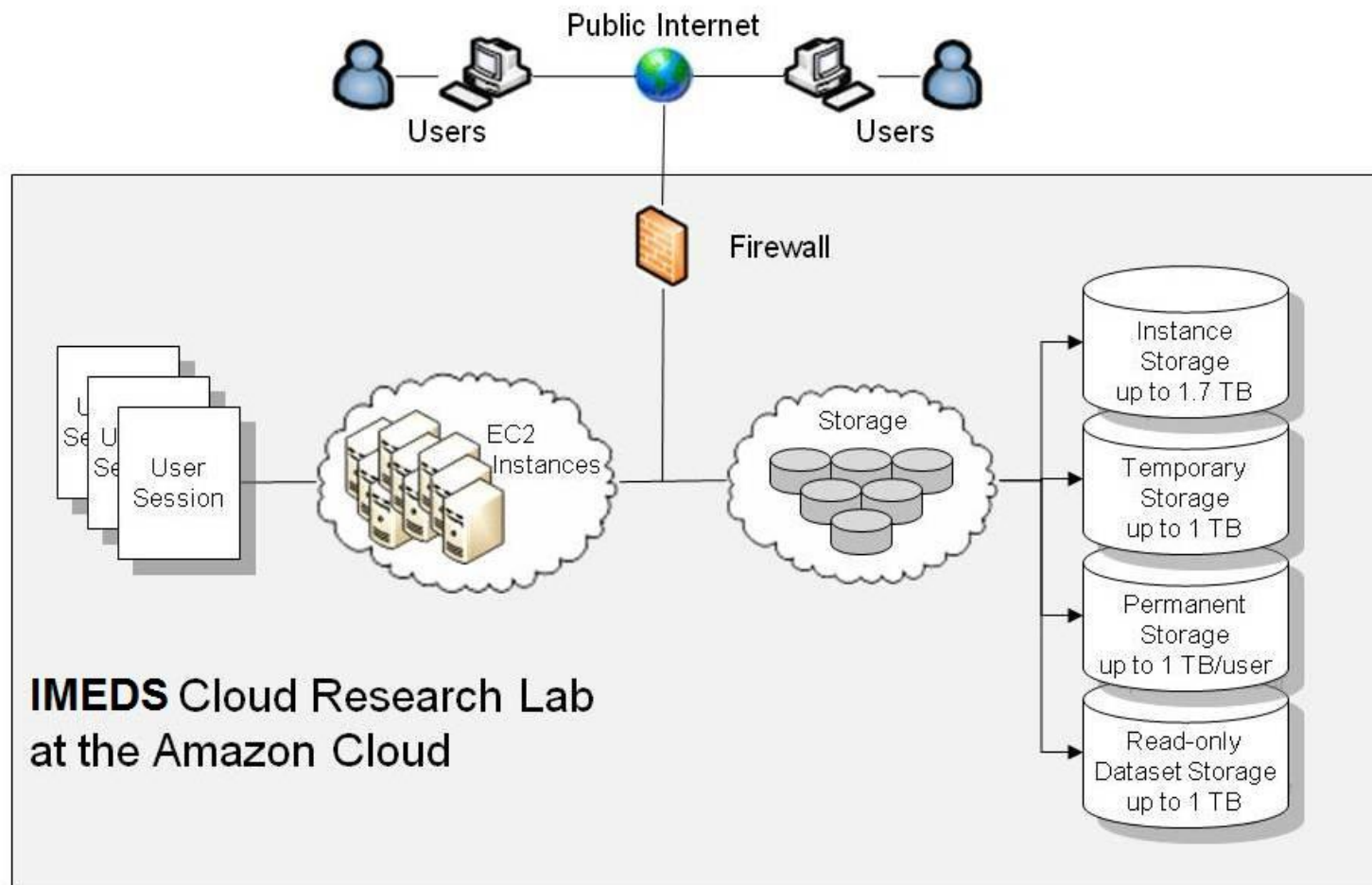
IMEDS Research Lab Overview

- Amazon Cloud infrastructure
- Variety of computer images with pre-installed tools
- Datasets – Claims, EHR. Over 200M persons
- Datasets in OMOP and Mini-Sentinel CDM formats
- Collaborative environment
- Secured
- Accessible from anywhere
- Unlimited research resources
- Browser access
- Support for on-boarding and user group forum

Research/Development at IMEDS Research Lab

- Statistical methods development
- Health Outcome research
- Tools and visualization development
- Reproduce research results produced elsewhere
- CDM and Vocabulary queries
- IMEDS Vocabulary development and distribution

IMEDS Research Lab access



Datasets in the IMEDS Research Lab

<http://imeds.reaganudall.org/On-Boarding>

Abbreviation	Name	Description	Population	Observation time	Drugs	Conditions	Procedures	Observations
CCAE	MarketScan Commercial Claims and Encounters	Represents privately insured population and captures administrative claims with patient-level de-identified data from inpatient and outpatient visits and pharmacy claims of multiple insurance plans.	Total:141.8m %male:48.9 Mean Age(SD): 36 (18.57)	Pt-years: 339.8m 2003-2013	Records : 2,876.1m NDC from pharmacy dispensing claims HCPCS/CPT/ICD9P procedures from inpatient/outpatient medical claims	Records : 4,604.6m ICD9 from inpatient/outpatient medical claims	Records : 9,845.2m HCPCS/CPT/ICD9P procedures from inpatient/outpatient medical claims	Records : 385.4m LOINC from outpatient laboratory services
MDCD	MarketScan Multi-State Medicaid	Contains administrative claims data for Medicaid enrollees from multiple states, including inpatient, outpatient, and pharmacy services.	Total:19.8m %male:42.4 Mean Age(SD): 28.0 (23.1)	Pt-years: 56.5m 2002-2012	Records : 878m NDC from pharmacy dispensing claims HCPCS/CPT/ICD9P procedures from inpatient/outpatient medical claims	Records : 1,431.9m ICD9 from inpatient/outpatient medical claims	Records : 2,585.8m HCPCS/CPT/ICD9P procedures from inpatient/outpatient medical claims	Not available
MDCR	MarketScan Medicare Supplemental Beneficiaries	Captures administrative claims for retirees with Medicare supplemental insurance paid by employers, including services provided under Medicare-covered payment, employer-paid portion, and any out-of-pocket expenses.	Total:10.6m %male:44.8 Mean Age(SD): 78 (9.5)	Pt-years: 32m 2003-2013	Records : 805.6m NDC from pharmacy dispensing claims HCPCS/CPT/ICD9P procedures from inpatient/outpatient medical claims	Records : 1,121.2m ICD9 from inpatient/outpatient medical claims	Records : 1,886.2m HCPCS/CPT/ICD9P procedures from inpatient/outpatient medical claims	Records : 20.7m LOINC from outpatient laboratory services
MSLR	MarketScan Lab Supplemental	Represents privately insured population that has at least one recorded laboratory value, with administrative claims from inpatient, outpatient, and pharmacy services supplemented by laboratory results.	Total:4.4m %male:35.9 Mean Age(SD): 44.8 (17.2)	Pt-years: 8.9m 2003-2011	Records : 124.5m NDC from pharmacy dispensing claims HCPCS/CPT/ICD9P procedures from inpatient/outpatient medical claims	Records :224.5m ICD9 from inpatient/outpatient medical claims	Records : 415.8m HCPCS/CPT/ICD9P procedures from inpatient/outpatient medical claims	Records : 234.9m LOINC from outpatient laboratory services
GE	GE Centricity	The GE MQIC (Medical Quality Improvement Consortium) represents the group of providers who use the GE Centricity Electronic Medical Record and who contribute their data for secondary analytic use. The GE MQIC database reflects events in usual care, including patient problem lists, prescribing patterns and over-the-counter use (OTC) of medications, and other clinical observations as experienced in the ambulatory care setting.	Total:33.3m %male:42.7 Mean Age(SD): 44 (22.5)	Pt-years: 73.6m 1986-2014* *1986 is first year with more than 10,000 people	Records : 824.4m GPI from medication history and prescriptions written	Records : 322.6m ICD9 from problem list	Records : 637m CPT from procedure list	Records : 4,620.3m LOINC for laboratory values, SNOMED for chief complaints, signs and symptoms
OSIM2		OMOP Simulated data, modeled after MSLR	Total : 10m % male: 35% Mean Age (SD): 37.6 (17.7)	Pt-years: 16.4m 2003-2007	Records: 118.5m RxNorm concepts, modeled after DRUG_ERA table from MSLR_CDM_MEDDRA	Records: 174.7m MedDRA concepts, modeled after CONDITION_ERA table from MSLR_CDM_MEDDRA	Not simulated	Not simulated

Access to the IMEDS Research Lab

<http://imeds.reaganudall.org/On-Boarding>

- Submit a brief proposal to gain authorized access
 - <http://imeds.reaganudall.org/LabUsers>
 - Send email to ewealebob@reaganudall.org
- Review by the Senior Director, IMEDS Methods-Research and the IMEDS Scientific Advisory Committee
- Sign and agree to the policies and procedures
- Gain access to the lab and begin your research

IMEDS Research Lab – Running Instances

- Access to the Lab: <https://174.129.41.64>



IMEDS
Research Lab

[Mark Khayter](#)

[log out](#)





























RUNNING INSTANCES

Running Instances

Launch

Client Install

Running Instances

Name	Connect	Host	Start Date	Instance Type	Dataset	Image	Actions
WINDOWS 32 FDA Demo	  	ec2-107-21-130-242.compute-1.amazonaws.com	29-Mar-2014 12:34:45 PM	Standard Small		Windows 32	
SAS Truven ALL	  	ec2-174-129-148-223.compute-1.amazonaws.com	16-Mar-2014 09:05:13 PM	Standard Large		SAS	
SAS MSLR_CDM_MEDDRA 8	  	ec2-54-226-211-180.compute-1.amazonaws.com	27-Jan-2014 12:33:15 PM	Standard Extra Large	MSLR_CDM_MEDDRA	SAS	
SAS Truven 1995_2013	  	ec2-50-16-155-106.compute-1.amazonaws.com	19-Jan-2014 09:10:47 PM	Standard Large		SAS	
LINUX 32 25	  	ec2-107-20-65-31.compute-1.amazonaws.com	14-Jan-2014 11:17:29 AM	Standard Small		Linux 32	
CPRD 2013 Aug Raw GE Patients	  	ec2-23-20-117-127.compute-1.amazonaws.com	10-Aug-2013 05:00:17 PM	Micro		Linux 32	
Ivan Martijn Storage	  	ec2-50-16-152-224.compute-1.amazonaws.com	23-Dec-2011 02:48:54 PM	Standard Small		Programming	

IMEDS Research Lab - Launch



IMEDS
Research Lab

[Mark Khayter](#)

[log out](#)

LAUNCH

Running Instances

Launch

Client Install

Configure Instances

Image:

- Dataset:
- MSLR V2 Linux/Win
 - MiniCentinel CCAE Set 1
 - MiniCentinel CCAE Set 2
 - GE
 - MSLR
 - MSLR Sorted
 - MDCR
 - MDCD
 - CCAE
 - GE_MEDDRA
 - GE-ICD9
 - GE_CDM4 In V2 format
 - MSLR_CDM_MEDDRA
 - MSLR-ICD9
 - Vocabulary
 - MDCD_CDM_MEDDRA
 - Vocabulary_V4
 - MDCR_CDM_MEDDRA
 - CCAE_CDM_MEDDRA
 - GE_CDM_MEDDRA

Instance Type:

Description: 3.75 GiB memory
2 EC2 Compute Unit (1 virtual core with
2 EC2 Compute Unit)
410 GB instance storage
32-bit or 64-bit platform

Load Method Code:

Other Users Storage:

My Personal Storage:

Temporary Storage:

Number of instances *:

Instance Name 1 *:

Cost and Budget

Estimated Instance \$/h:	0.12
Estimated Storage \$/h:	0.01
Estimated Total \$/h:	0.13
Remaining Budget \$:	1,149
Instances Limit:	101
Running Instances:	7

Submit

IMEDS Research Lab – Edit Account



IMEDS
Research Lab

[Mark Khayter](#)

[log out](#)

EDIT ACCOUNT

Running Instances

Launch

Client Install

Email*:

First Name*:

Last Name*:

Phone*:

Organization:

Title:

Password:

Password Again:

Password will be reset only if it is entered

Submit

Dataset access:

- MSLR V2 Linux/Win
- MiniCentinel CCAE Set 1
- MiniCentinel CCAE Set 2
- GE
- MSLR
- MSLR Sorted
- MDCR
- MDCD
- CCAE
- GE_MEDDRA
- GE-ICD9
- GE CDM4 In V2 format
- MSLR_CDM_MEDDRA
- MSLR-ICD9
- Vocabulary
- MDCD_CDM_MEDDRA
- Vocabulary_V4

Image access:

- Linux 32
- Linux 64
- Linux 64 SAS-T
- SAS
- Linux 64 SAS
- Windows 32
- Windows 32 Ora/RedShift
- Windows 64
- GPU_SAS
- Programming GPU
- Windows SAS

Charge limit (\$):

Remaining (\$):

Max instances:

Allow sharing my storage with

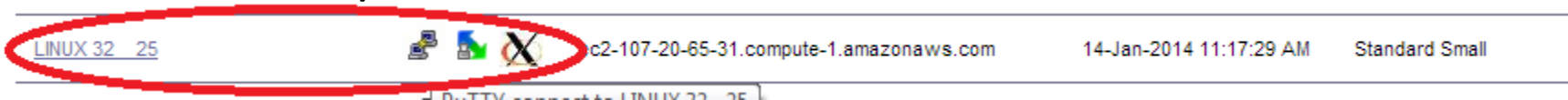
- creich
- dtorok
- abarnard
- ahartzema
- bletham
- dmadigan
- emoyneur
- enikhoma
- ewelebob
- gtrifiro
- izorych
- jasmandawu
- jduke
- jweiss
- mkhayter1
- mkhayter2
- moverhage
- mschuemie
- tmccormick
- vhuser
- vscosta
- Ywang
- zshahn

Users sharing their storage with me

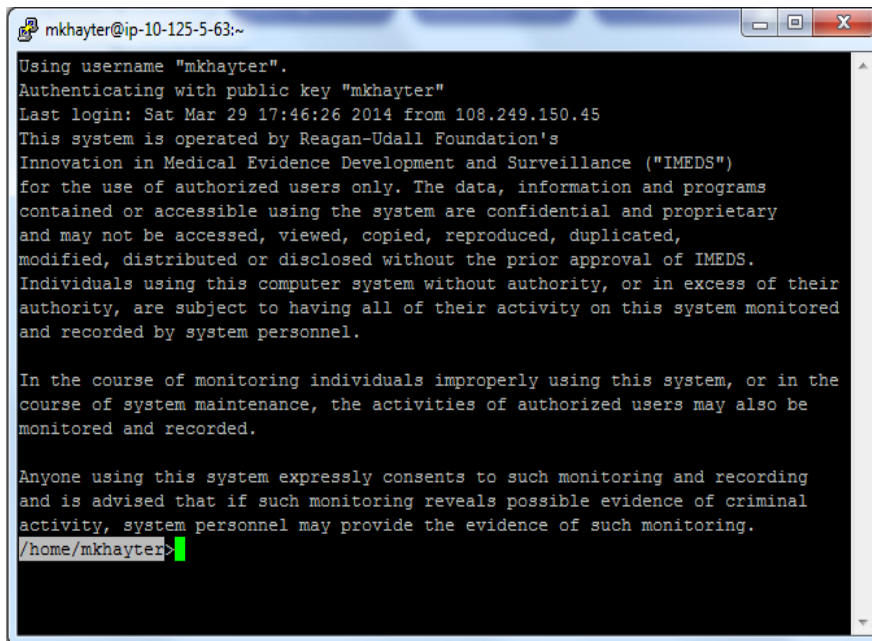
- creich
- dmadigan
- dtorok
- mschuemie
- msuchard
- pryan
- rhansen
- rharpaz

IMEDS Research Lab – Linux Access

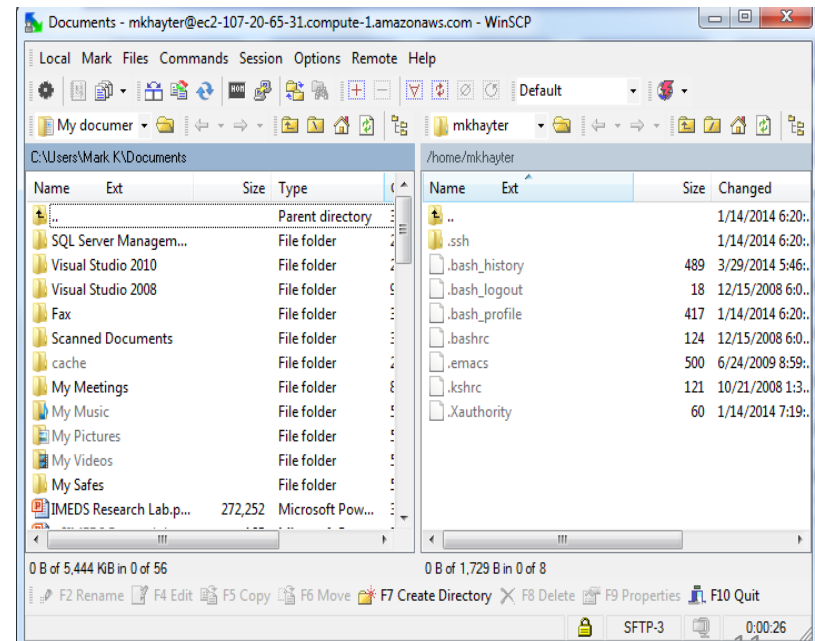
In Running Instances, click for Linux instance on PuTTY, WinSCP, or Xwindows option.



PuTTY

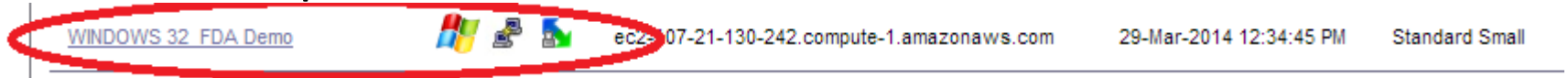


WinSCP

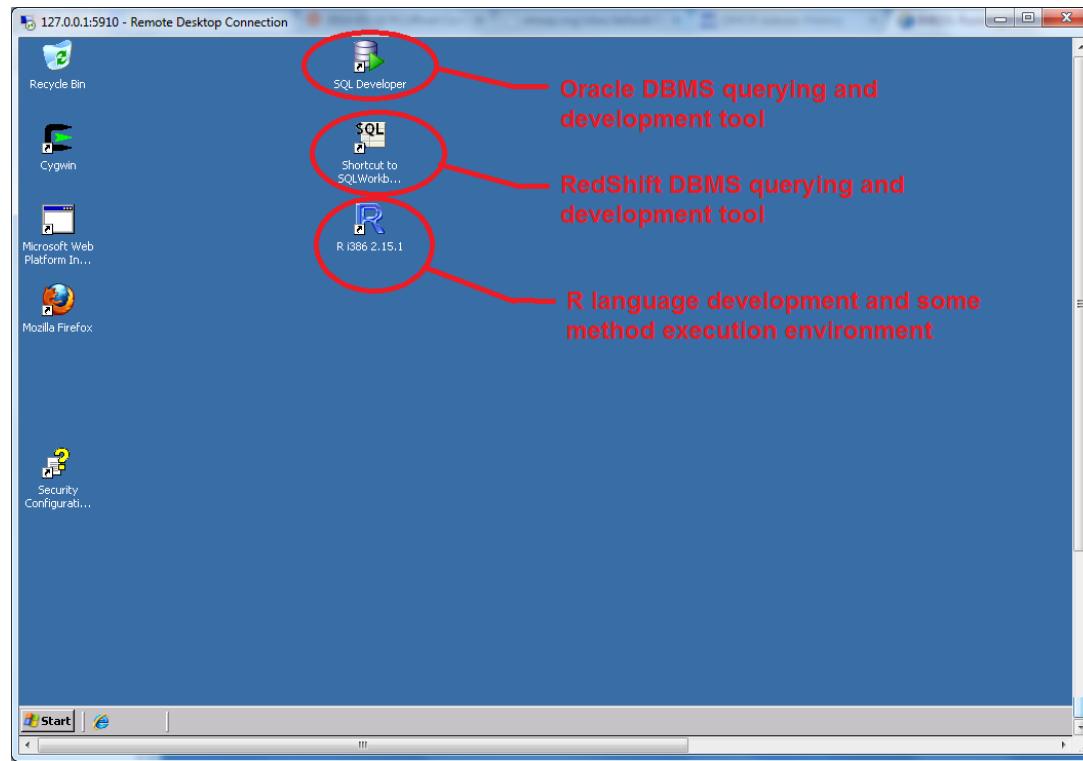


IMEDS Research Lab – Windows access

In Running Instances, click for Windows instance on RDC, PuTTY, or WinSCP option.



Remote Desktop Connect (RDC)



SQL – Counting persons

SET search_path **TO** ccae_cdm4;

/*

Count the total number of persons in the person table

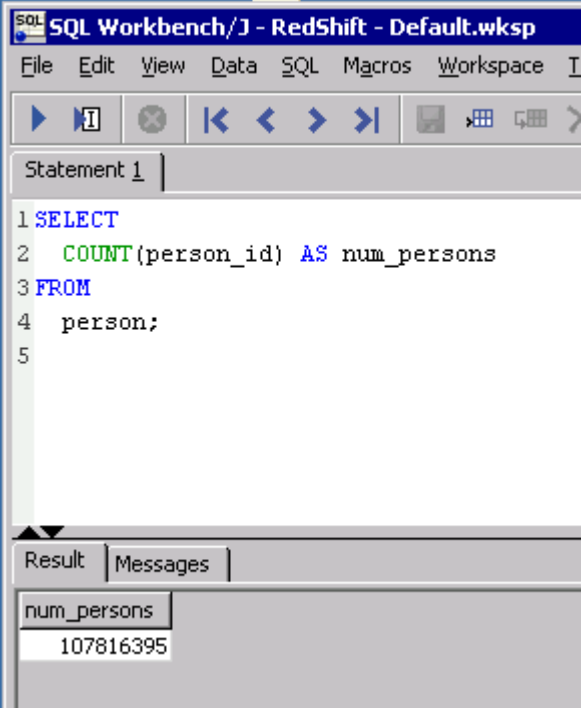
*/

SELECT

COUNT(person_id) **AS** num_persons

FROM

person;



The screenshot shows the SQL Workbench/J interface. The title bar reads "SQL Workbench/J - RedShift - Default.wksp". The menu bar includes "File", "Edit", "View", "Data", "SQL", "Macros", and "Workspace". The toolbar contains various navigation and execution icons. The main text area, labeled "Statement 1", contains the following SQL code:

```
1 SELECT
2   COUNT(person_id) AS num_persons
3 FROM
4   person;
5
```

Below the text area, there are tabs for "Result" and "Messages". The "Result" tab is active, displaying a table with one column named "num_persons" and one row containing the value "107816395".

Contact Information

- Email:
 - mkhayter@gmail.com
 - ewealebob@reaganudall.org
- Phone:
 - Cell Mark Khayter: 617-512-5967
 - Cell Emily Welebob: 703-508-6225

Lab Researchers Presenting Today

- Trevor Shaddox (UCLA)
- Jasmanda Wu (Sanofi)
- Richard Hansen (Auburn)
- Zach Shahn (Columbia)