OHDSI Technology Stack

Karthik Natarajan (Columbia University)
Mark Velez (Columbia University)
Lee Evans (LTS Computing LLC)
Taha Abdul-Basser (Columbia University)

Please copy the contents of the USB drive to your hard disk now.
You will need ~45GB free disk space available.

Introduction

Frank Defalco
Janssen

Lee Evans
LTS Computing LLC

Taha Abdul-Basser
Columbia University

Karthik Natarajan
Columbia University

Mark Velez
Columbia University

Introduction

Frank Defalco
Janssen

Lee Evans
LTS Computing LLC

Taha Abdul-Basser
Columbia University

Karthik Natarajan
Columbia University

Mark Velez
Columbia University

Jungmi Han
Columbia University

Agenda

Start time	Section
13:00	Introduction
	Atlas Overview
	Lab 1: OHDSI in a Box
14:00	Break
	Architectural Overview
	Lab 2: Installing OHDSI Platform
15:45	Break
	Lab 3: Using Achilles
	Wrap up and Q&A
17:00	End

Logistics

- This session is being recorded
- You should have a handout (for Lab 2) and flash drive
- Flash drives must be returned
- We welcome questions, but may table some

Introduction

- Course Objectives
  - Understand basic use cases of OHDSI stack
  - Understand components of OHDSI stack
  - Learn how to configure & deploy OHDSI stack
  - Prepare you to be an implementer or contributor

Introduction

Introduction

- How do we decide what to watch on Netflix?

Introduction

- How do Netflix users decide what to watch?

Introduction

- How do Netflix users decide what to watch?

About 75% of Netflix viewing is driven by the recommendation algorithm (source)

Introduction

- How do Netflix users decide what to watch?

These are **personalized** predictions based on **evidence** drawn from data collected as people use the system.

Introduction

- How do people decide what medical treatment to pursue?

Introduction

- How do people decide what medical treatment to pursue?
  - Using evidence from research, however indirect
- Prospective research: inherently stronger evidence, but does it scale? (e.g. cost, generalizability, research questions)
- Observational research: scales, but strength of evidence depends (e.g. sample size, methods)

Introduction

Netflix is just one of many popular systems today that aid decision-making by leveraging observational data.

What resources are available for medical decisions?

Introduction

Netflix is just one of many popular systems today that aid decision-making by leveraging observational data.

What resources are available for medical decisions?

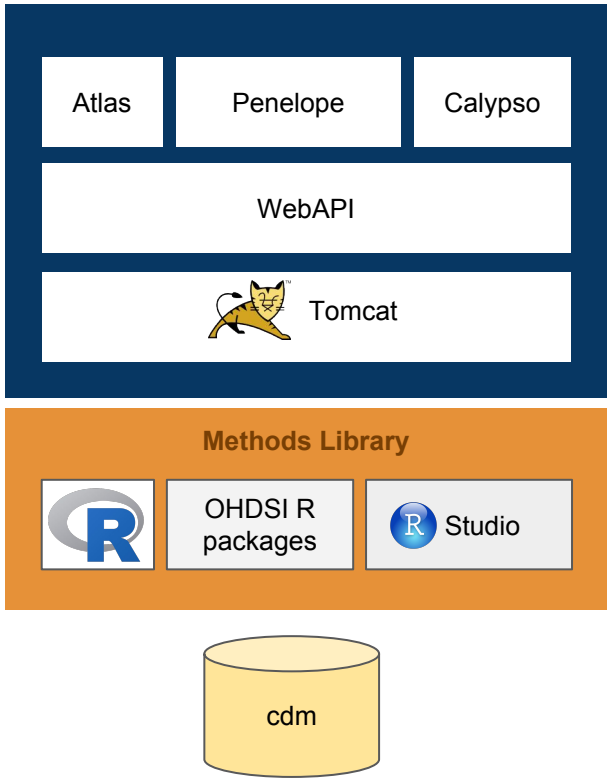
The full potential of observational health data has not been realized.

Introduction

- OHDSI mission: empower a community to collaboratively generate the evidence that promotes better health decisions and better care
- How? Open standards for open science:
  - Data (OMOP Common Data Model & Vocabulary)
  - Methods (CohortMethod, PatientLevelPrediction)
  - Tools (Atlas, Laertes)
    - characterize data
    - generate, evaluate, disseminate evidence

Introduction

- OHDSI technology stack
  - Database (CDM, Vocabulary)
  - REST services (WebAPI)
  - Packages (e.g. R CohortMethod)
  - Web applications (Atlas, Penelope)
- All open-source and on GitHub



Atlas Overview

Atlas Overview

- Viewing data source profiles with Atlas
- Managing Concept Sets/Vocabulary Searches
- Defining and generating Cohorts

ACHILLES

- Interactive platform to visualize data in CDM
  - patient demographics
  - prevalence of conditions, drugs and procedures
  - distribution of values for clinical observations

- https://github.com/OHDSI/Achilles

ATLAS

- Home
- Data Sources
- Vocabulary
- Concept Sets
- Cohorts
- Profiles
- Jobs
- Configuration
- Feedback

Data Sources

Data Sources | Reports

Demo_data_1_percent_synthetic_patients

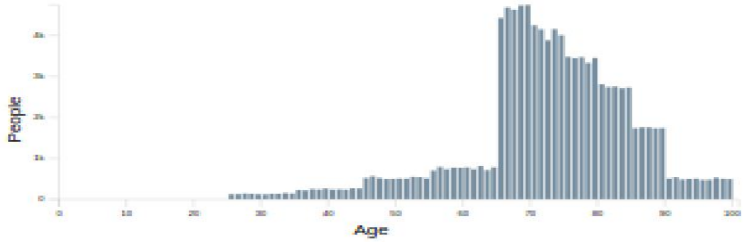
Dashboard

CDM Summary

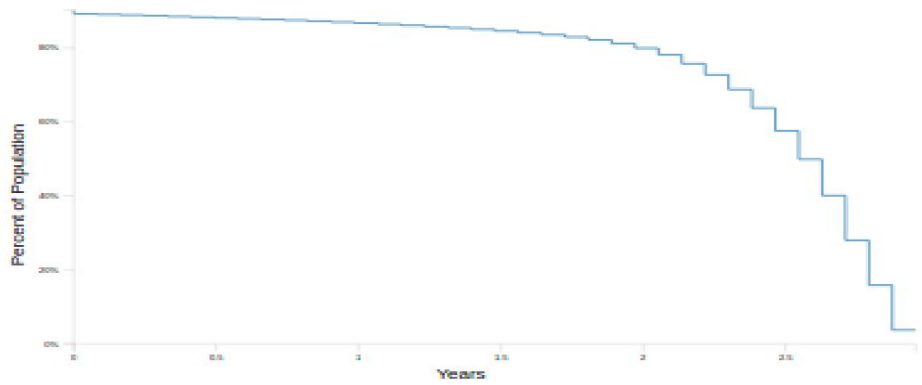
Source name: synpuf_1percent
Number of 116.35k persons:

Population by Gender

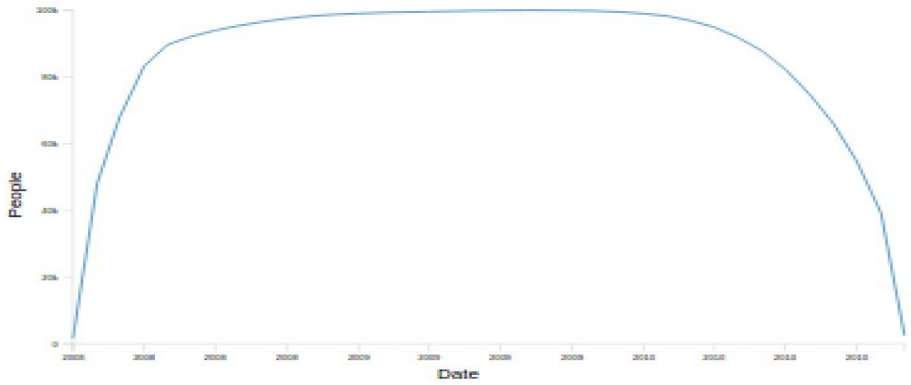
Age at First Observation



Cumulative Observation



Persons With Continuous Observation By Month



ATLAS

Data Sources

- Home
- Data Sources
- Vocabulary
- Concept Sets
- Cohorts
- Profiles
- Jobs
- Configuration
- Feedback

Data Sources | Reports

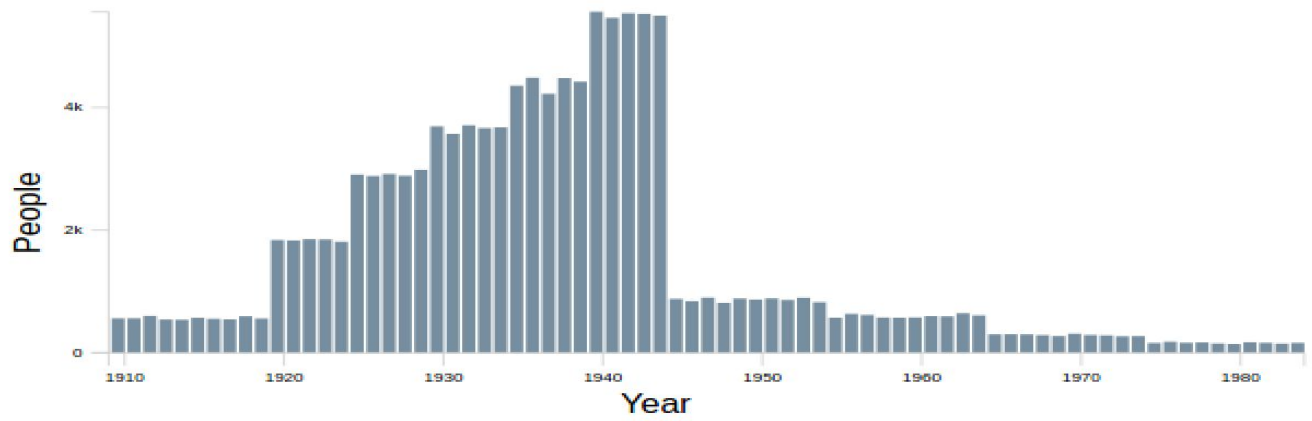
Demo_data_1_percent_synthetic_patients

Person

Person Summary

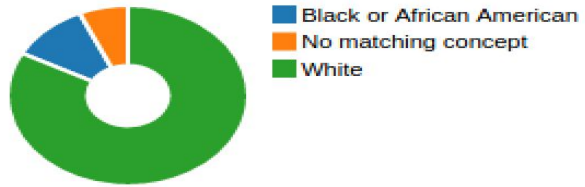
Source name: synpuf_1percent
Number of persons: 116.35k

Year of Birth

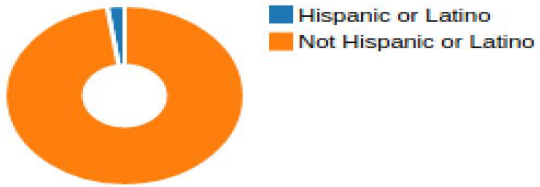


Population by Gender

Population by Race



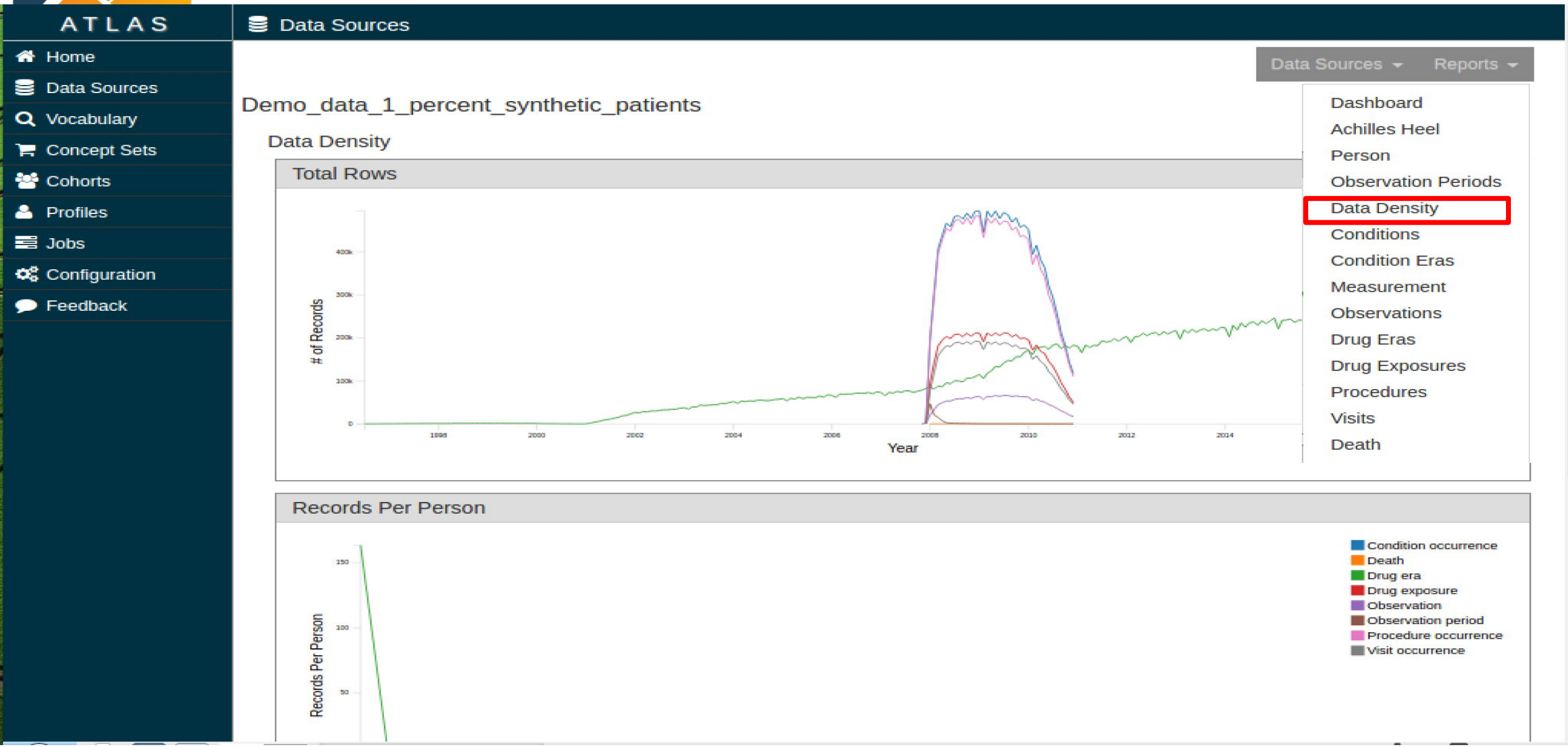
Population by Ethnicity

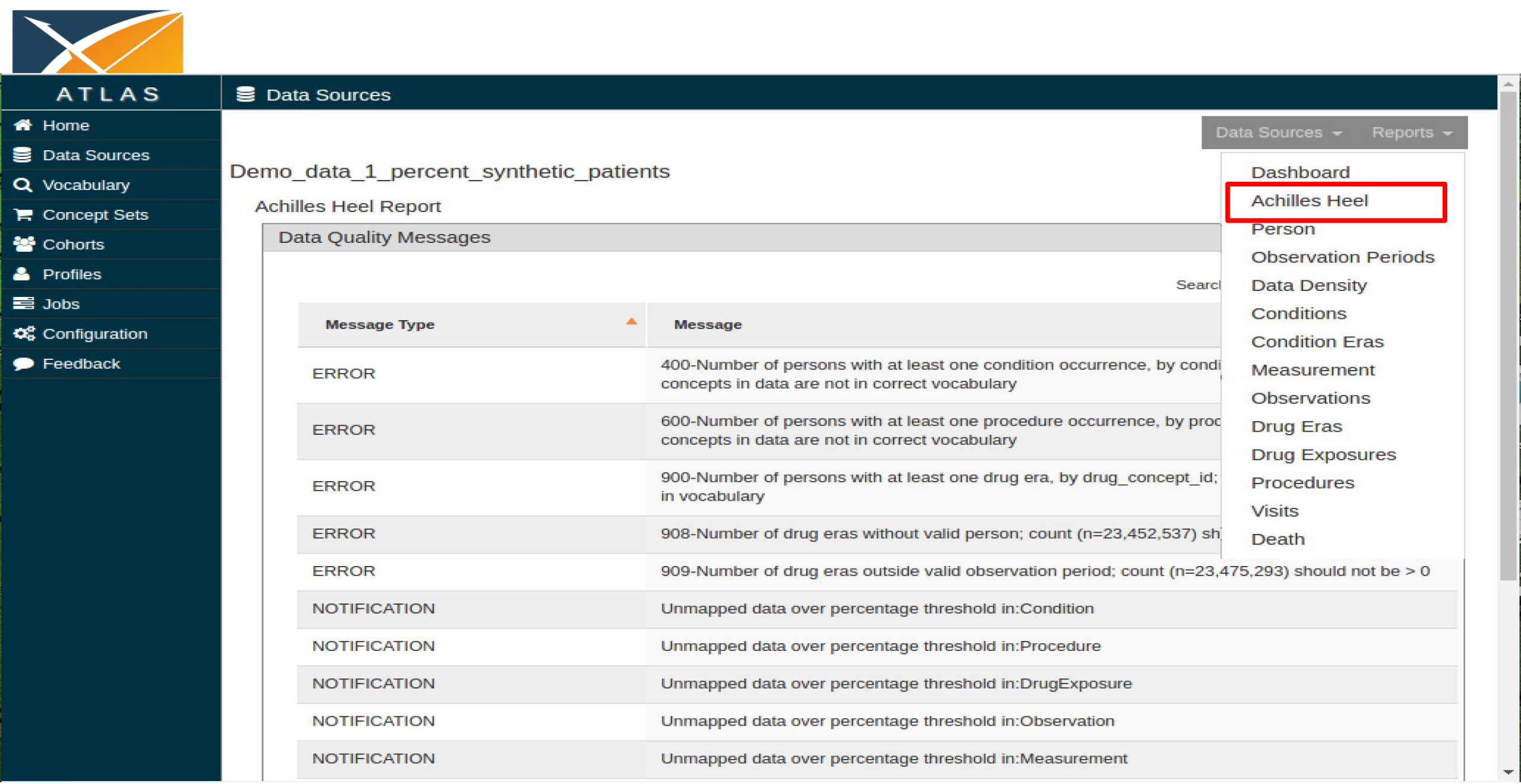


ATLAS - Google ... EN 14:10


ATLAS
Data Sources
Home
Data Sources
Vocabulary
Concept Sets
Cohorts
Profiles
Jobs
Configuration
Feedback
Data Sources
Reports
Demo_data_1_percent_synthetic_patients
Conditions
Condition Prevalence
Treemap
Table
Box Size: Prevalence, Color: Records per Person (Blue to Orange = Low to High), Use Ctrl-Click to Zoom, Alt-Click to Reset Zoom
Dashboard
Achilles Heel
Person
Observation Periods
Data Density
Conditions
Condition Eras
Measurement
Observations
Drug Eras
Drug Exposures
Procedures
Visits
Death


ATLAS
Data Sources
Home
Data Sources
Vocabulary
Concept Sets
Cohorts
Profiles
Jobs
Configuration
Feedback
Data Sources
Reports
Demo_data_1_percent_synthetic_patients
Conditions
Condition Prevalence
Treemap
Table
Vascular disorders
Arteriosclerosis, stenosis, vascular insufficiency and necrosis
Non-site specific necrosis and vascular insufficiency NEC
Arteriosclerosis
Coronary arteriosclerosis
Prevalence: 44.07%
Number of People: 51,278
Records per Person: 2.51
Box Size: Prevalence, Color: Records per Person (Blue to Orange = Low to High), Use Ctrl-Click to Zoom, Alt-Click to Reset Zoom
Coronary arteriosclerosis
Condition Prevalence
MALE
FEMALE
Age Decile
0-9
10-19
20-29
30-39
40-49
50-59
60-69
70-79
80-89
90-99


ATLAS
Data Sources
Home
Data Sources
Vocabulary
Concept Sets
Cohorts
Profiles
Jobs
Configuration
Feedback
Data Sources
Reports
Demo_data_1_percent_synthetic_patients
Data Density
Total Rows
# of Records
Year
Dashboard
Achilles Heel
Person
Observation Periods
Data Density
Conditions
Condition Eras
Measurement
Observations
Drug Eras
Drug Exposures
Procedures
Visits
Death
Records Per Person
Records Per Person
Condition occurrence
Death
Drug era
Drug exposure
Observation
Observation period
Procedure occurrence
Visit occurrence


ATLAS
Home
Data Sources
Vocabulary
Concept Sets
Cohorts
Profiles
Jobs
Configuration
Feedback
Data Sources
Data Sources
Reports
Demo_data_1_percent_synthetic_patients
Achilles Heel Report
Data Quality Messages
Dashboard
Achilles Heel
Person
Observation Periods
Data Density
Conditions
Condition Eras
Measurement
Observations
Drug Eras
Drug Exposures
Procedures
Visits
Death
Message Type
Message
ERROR
400-Number of persons with at least one condition occurrence, by condi
concepts in data are not in correct vocabulary
ERROR
600-Number of persons with at least one procedure occurrence, by proc
concepts in data are not in correct vocabulary
ERROR
900-Number of persons with at least one drug era, by drug_concept_id;
in vocabulary
ERROR
908-Number of drug eras without valid person; count (n=23,452,537) sh
ERROR
909-Number of drug eras outside valid observation period; count (n=23,475,293) should not be > 0
NOTIFICATION
Unmapped data over percentage threshold in:Condition
NOTIFICATION
Unmapped data over percentage threshold in:Procedure
NOTIFICATION
Unmapped data over percentage threshold in:DrugExposure
NOTIFICATION
Unmapped data over percentage threshold in:Observation
NOTIFICATION
Unmapped data over percentage threshold in:Measurement

Queries Can Be Automated

Definition | Concept Sets | Generation | Reporting | Explore | **Export**

Print Friendly | JSON | SQL

Initial Event Cohort

People having any of the following:

- a drug exposure of Warfarin[2]
  - for the first time in the person's history
  - with age >= 65

with continuous observation of at least 180 days prior and 0 days after event index date, and limit initial events to: **earliest event per person.**

For people matching the Primary Events, include:
People having all of the following criteria:

- at least 1 occurrences of a condition occurrence of Atrial fibrillation[1]
  occurring between all days Before and 1 days Before event index date

Limit cohort of initial events to: **earliest event per person.**
Limit qualifying cohort to: **earliest event per person.**
No end date strategy selected. By default, the cohort end date will be the end of the observation period that contains the index event.

Appendi

Definition | Concept Sets | **Generation** | Reporting | Explore | Export

1. Atrial

Available CDM Sources

	Source Name	Generation Status	Distinct People
▶ Generate	OHDSI CDM V5 Database	COMPLETE	8207

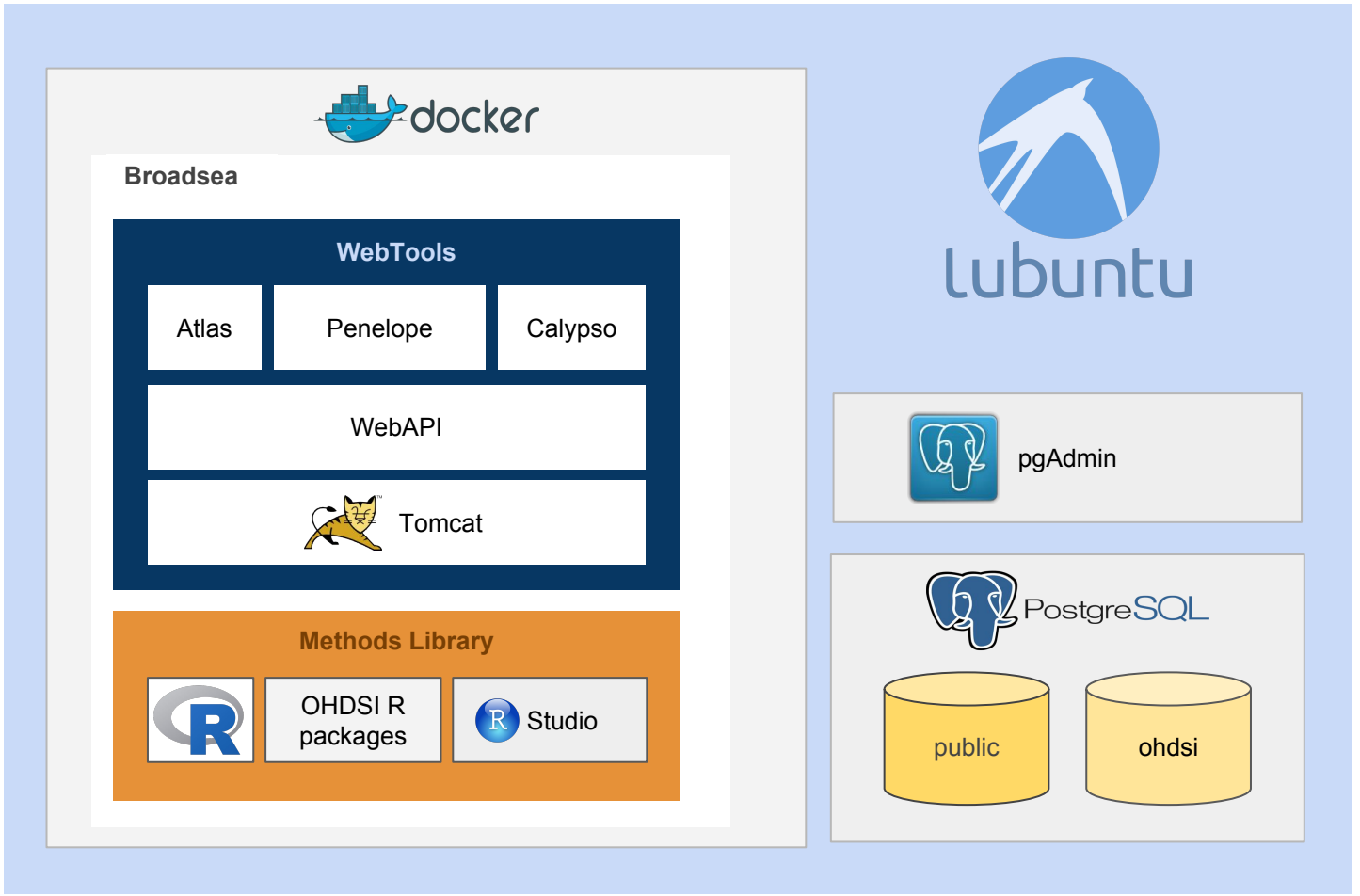
2. Warfa

1310149	Warfarin		Drug	RxNorm	NO	YES	NO

Lab 1

OHDSI in a Box

Lab 1: OHDSI in a Box

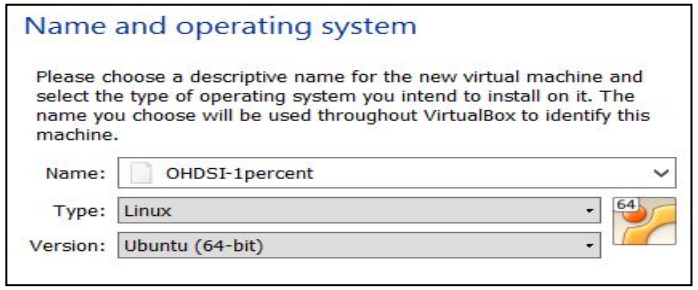


OHDSI in a Box: Setup

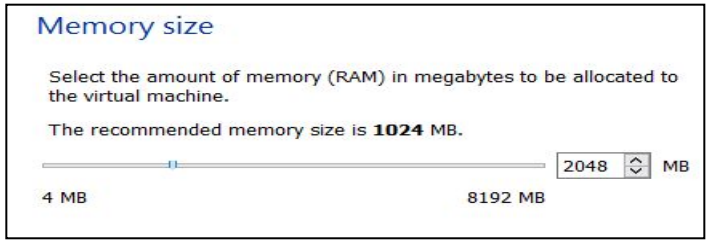
1. Open VM VirtualBox Manager

2. Click on New

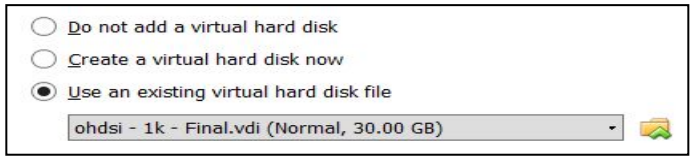
3.

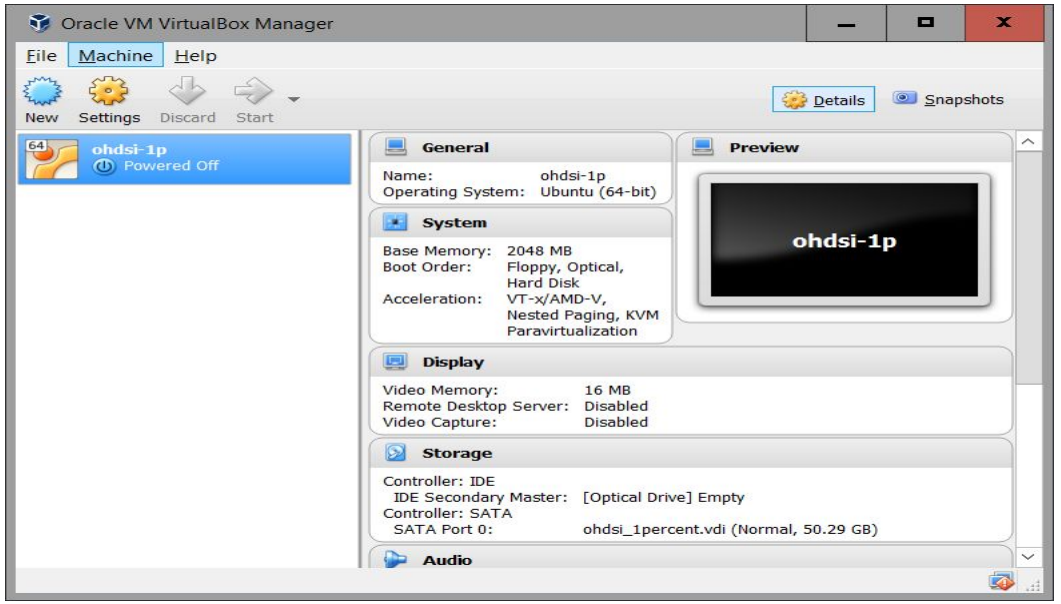


4.

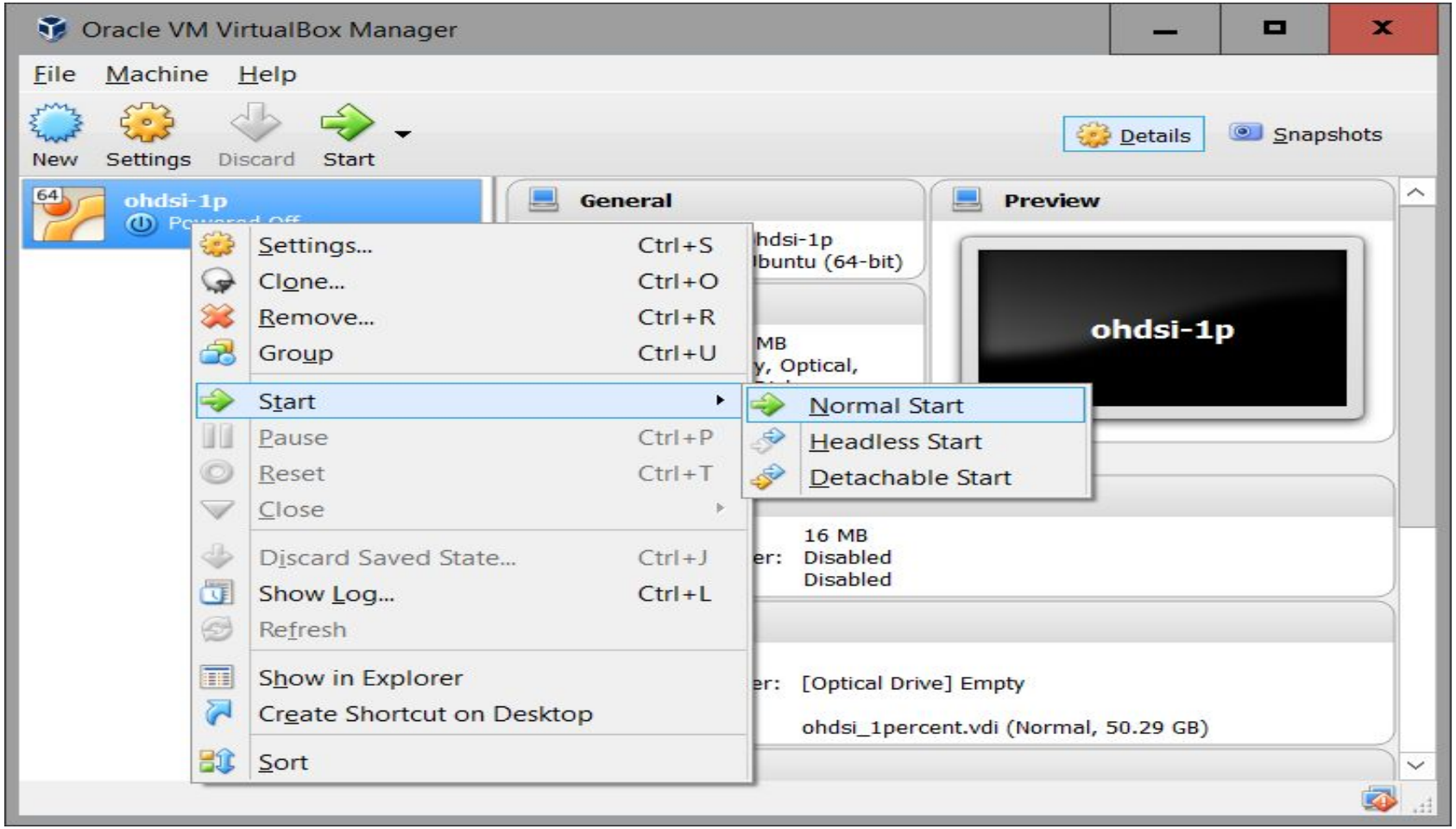


5.

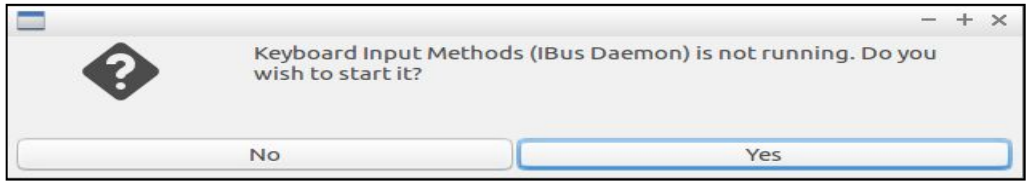


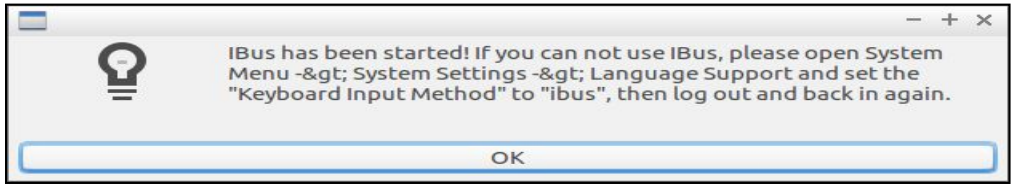


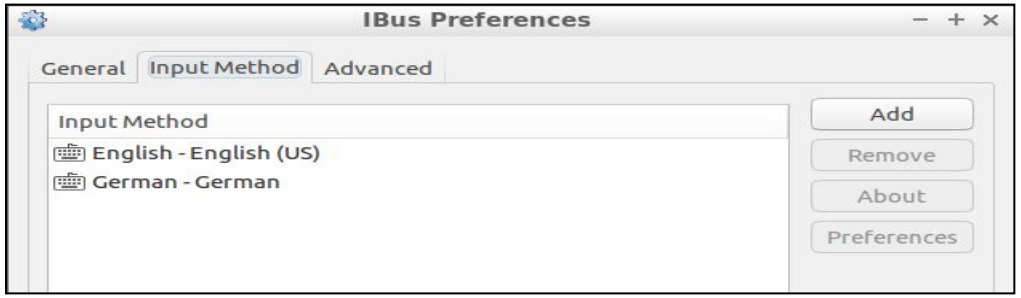
OHDSI in a Box: Start Up

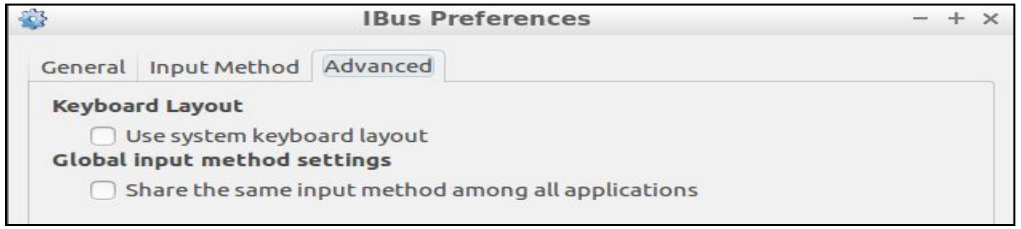


OHDSI in a Box: International Keyboards

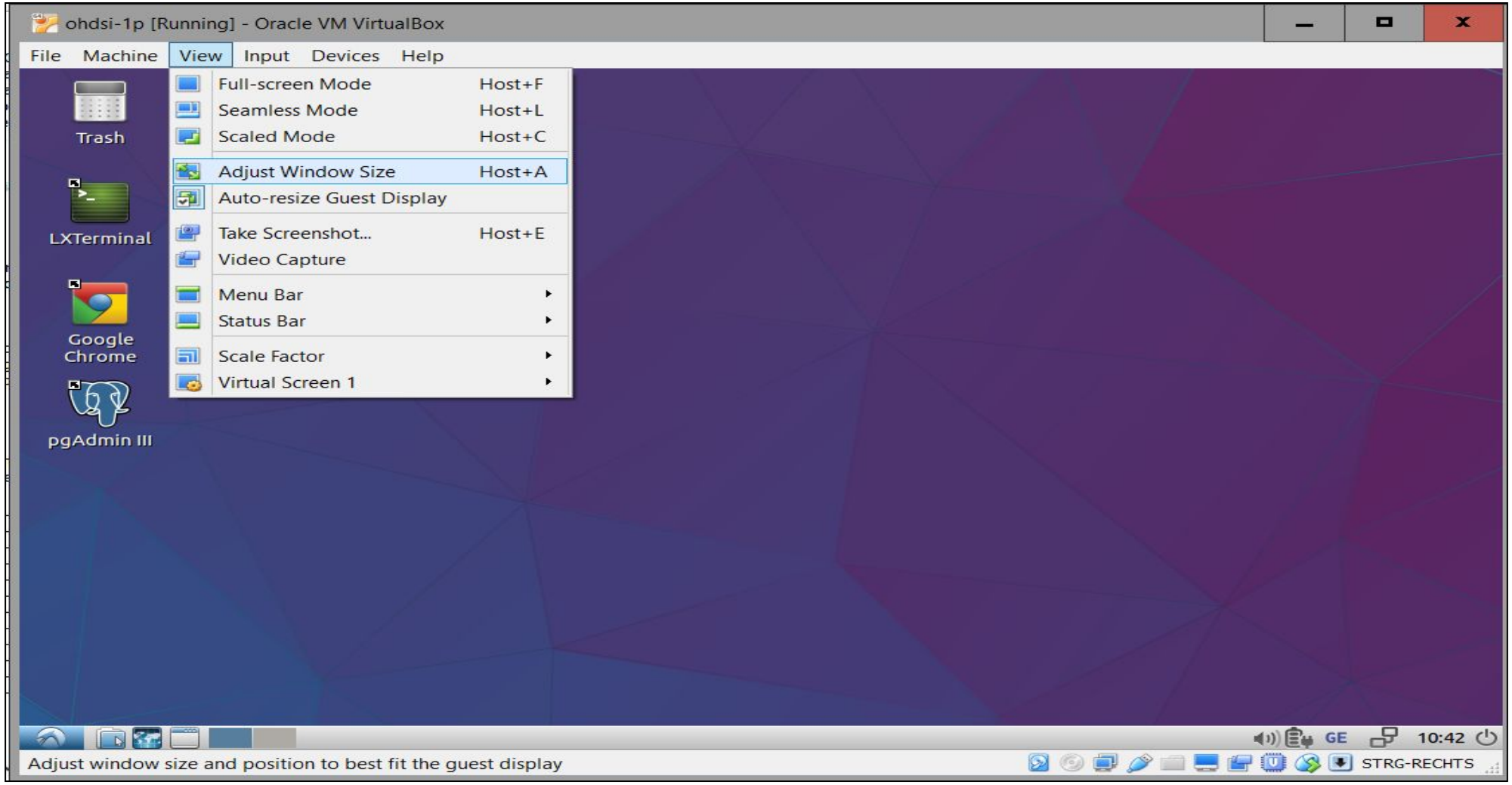




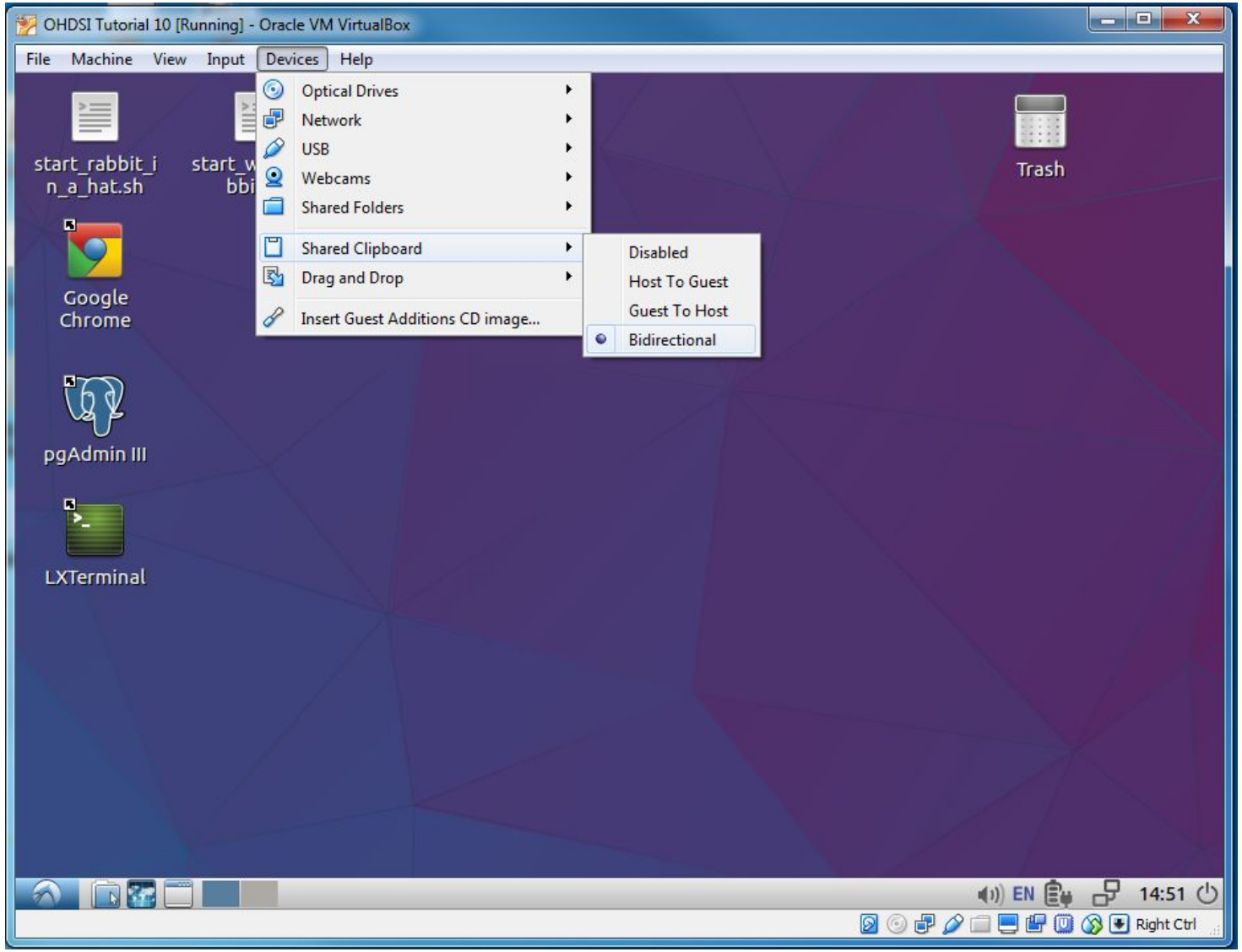




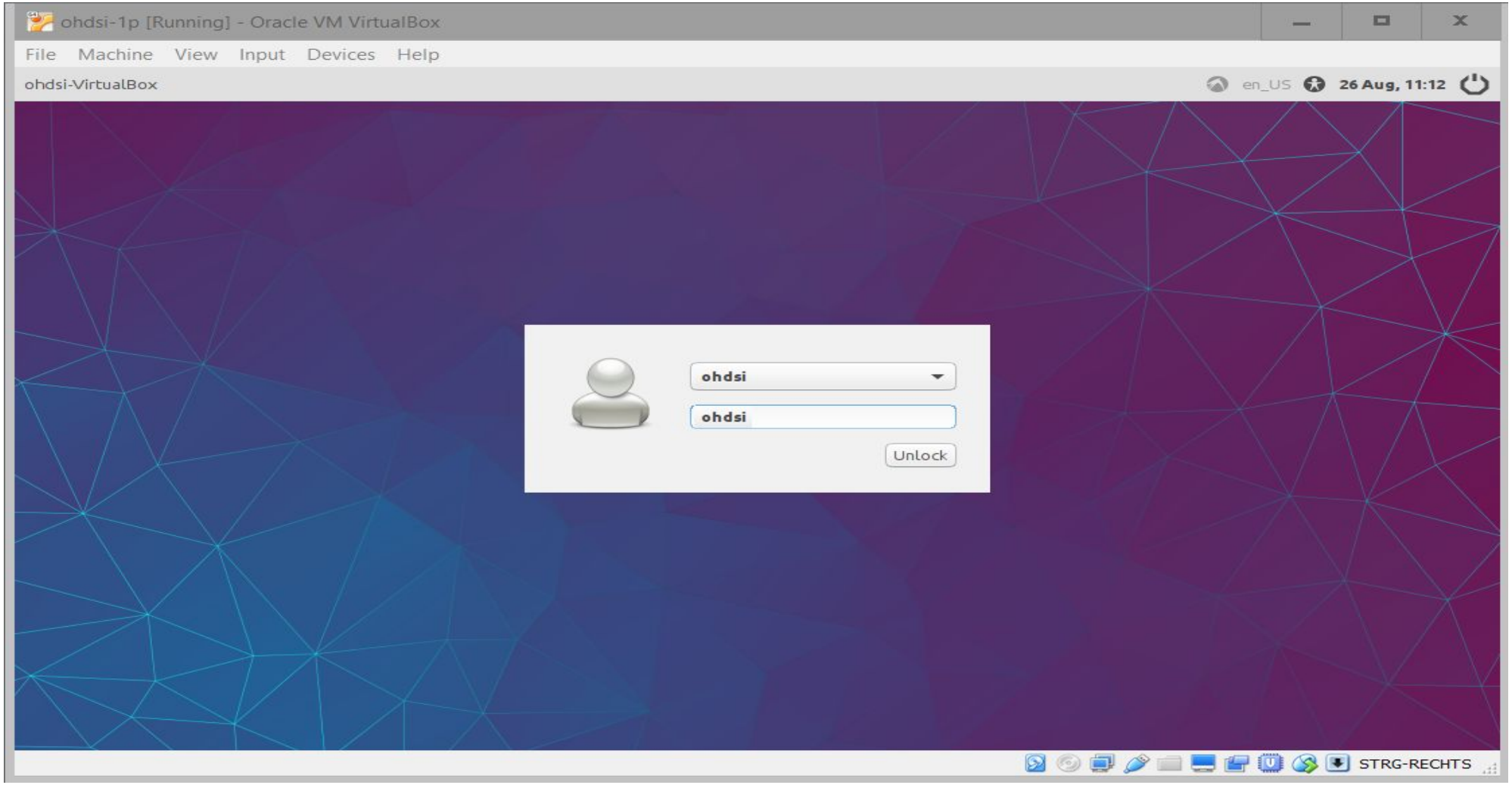
OHDSI in a Box: Adjust Resolution



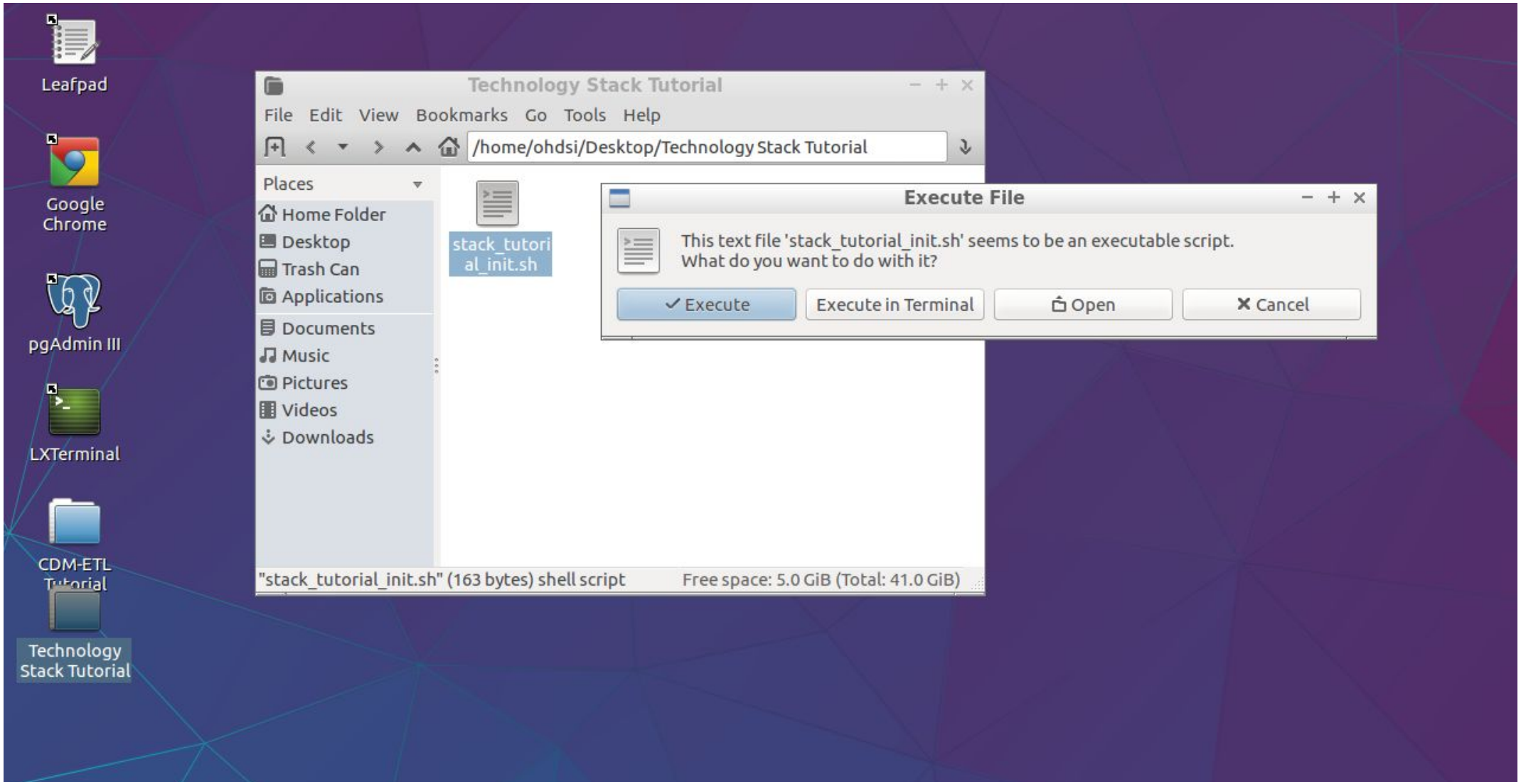
OHDSI in a Box – Clipboard



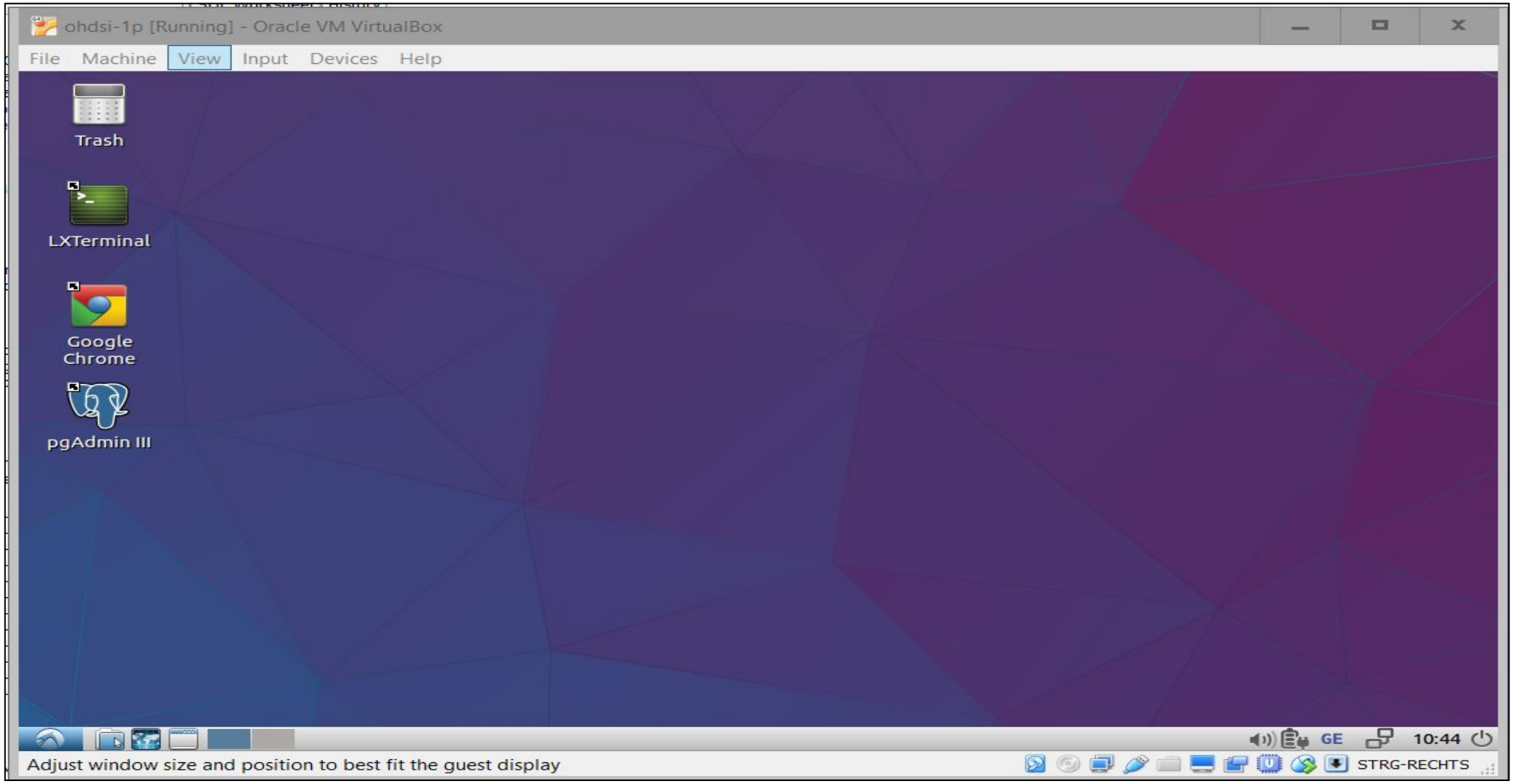
OHDSI in a Box: Timeout



Initialize VM



OHDSI in a Box – Ready



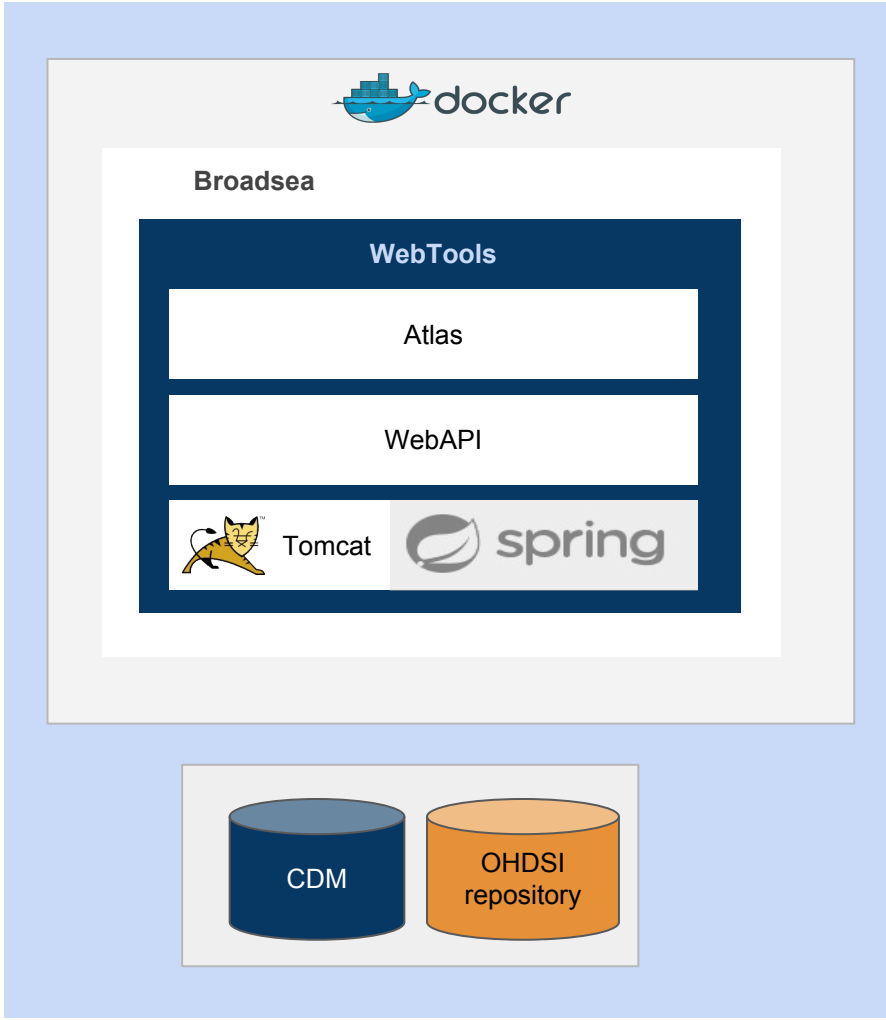
BREAK

resume at {14:15}

Architectural Overview

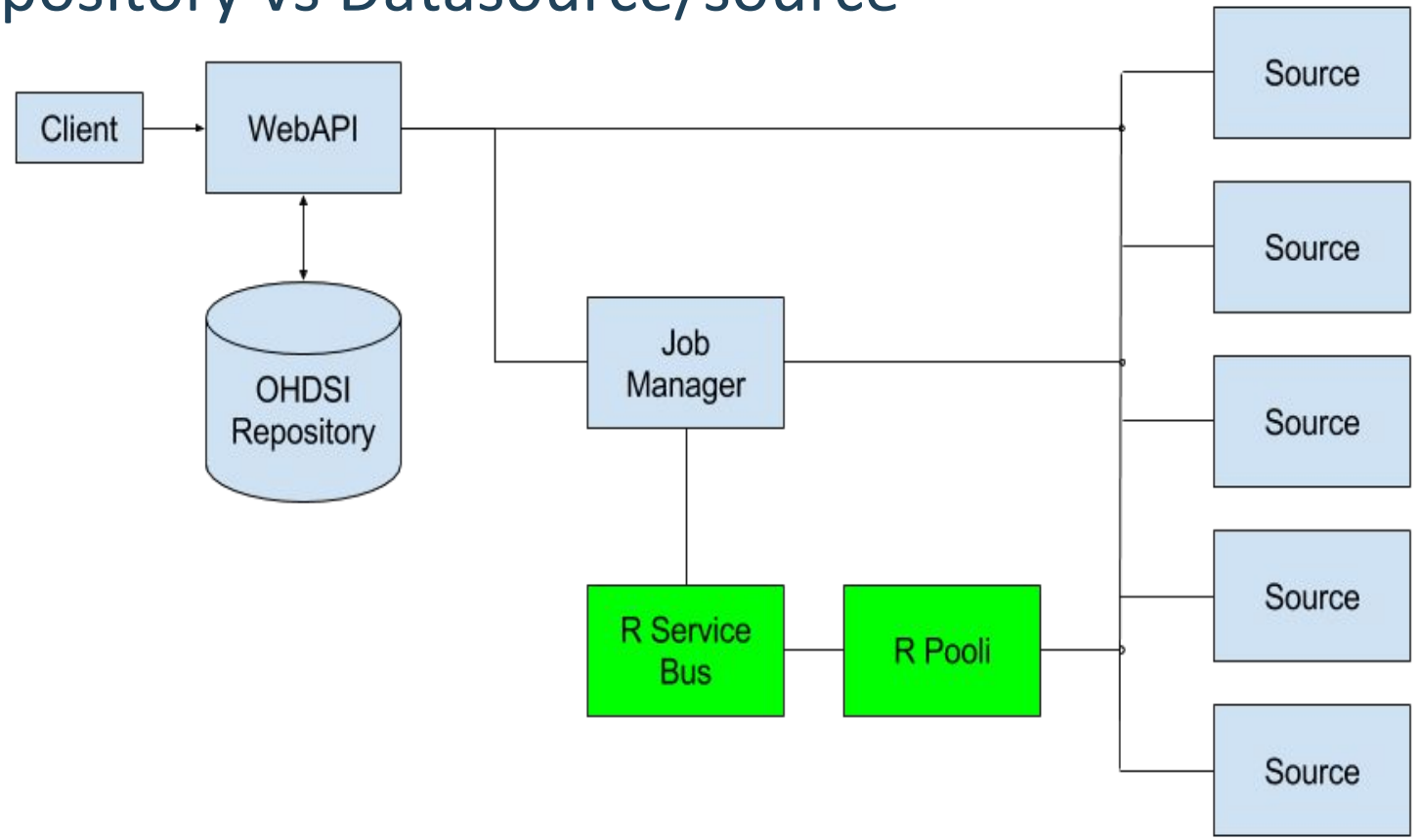
Components

- Atlas
  - HTML5 web application
- WebAPI
  - REST web services module
- OMOP CDM/R
  - patient-centric observational healthcare data model
  - Multiple RDBMS
    - PostgresQL, SQL Server, Oracle



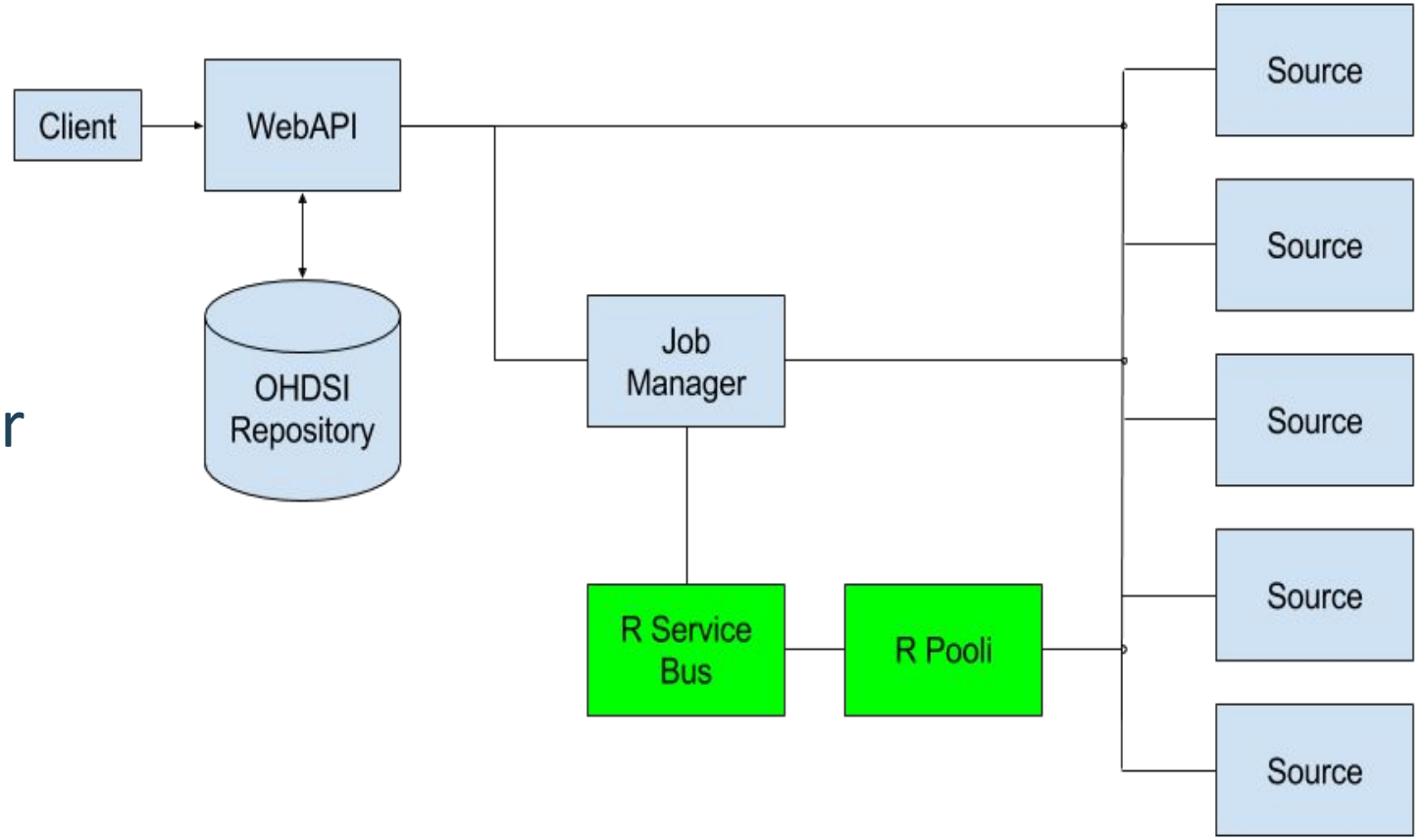
OHDSI Repository

- CDM vs OHDSI schema/repository vs Datasource/source
  - *Source* table
- Daimon
  - *Source_Daimon*
    - CDM
    - Vocabulary
    - Results
    - Evidence
- Multi-DS configurations
- Spring Batch table
- Results schema



WebAPI

- WebAPI
  - REST web services (JSON/HTTP)
  - Java application using Spring Boot
  - Java application server



Atlas

- Served by web server (HTTP server)
- Runs in web browser (supported: Chrome)
- HTML5 application (HTML5, JavaScript, CSS)
  - Built with various popular JS libraries/frameworks
  - Cascading Style Sheets (CSS) - Bootstrap
  - Asynchronous module definition (AMD) - Require JS
  - Model-View-View-Model - Knockout JS
    - Components
      - Views
      - Templates

Lab 2

Deploy, Configure & Start the Broadsea containers with a Postgres CDM v5 database

OHDSI Broadsea Acknowledgements

Lee Evans
Developer of the Broadsea OHDSI WebAPI & Web Tools Docker Container
http://www.ltscomputingllc.com/

Marc Suchard
Developer of the Broadsea OHDSI Web Methods (Rstudio / R) Docker Container

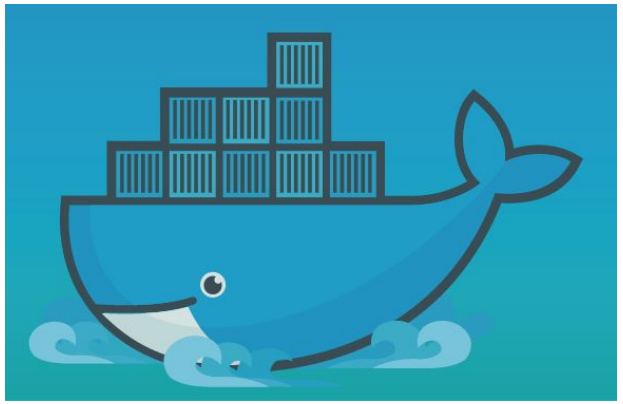
Jon Duke
Initial idea for the OHDSI WebAPI & Web Apps Docker Container
Tested Broadsea with Regenstreif Oracle OMOP CDM Version 5

What is Broadsea?

Deploy the full OHDSI technology stack
& connect it to your OMOP CDM Version 5 Database

Broadsea = +

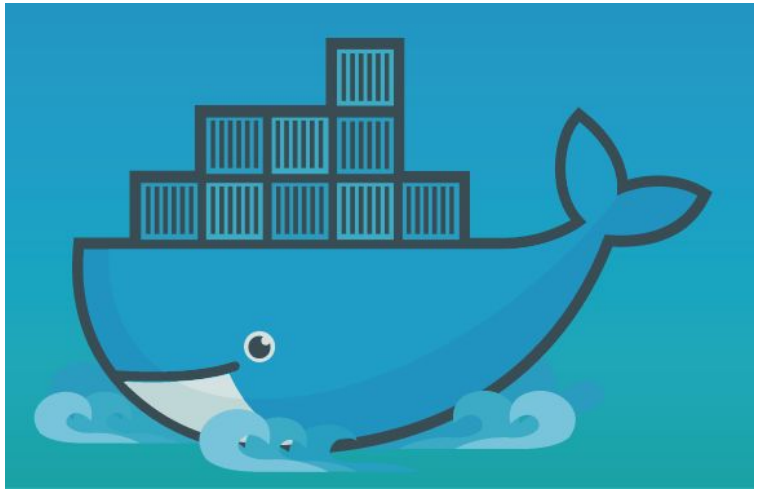
OHDSI R packages
& Web Apps

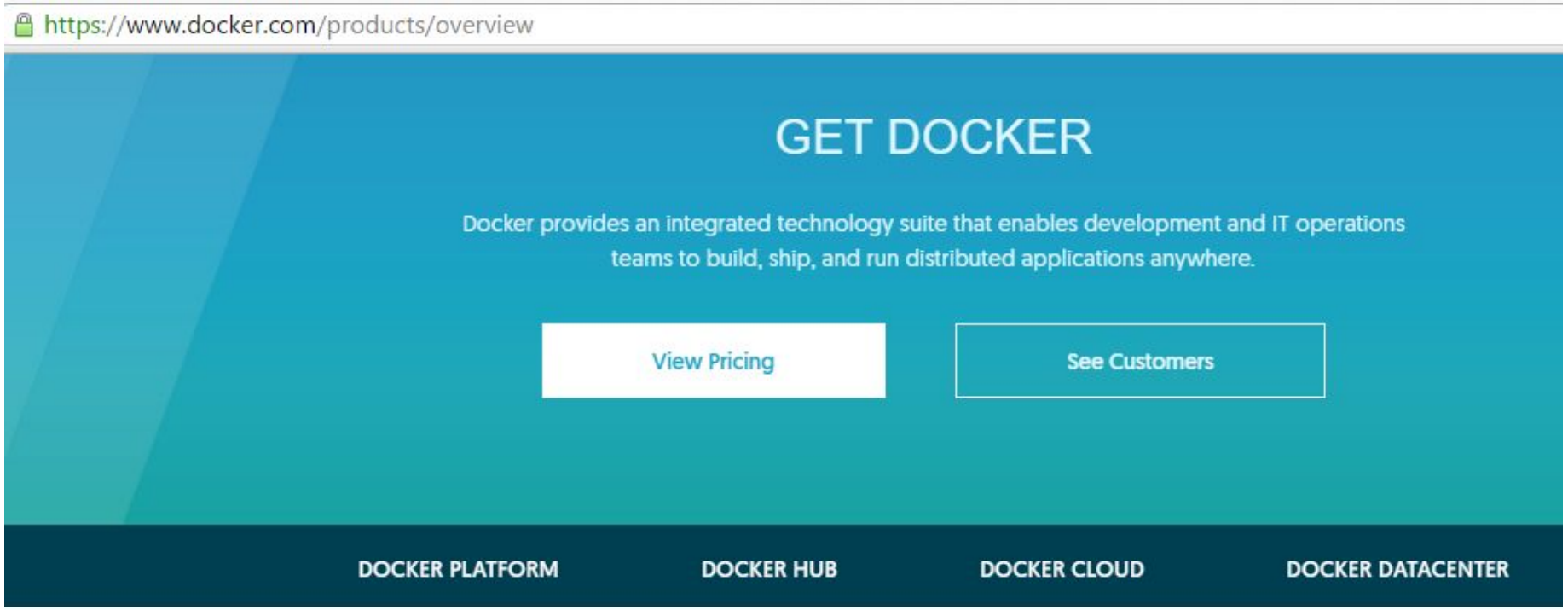


Deployed as
Docker Containers

What is a Docker container?

- An application packaged with all it’s dependencies
- Built from a configuration file (Dockerfile)
- Download container images from Docker Hub website
- Docker engine must be installed on your machine
- Cross-platform
- Free to install and use
- See this link for more info:
  - https://www.docker.com/what-docker





INSTALL THE PLATFORM

Install Docker with easy to use installers for the major desktop and cloud platforms.

MAC WINDOWS LINUX

The OHDSI Broadsea Docker containers

Note. The containers are large – around 3 GB and 275 MB
and take some time to download from the DockerHub web site

ohdsi/broadsea-webmethods	ohdsi/webapi-webtools
• RStudio server • R • OHDSI R packages	• WebAPI • Atlas • Calypso

Configure & Run Broadsea

```
# Broadsea OHDSI RStudio & R packages and OHDSI WebAPI & web applications

version: '2'

services:

  ohdsi-rstudio-r:
    image: ltscomputingllc/ohdsi-rstudio-r
    ports:
      - "8787:8787"
    volumes:
      - ./home/rstudio:/home/rstudio

  ohdsi-webapi-webapps:
    image: ltscomputingllc/ohdsi-webapi-webapps:1.1.1
    ports:
      - "8080:8080"
    volumes:
     - .:/tmp/drivers/:ro
     - .:/tmp/achilles-data-reports/:ro
    environment:
      - WEBAPI_URL=http://192.168.99.100:8080
      - datasource_driverClassName=org.postgresql.Driver
      - datasource_url=jdbc:postgresql://192.168.1.8:5556/postgres
      - datasource_username=postgres
      - datasource_password=abc123
      - spring.jpa.properties.hibernate.default_schema=ohdsi
      - flyway_datasource_driverClassName=org.postgresql.Driver
      - flyway_datasource_url=jdbc:postgresql://192.168.1.8:5556/postgres
      - flyway_schemas=ohdsi
      - flyway.placeholders.ohdsSchema=ohdsi
      - flyway_datasource_username=postgres
      - flyway_datasource_password=abc123
      - flyway.locations=classpath:db/migration/postgresql
```

- Edit docker-compose.yml file
  - WebAPI URL
  - DB connection info
- docker-compose run –d
- Configure source & source_daimon tables with CDM connections
- docker-compose down
- docker-compose run -d

How can I use Broadsea at my site?

- See the Documentation at https://github.com/OHDSI/Broadsea
- Tested on Windows, Mac OS X, Linux, PostgreSQL, SQL Server, Oracle
- Example docker-compose.yml files available in the github repo
- Docker engine must be installed on your machine
- OMOP CDM Version 5 database required

Broadsea will continue to evolve over time as new OHDSI web tools & R packages are developed.

Lab 3

Using Achilles

Using Achilles

- Objective
  - Run Achilles to generate summarization and characterization data from OMOPv5 CDM

Using Achilles

- R modules to generate summarization/characterization data for OMOP v5 CDM
  - patient demographics
  - prevalence of conditions, drugs and procedures
  - distribution of values for clinical observations

- Docs and code available on GitHub

Using Achilles

- Open RStudio

```
http://localhost:8787
```

Using Achilles

- Load dependencies

```
library('devtools')
library('Achilles')
```

Using Achilles

Set up connection

```
connectionDetails<- createConnectionDetails(dbms
= "postgresql", user = “ohdsi”, password =
“ohdsi”, server = “localhost”)
```

Using Achilles

- Run Achilles

```
achilles(connectionDetails,\
cdmDatabaseSchema="public",\
resultsDatabaseSchema="ohdsi",\
sourceName="ohdsi_ds")
```

Lab 3: Using Achilles

- Export results to JSON file

```
exportToJson(connectionDetails,"public",
"ohdsi")
```

Lab 3: Using Achilles

- Test and validate

  `Atlas > Datasources`

ATLAS

- Home
- Data Sources
- Vocabulary
- Concept Sets
- Cohorts
- Profiles
- Jobs
- Configuration
- Feedback

Data Sources

Data Sources | Reports

Demo_data_1_percent_synthetic_patients

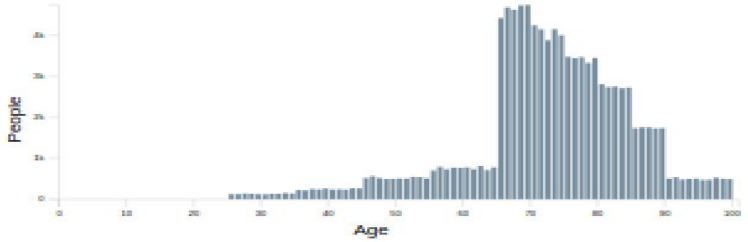
Dashboard

CDM Summary

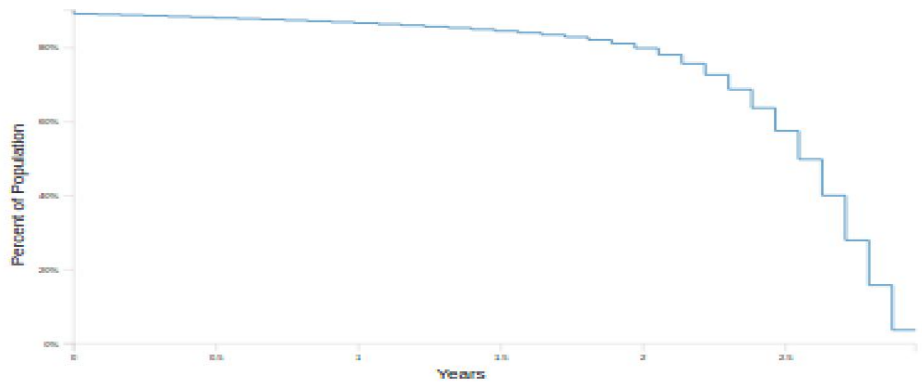
Source name: synpuf_1percent
Number of 116.35k persons:

Population by Gender

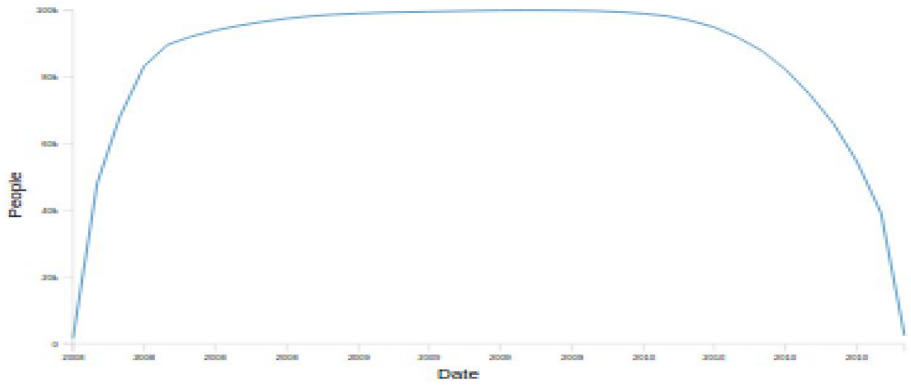
Age at First Observation



Cumulative Observation



Persons With Continuous Observation By Month



ATLAS

Data Sources

- Home
- Data Sources
- Vocabulary
- Concept Sets
- Cohorts
- Profiles
- Jobs
- Configuration
- Feedback

Data Sources | Reports

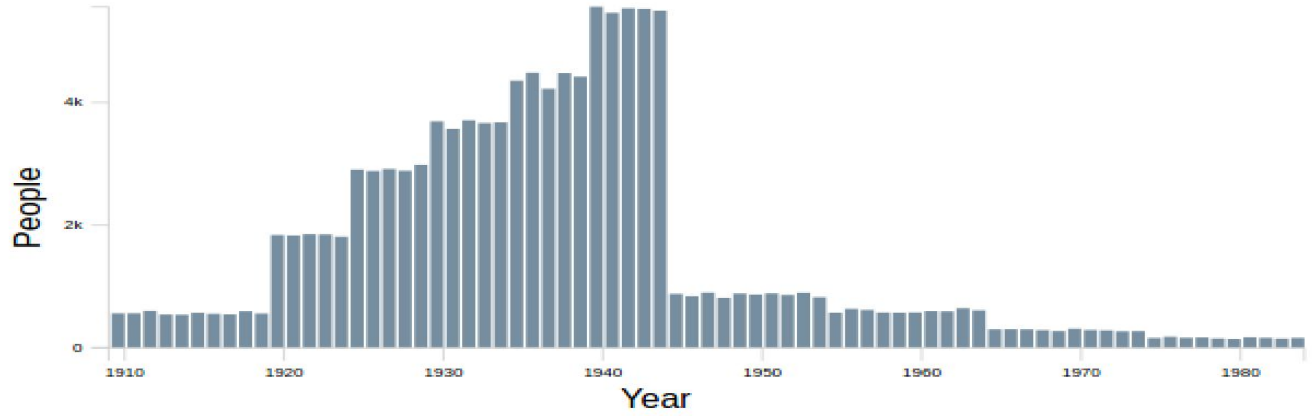
Demo_data_1_percent_synthetic_patients

Person

Person Summary

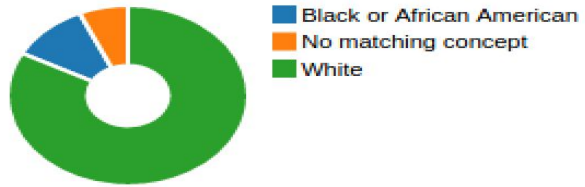
Source name: synpuf_1percent
Number of persons: 116.35k

Year of Birth

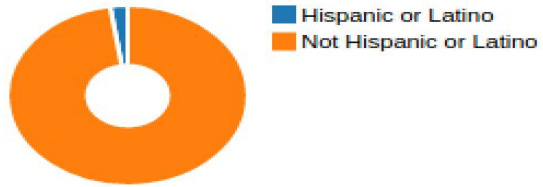


Population by Gender

Population by Race



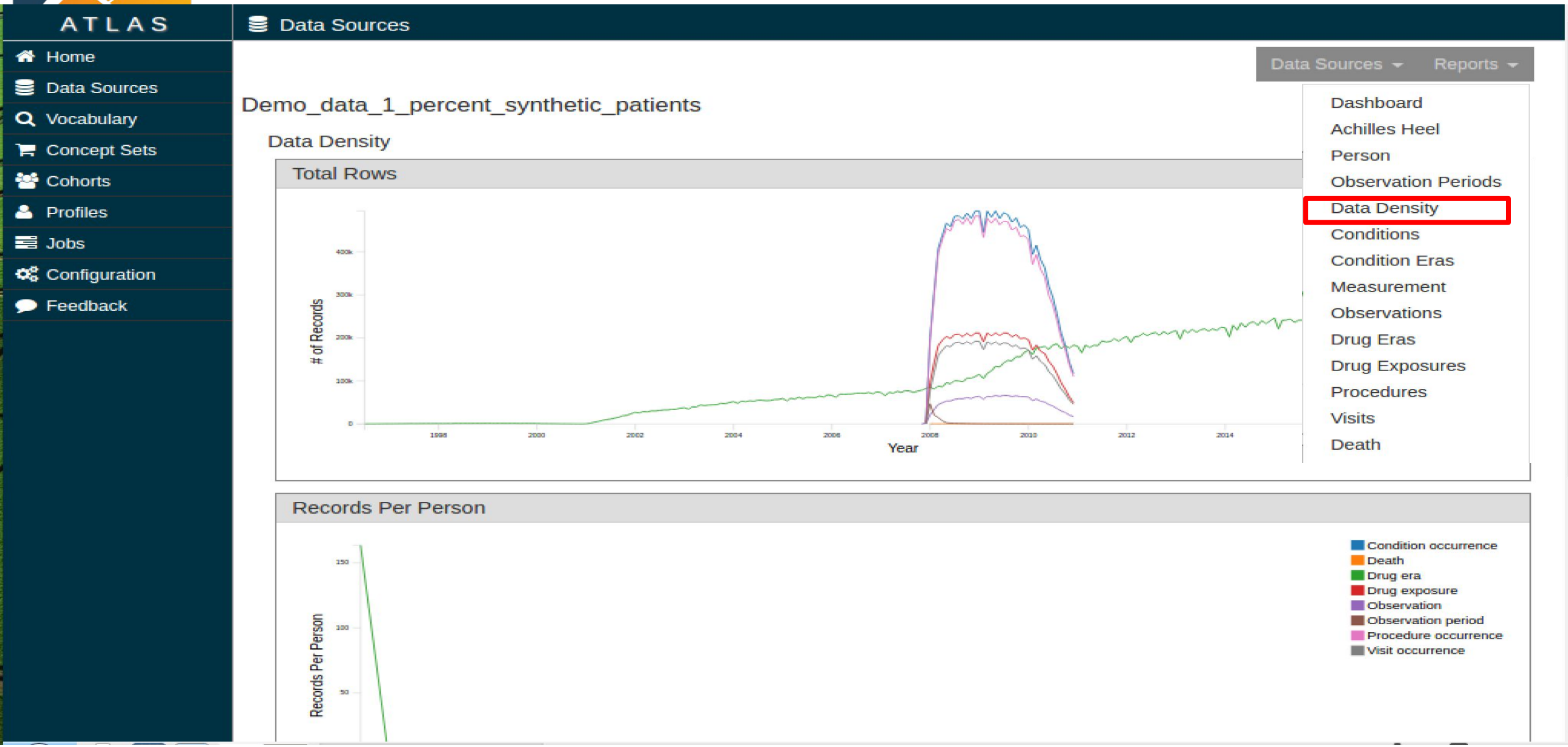
Population by Ethnicity

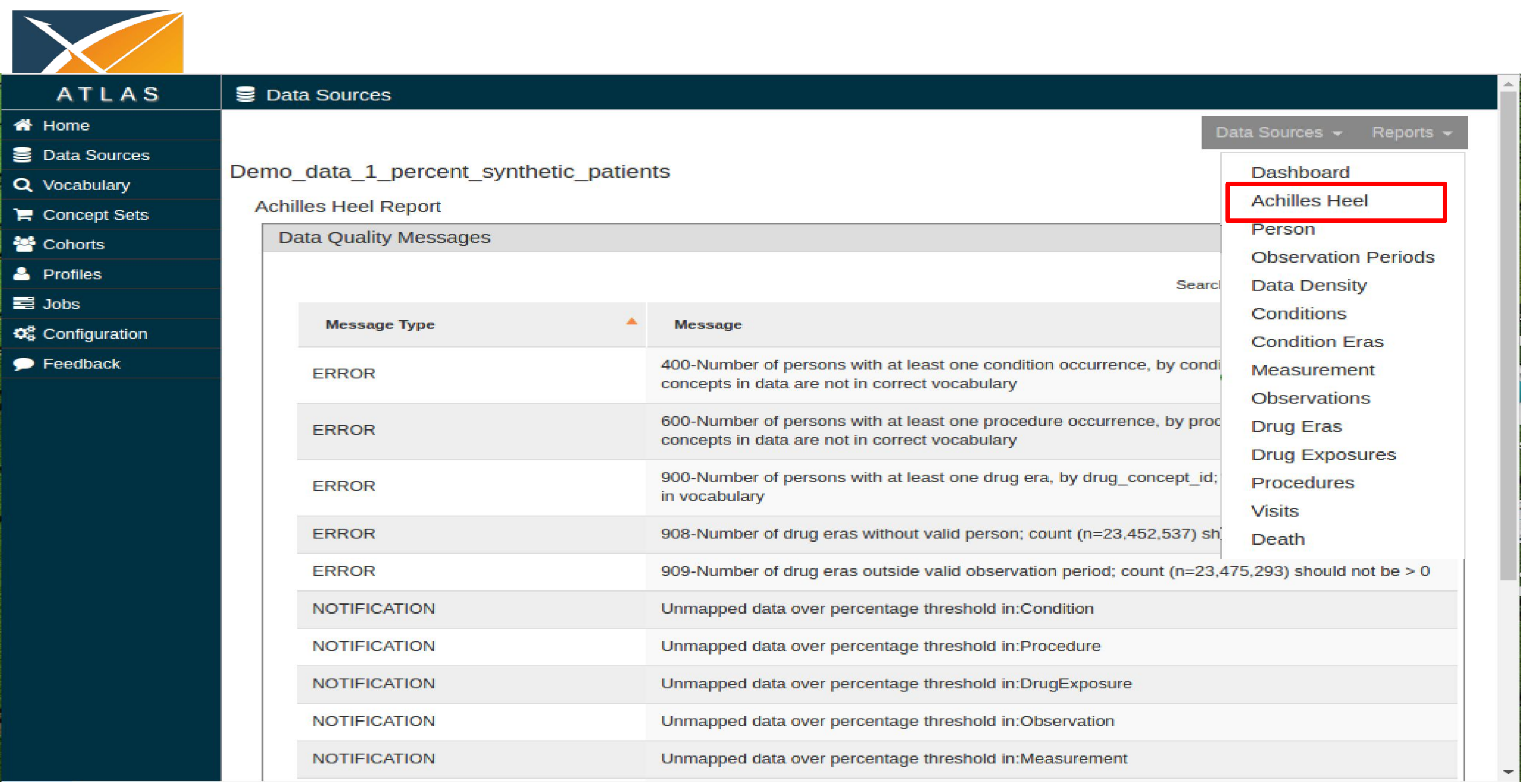


ATLAS - Google ... EN 14:10


ATLAS
Data Sources
Home
Data Sources
Vocabulary
Concept Sets
Cohorts
Profiles
Jobs
Configuration
Feedback
Data Sources
Reports
Demo_data_1_percent_synthetic_patients
Conditions
Condition Prevalence
Treemap
Table
Box Size: Prevalence, Color: Records per Person (Blue to Orange = Low to High), Use Ctrl-Click to Zoom, Alt-Click to Reset Zoom
Dashboard
Achilles Heel
Person
Observation Periods
Data Density
Conditions
Condition Eras
Measurement
Observations
Drug Eras
Drug Exposures
Procedures
Visits
Death


ATLAS
Data Sources
Home
Data Sources
Vocabulary
Concept Sets
Cohorts
Profiles
Jobs
Configuration
Feedback
Data Sources
Reports
Demo_data_1_percent_synthetic_patients
Conditions
Condition Prevalence
Treemap
Table
Vascular disorders
Arteriosclerosis, stenosis, vascular insufficiency and necrosis
Non-site specific necrosis and vascular insufficiency NEC
Arteriosclerosis
Coronary arteriosclerosis
Prevalence: 44.07%
Number of People: 51,278
Records per Person: 2.51
Box Size: Prevalence, Color: Records per Person (Blue to Orange = Low to High), Use Ctrl-Click to Zoom, Alt-Click to Reset Zoom
Coronary arteriosclerosis
Condition Prevalence
MALE
FEMALE
Age Decile
0-9
10-19
20-29
30-39
40-49
50-59
60-69
70-79
80-89
90-99


ATLAS
Data Sources
Home
Data Sources
Vocabulary
Concept Sets
Cohorts
Profiles
Jobs
Configuration
Feedback
Data Sources
Reports
Demo_data_1_percent_synthetic_patients
Data Density
Total Rows
# of Records
Year
Dashboard
Achilles Heel
Person
Observation Periods
Data Density
Conditions
Condition Eras
Measurement
Observations
Drug Eras
Drug Exposures
Procedures
Visits
Death
Records Per Person
Records Per Person
Condition occurrence
Death
Drug era
Drug exposure
Observation
Observation period
Procedure occurrence
Visit occurrence


ATLAS
Home
Data Sources
Vocabulary
Concept Sets
Cohorts
Profiles
Jobs
Configuration
Feedback
Data Sources
Data Sources
Reports
Demo_data_1_percent_synthetic_patients
Achilles Heel Report
Data Quality Messages
Dashboard
Achilles Heel
Person
Observation Periods
Data Density
Conditions
Condition Eras
Measurement
Observations
Drug Eras
Drug Exposures
Procedures
Visits
Death
Message Type
Message
ERROR
400-Number of persons with at least one condition occurrence, by condi
concepts in data are not in correct vocabulary
ERROR
600-Number of persons with at least one procedure occurrence, by proc
concepts in data are not in correct vocabulary
ERROR
900-Number of persons with at least one drug era, by drug_concept_id;
in vocabulary
ERROR
908-Number of drug eras without valid person; count (n=23,452,537) sh
ERROR
909-Number of drug eras outside valid observation period; count (n=23,475,293) should not be > 0
NOTIFICATION
Unmapped data over percentage threshold in:Condition
NOTIFICATION
Unmapped data over percentage threshold in:Procedure
NOTIFICATION
Unmapped data over percentage threshold in:DrugExposure
NOTIFICATION
Unmapped data over percentage threshold in:Observation
NOTIFICATION
Unmapped data over percentage threshold in:Measurement

Wrap up and Q&A

Technology Roadmap

- R Service integration
- Security (authentication, authorization, access control) - SHIRO
- Incidence analysis
- Data source visualizations direct from CDM results
- Cartoons
- UI/UX enhancements

Call to Action

Get Involved!

- Github.com
- OHDSI Forum
- OHDSI Wiki

Q & A