

# Demystifying Healthcare Information Systems

Chen Xie - MIT Lab for Computational  
Physiology

A dark blue diagonal gradient bar that starts from the bottom left corner and extends towards the top right corner, covering the bottom half of the slide.

# This Workshop

A technical and non-technical overview of the components of healthcare information systems (HIS).

Break down the common buzzwords

Leave knowing how the pieces fit together

Code: <https://github.com/cx1111/demystifying-healthcare-information-systems>

# Sections

1. Parts of a health information system
2. Storing data on computers: files and relational databases
3. Servers and cloud
4. Pooling and harmonizing data (Data Lake)
5. Analyzing and applying data (Big Data, AI)
6. Public data interfaces (API)

# Quickfire

- A server is a computer.
- A cloud computer is a computer rented from a large company stored in their remote server farm/data center.
- A data lake is a pooled collection of disparate collected data.
- Big data is data too big to fit on your everyday computer. Or it is an arbitrary term. What is big? Quality of data?

# Quickfire

- Artificial intelligence is intelligence being demonstrated by machines. Machines appearing smart. Usually built on top of machine learning/other mathematical methods.
- An application programming interface (API) is a set of clearly defined methods of communication among various software components.

# Parts of a Health Information System

## Data Sources

Medical Devices

Electronic Health Records

Wearables

Claims Data

## Data Storage

Server  
(Data Lake)

## Data Preparation

## Data Analysis

## Data Driven Application

Maybe done on cloud servers

```
graph LR; subgraph DS [Data Sources]; MD[Medical Devices]; EHR[Electronic Health Records]; W[Wearables]; CD[Claims Data]; end; DS --> S[(Server Data Lake)]; S --> P[Data Preparation]; P --> A[Data Analysis]; A --> DA[Data Driven Application];
```

Files

# Files

- A file is a fixed length series of zeros and ones (bits).
- Different formats encode information differently. Ie: ASCII, two's complement binary. Use the right software to read/write the right file. Better if the format is open like csv or json so we're not restricted by software choices.
- Flat files are easy to move between computers. All computers use and see the same zeros and ones in a file.



# Relational Databases

# Relational Databases

“A relational database (RDB) is a digital database based on the relational model of data”. I.e. data elements are grouped, like in a table row.

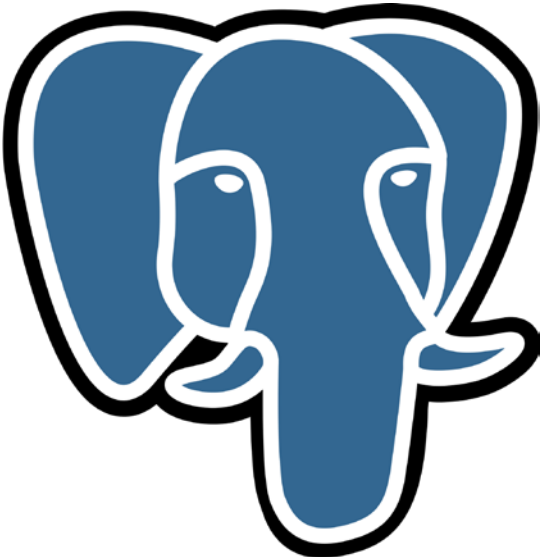
- Effectively: storing data in tables with columns representing attributes and rows representing instances of data.
- RDBs can be queried using a Structured Query Language (SQL).
- A common and intuitive way to store and use data.
- Actually RDBs store data in files too, but the software manages it all.

# Popular Relational Databases

ORACLE®



MySQL®



Microsoft®  
SQL Server®

# Notes about Relational Databases

- Most digital applications store their data using relational databases.
- There are non-relational databases for instances where data does not fit into a relational format.
- Watch out for expensive enterprise systems which have not been shown to outperform open systems.

# Flat Files vs Relational Databases

- Pros of RDBs
  - Can extract content using SQL.
  - Software ensures data integrity.
  - Quick search, insertion, and retrieval of content from large databases.
  - Easy to modify the structure/add data to relational databases using the software.
- Pros of Files
  - Takes up less space/overhead compared to RDBs.
  - For simple data, flat files are easier to manage.
  - Easier to share flat files. Just send them individually.

# Servers and Cloud

# Servers

“A server is a computer program or a device that provides functionality for other programs or devices, called "clients". - Wikipedia

Effectively the device server:

- The server device/computer ‘serves’ content to others.
- Big powerful computer in storage rack. Designed to be run 24/7 and cooled.
- Contains redundant arrays of independent disk (RAID).

# Servers

## Device server:

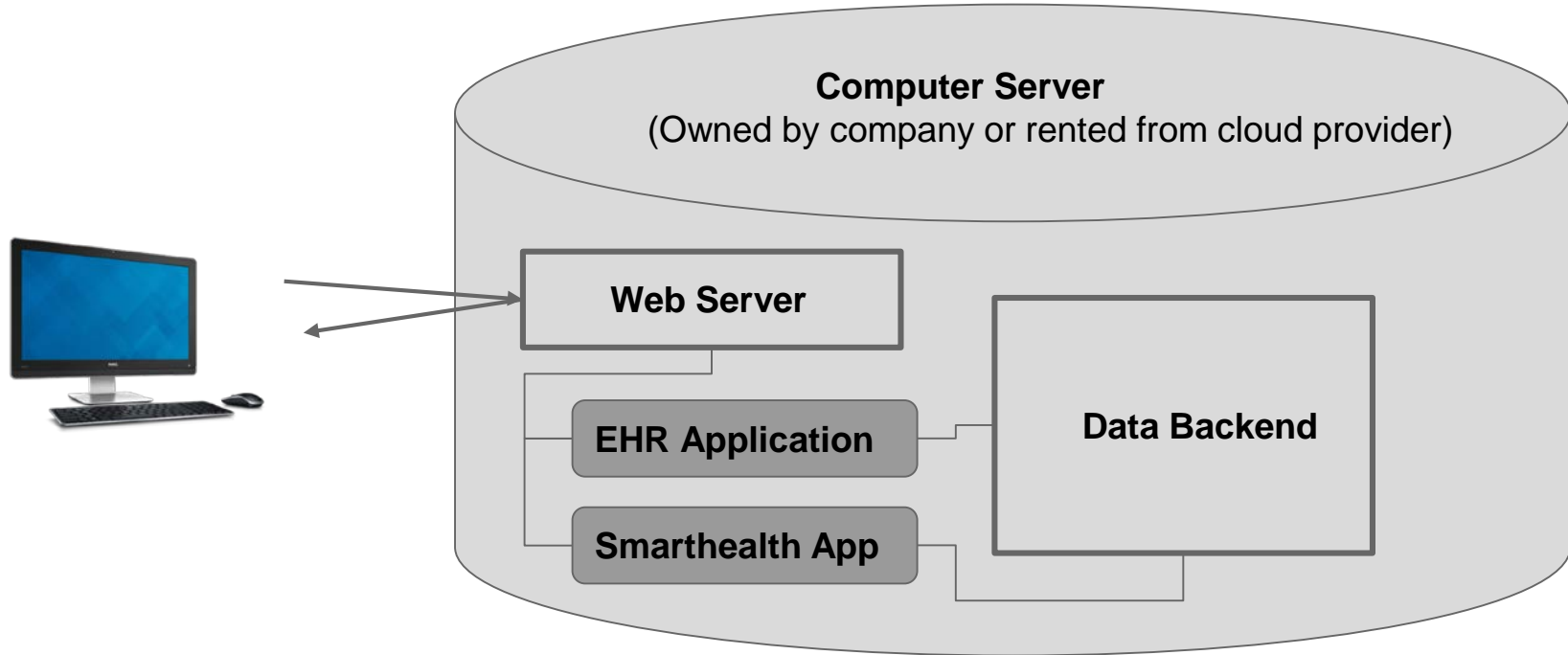
- The device/computer 'serves' content to others.
- Big powerful computer in storage rack. Designed to be run 24/7 and cooled.
- Contains redundant arrays of independent disk (RAID).

## Application server:

- A program that serves content to clients (users or other programs). I.e. web server, database server, smart health application server.



# Overall Picture



# Servers



# Cloud

“Cloud computing is shared pools of configurable computer system resources and higher-level services that can be rapidly provisioned with minimal management effort, often over the Internet.” - Wikipedia

- Effectively: Hiring computers from a company.
- Ie. Moving to the cloud, deploying from the cloud, etc...
- Computers are just computers...

# Popular Cloud Computing Providers



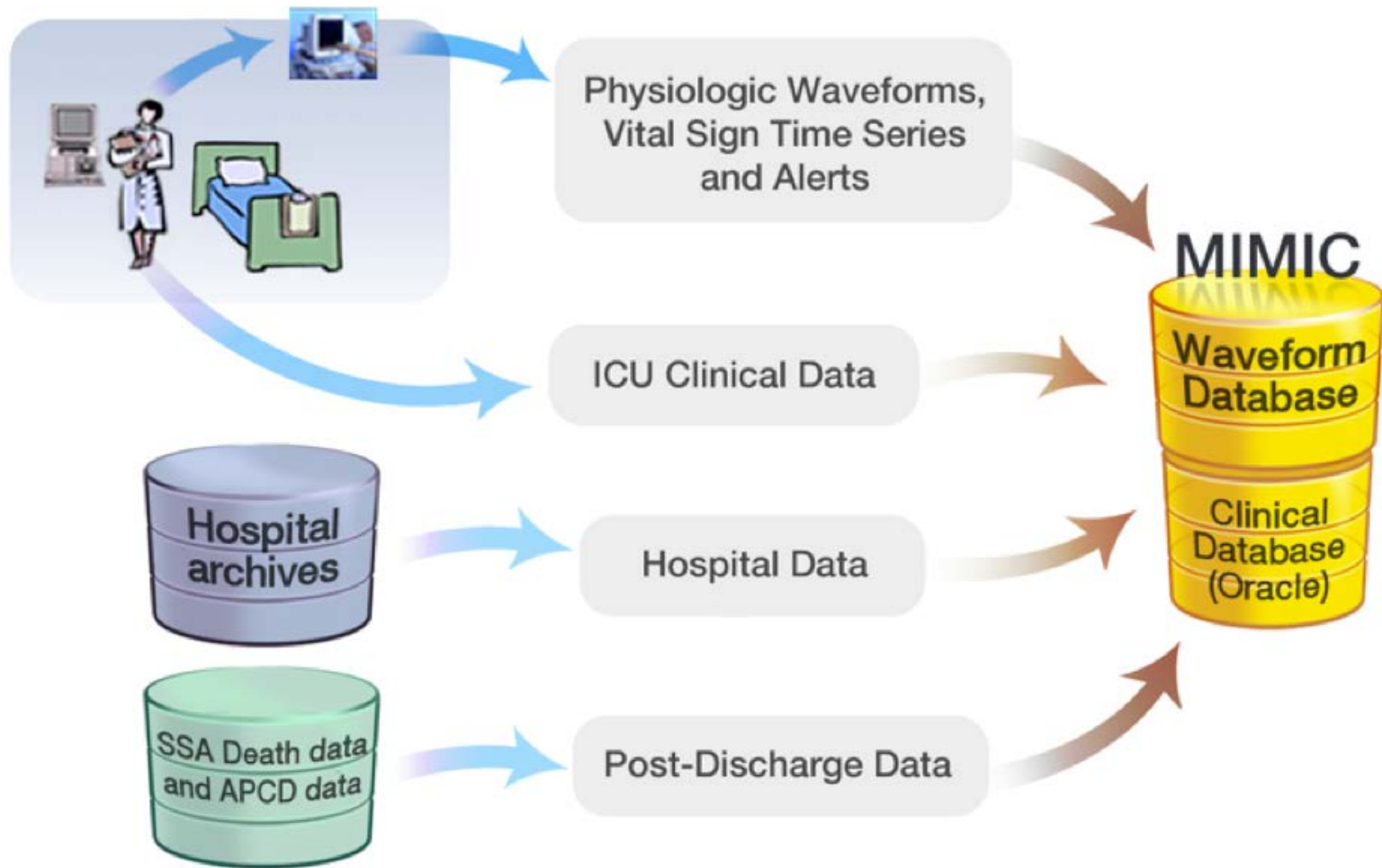
Google Cloud Platform

# Utilizing Third Party Cloud Providers

- Pros
  - Avoid needing to pay for large amounts of computing hardware infrastructure and maintenance.
  - Flexible scaling of resources. Pay for what you need.
- Cons
  - May cost more in the long term for same compute resources.
  - Organization regulations may prohibit third party compute providers.

<https://cloud.google.com/products/calculator/>

# Pooling and Harmonizing Data

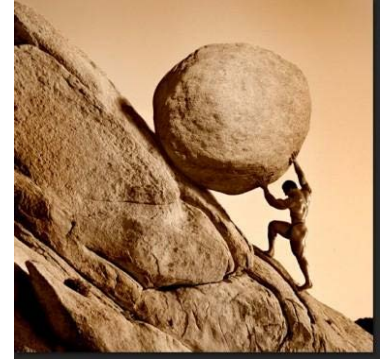


# Pooling and Harmonizing Data

- Data collected almost always comes in several disparate formats.
- A data lake is (one or more computers) where you pool all the different sources of data.
- In order for the data to be of use, it should be processed and converted into a usable layout/format.



Data Collection  
and Curation



Machine Learning and  
AI



Bringing the Algorithms to the  
Bedside or Policy makers



# Analyzing and Applying Data



# WHAT IS A.I.?

Artificial intelligence (AI) is defined by Marvin Minsky as the science of making machines do things that would require intelligence if done by man.

A **computer** performs a task by following the rules that it is fed.

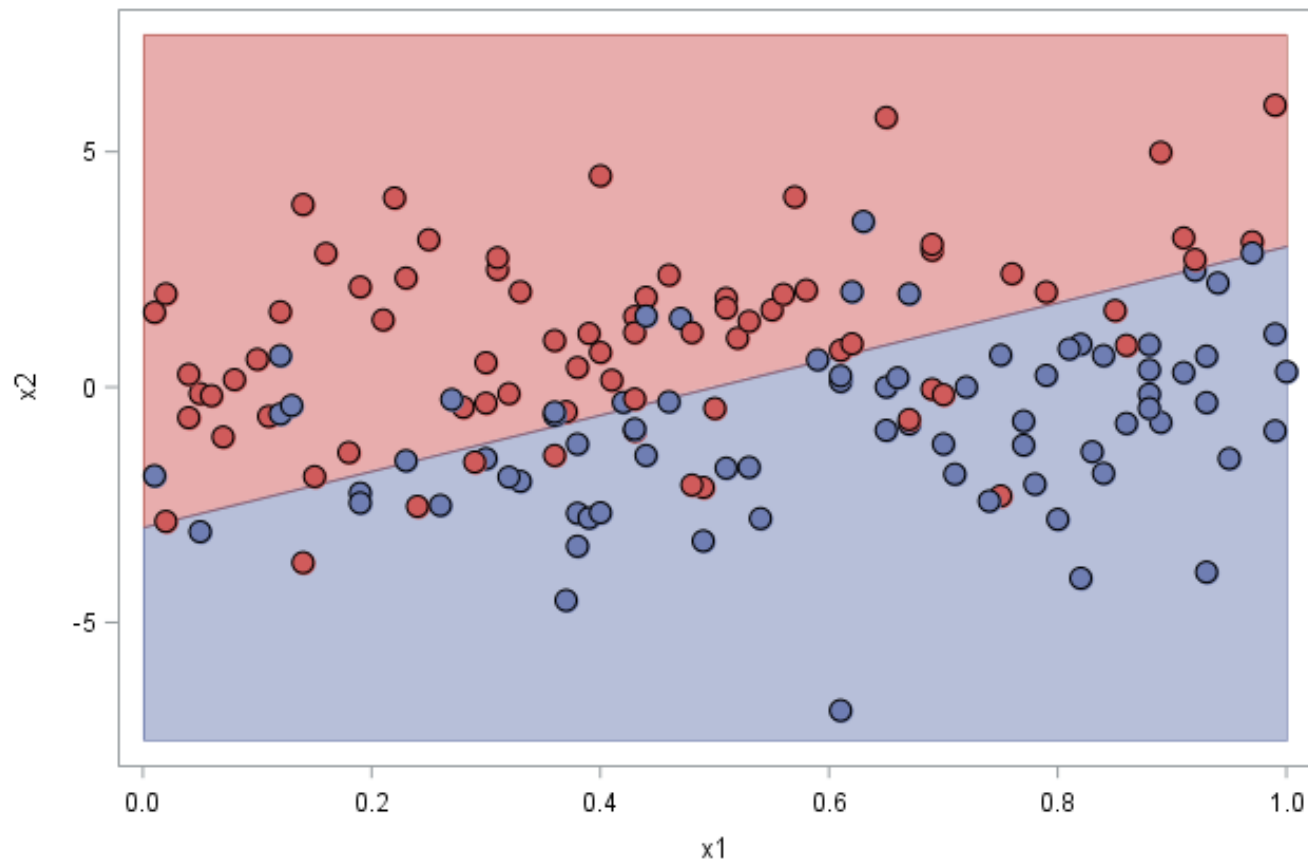


An **A.I.** performs a task by learning the rules from data that it is fed.

# Analyzing Data to Extract Insights

- Machine learning: using algorithms to extract insights from data.
- Artificial intelligence: using a machine to do apparently intelligent things: Intelligent applications don't have to learn from the data, but they should.

Data with Binary Response



Predicted Regions 0 1

# Actionable Analytics from Healthcare Data

Use analytics tools to support clinicians. Balance between complexity and interpretability.

Ways to improve model interpretability:

- Show input variables used.
- Show level of uncertainty.
- Show sample size of training data used.
- Show how different inputs affect the outcome, or level of uncertainty (sensitivity analysis).
- Intuitive model concepts like patient matching and regression.

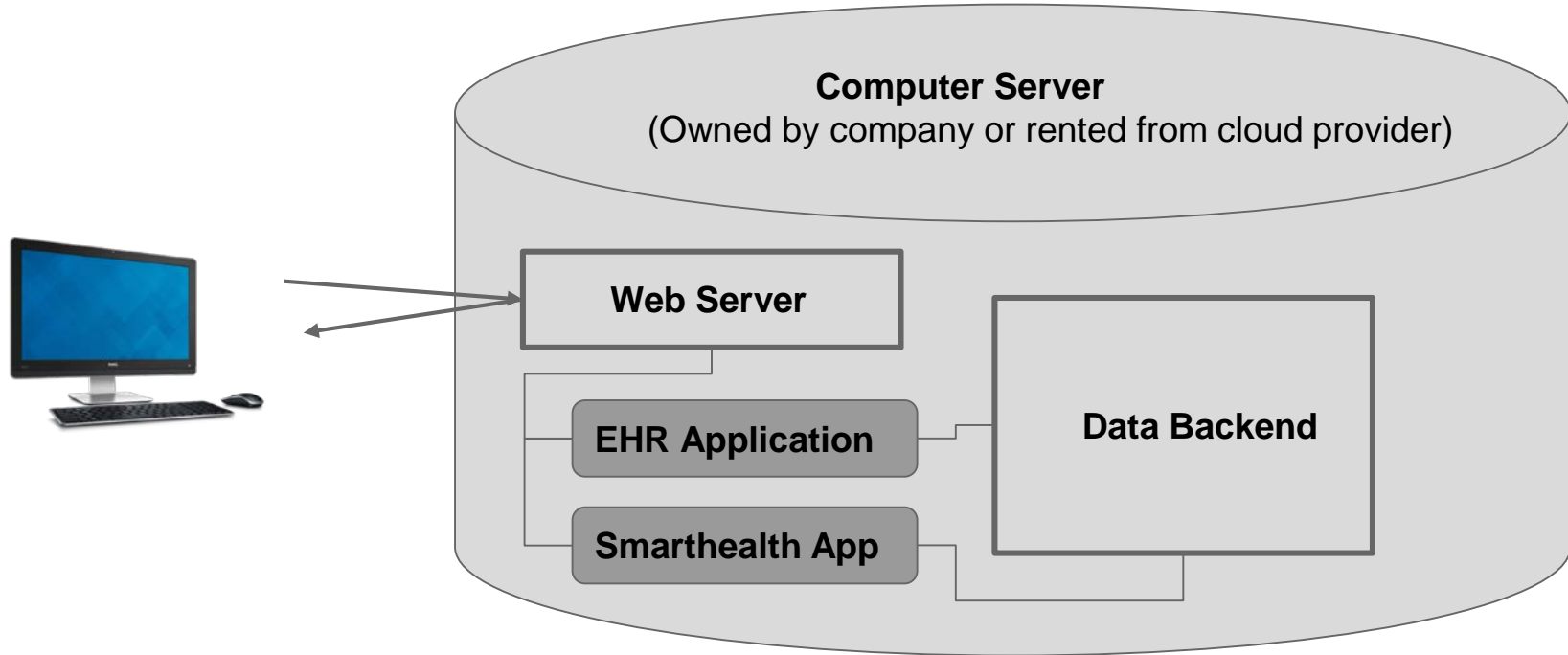
# Data Driven Applications



# Data Driven Applications

- We have the analytics done. Now we need to deliver it to users.
- Not just for publishing papers!
- Application server to receive requests and deliver content to client (user).
- Application server calls the analytics app to run its analysis and gives the results back to the client.
- Many shared resources on the same computer/computer system.

# Overall Picture



# Application Programming Index

“An application programming interface (API) is a set of subroutine definitions, communication protocols, and tools for building software”. - Wikipedia

Meaning: Interfaces/standards for different computer programs to communicate.

Examples:

- Health application developer makes a public API for users to query their own data.
- Whatsapp API allows automated sending of messages.
- Cloud platform API allows spinning up and tearing down of VMs.

# Parts of a Health Information System

## Data Sources

Medical Devices

Electronic Health Records

Wearables

Claims Data

## Data Storage

Server  
(Data Lake)

## Data Preparation

## Data Analysis

## Data Driven Application

Maybe done on cloud servers

```
graph LR; subgraph DS [Data Sources]; MD[Medical Devices]; EHR[Electronic Health Records]; W[Wearables]; CD[Claims Data]; end; DS --> S[(Server Data Lake)]; S --> P[Data Preparation]; P --> A[Data Analysis]; A --> DA[Data Driven Application]; C[Maybe done on cloud servers];
```