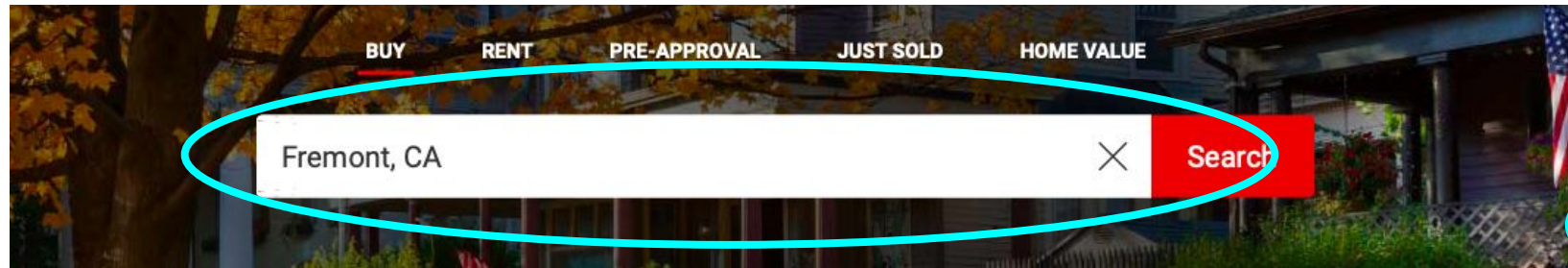




How We Are Bringing Science to the Art of Home Buying For Materiall

Materiall + UC Berkeley

Problem



realtor.com® Buy Sell Rent Mortgage Find Realtors® My Home News & Insights

Address, City, Zip, Neighborhood Price Property Type Beds Baths Listing Status More Filters Save Search

New listings in Fremont, CA

[View All 41 New Listings](#)



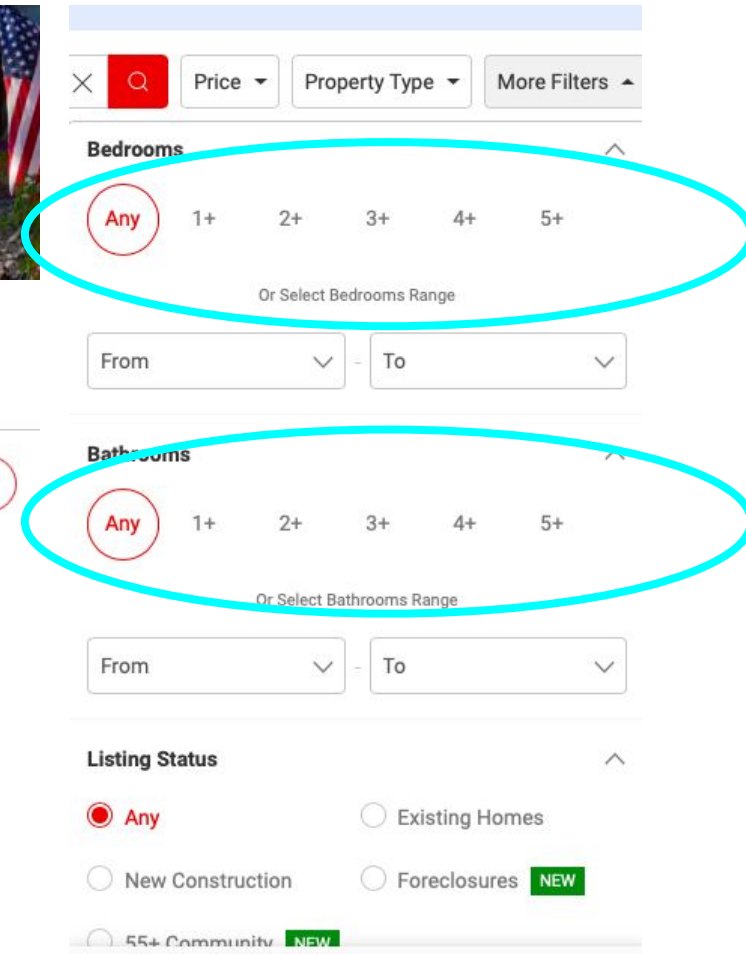
4 bed 3 bath 2,664 sqft
715 Washo Dr, Fremont, CA 94539



3 bed 2 bath 1,166 sqft
4111 Greenland Ter, Fremont, CA 94555



2 bed 2 bath 934 sqft
5027 Shalimar Cir, Fremont, CA 94555



Solution:

Computer Vision for Home Buying

1. Identify if room is Indoor or Outdoor

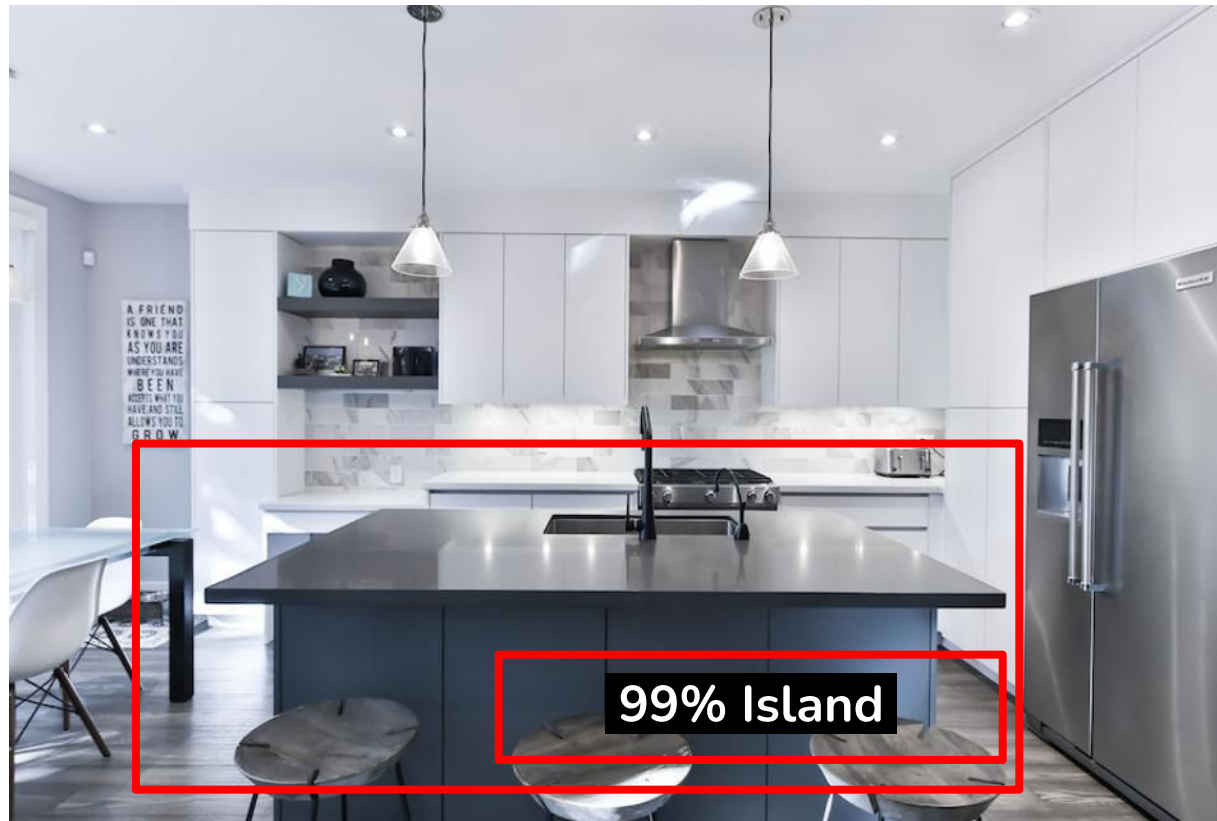
2. Identify rooms within a home



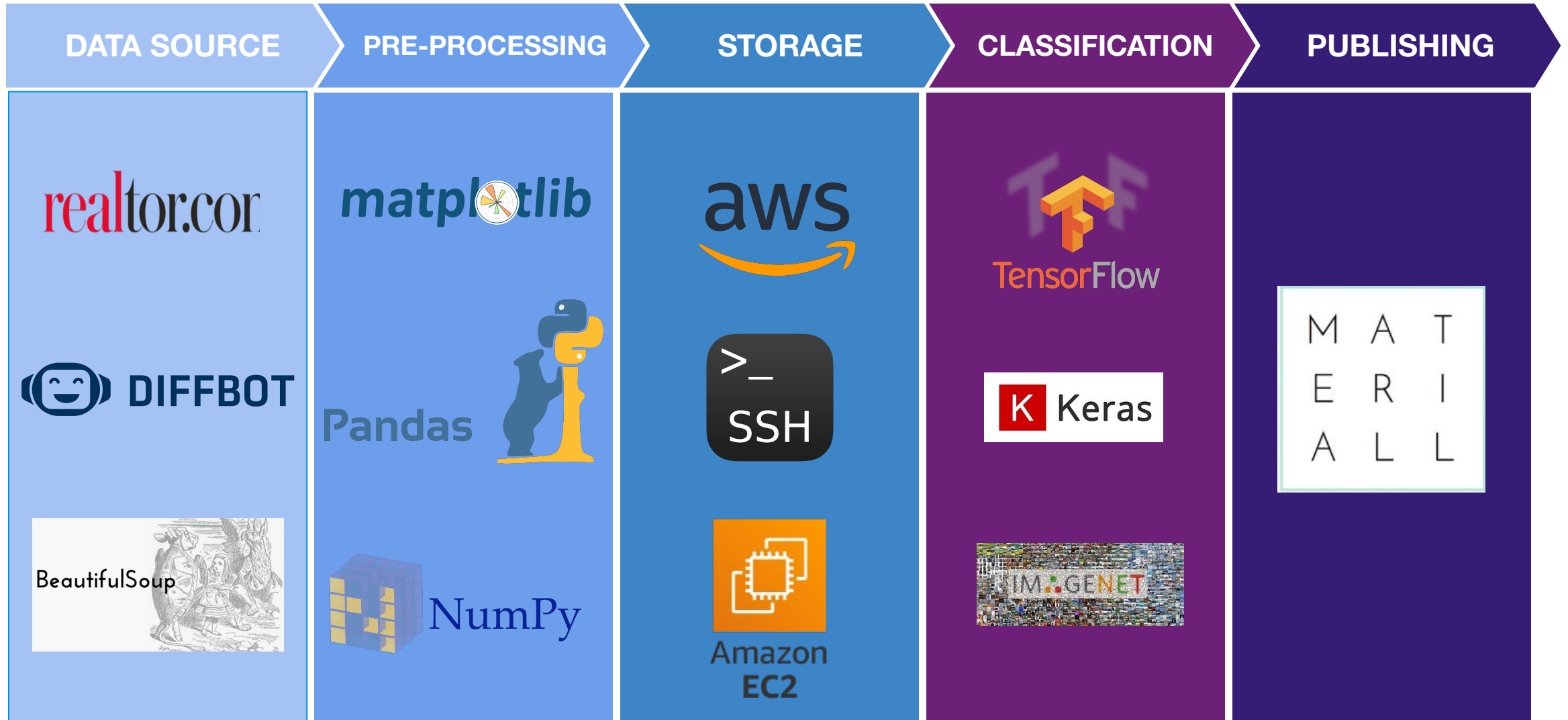
Solution:

Computer Vision for Home Buying

3. Identify features within a room



System Architecture



System Architecture

DATA SOURCE

PRE-PROCESSING

STORAGE

CLASSIFICATION

PUBLISHING

realtor.com



[CRAWLBOT](#) > [ALL CRAWLS](#) > CRAWL DETAIL

Realtor

[Crawl Status](#)

Crawl Settings

Crawl Name *

Realtor

No spaces or special characters.

Seed URLs *

<https://www.realtor.com/realestateandhome>

Separate multiple URLs with a space.

DIFFBOT API SETTINGS

Choose the [Automatic](#) or [Custom API](#) through which pages should be processed. All pages will be processed unless a Page Processing Limit is defined below.

Process pages via

REALTOR2

The [Analyze API](#) automatically identifies and extracts articles, images and product pages.

Diffbot Querystring

& links

Append this querystring to all Diffbot API calls, e.g. `&fields=tags`. [More info](#).

CRAWLING LIMITS

Define the pages or sections of a site you wish to spider.

Crawl Patterns

realestateandhomes-search
realestateandhomes-detail

Only crawl URLs that contain any of these text strings (one per line). [More info](#).

System Architecture

DATA SOURCE

PRE-PROCESSING

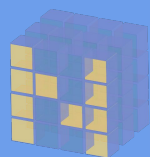
STORAGE

CLASSIFICATION

PUBLISHING

matplotlib

Pandas



NumPy



System Architecture

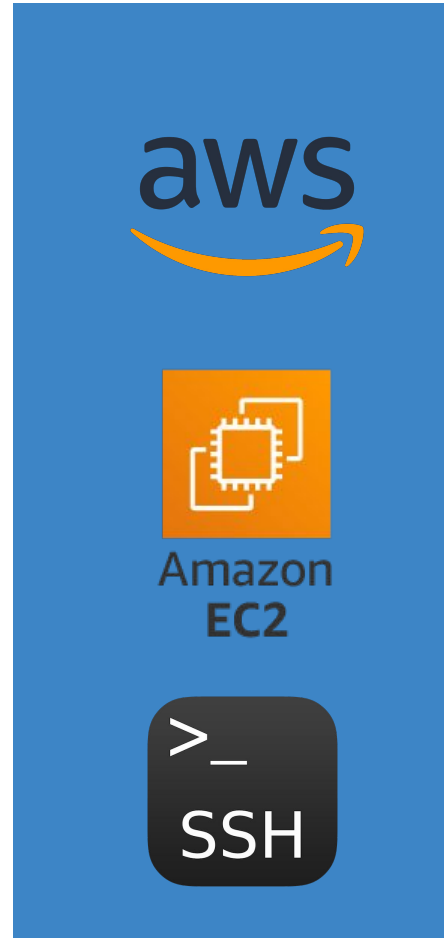
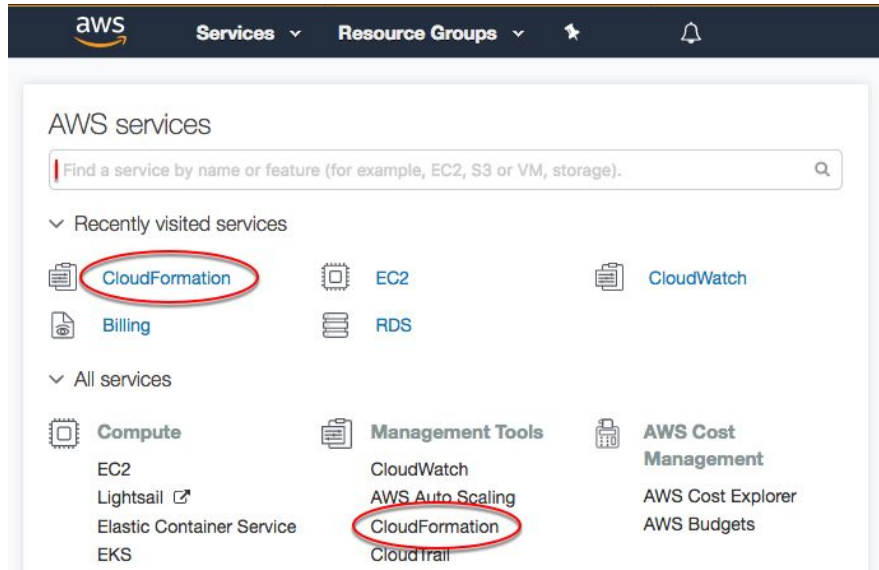
DATA SOURCE

PRE-PROCESSING

STORAGE

CLASSIFICATION

PUBLISHING



System Architecture

DATA SOURCE

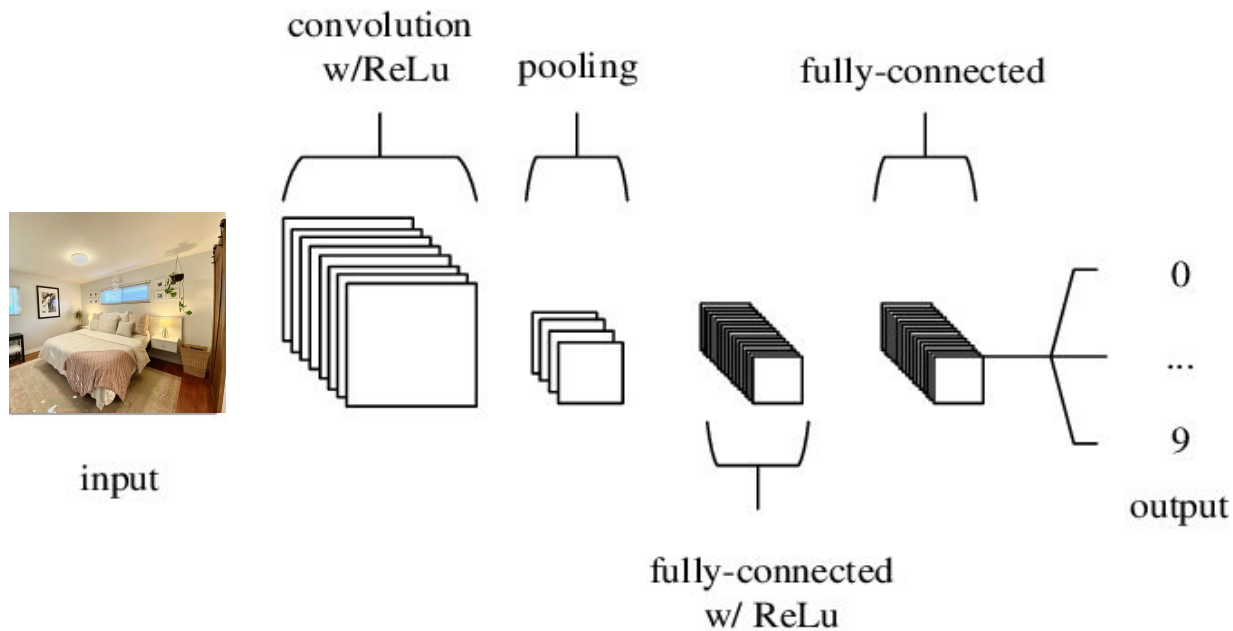
PRE-PROCESSING

STORAGE

CLASSIFICATION

PUBLISHING

- Keras VGG16 - Pre-trained CNN Model



System Architecture

DATA SOURCE

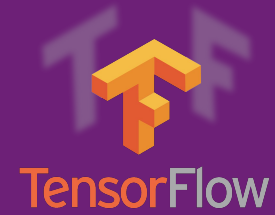
PRE-PROCESSING

STORAGE

CLASSIFICATION

PUBLISHING

- Keras VGG16
- Two main classification models
 - a. Indoor/outdoor
 - b. Indoor room (e.g. kitchen, bathroom)
- Main idea: automatically populating the descriptions based off the images of the home



System Architecture

DATA SOURCE

PRE-PROCESSING

STORAGE

CLASSIFICATION

PUBLISHING

M A T
E R I
A L L

[Solution](#) [Team](#) [Contact us](#)

OMNICHANNEL
**PRODUCT DISCOVERY
& ENGAGEMENT**

that ACTUALLY works!

M A T
E R I
A L L



The Team

Project Manager:
Parker Nelson

Communications Officer:
Vincent Lao

Engineers:
Bharadwaj Swaminathan
Daniel del Carpio
Samantha Tang

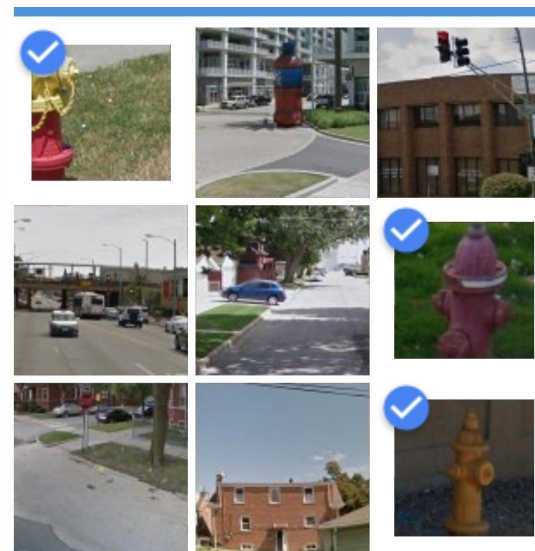
Learning Path - Web Scraping

WHO WOULD WIN?

Sophisticated webscraping tools



Some blurry fire hydrants



Learning Path - Web Scraping

realtor.com®

Buy Sell Rent Mortgage Find Realtors® My Home News & Insights Log In Sign up Advertise

< Back


Fremont, CA

California > Alameda County > Fremont > 4140 Hawkins St

Presented by:

David Ascencio with Allie

div.slick-slide.slick-active.slick-current 561 x 424.967



New For Sale - Active 1 / 2

Map

More about this property

Full Name

Email

Phone

I am interested in 4140 Hawkins St.

☐ Get pre-approved by a lender.

Email Agent

By proceeding, you consent to receive calls and texts at the number you provided, including marketing by autodialer and prerecorded and artificial voice, and email, from realtor.com and others about your inquiry and other home-related matters, but not as a condition of any purchase; this applies regardless of whether you check, or leave un-checked, any box above. More...

\$975,000

Est. Payment \$3,279/mo

4 bed 1 bath 1,600 sqft 5,300 sqft lot

4140 Hawkins St,
Fremont, CA, 94538

Inspector

Search HTML

```
event
<div class="slick-list"> event
  <div class="slick-track" style="width: 5049.11px; opacity: 1; transform: translate3d(-561.112px, 0px, 0px);"> event
    ::before
    <div class="slick-slide slick-cloned" data-index="-1" tabindex="-1" aria-hidden="true" style="width: 561px;"></div> event
    <div class="slick-slide slick-active slick-current" data-index="0" tabindex="-1" aria-hidden="false" style="outline: currentcolor none medium; width: 561px; --darkreader-inline-outline: currentcolor;" data-darkreader-inline-outline=""> event
      <div>
        <div class="main-carousel" data-testid="gallery-item" tabindex="-1" style="width: 100%; display: inline-block;"> event
          <picture>
            <source type="image/webp"> event
             event
          </picture>
        </div>
      </div>
    <div class="slick-slide slick-active" data-index="1" tabindex="-1" aria-hidden="false" style="outline: currentcolor none medium; width: 561px; --darkreader-inline-outline: currentcolor;" data-darkreader-inline-outline=""></div> event
    <div class="slick-slide" data-index="2" tabindex="-1" aria-hidden="true" style="outline: currentcolor none medium; width: 561px; --darkreader-inline-outline: currentcolor;" data-darkreader-inline-outline=""></div> event
    <div class="slick-slide" data-index="3" tabindex="-1" aria-hidden="true" style="outline: currentcolor none medium; width: 561px; --darkreader-inline-outline: currentcolor;" data-darkreader-inline-outline=""></div> event
    <div class="slick-slide slick-cloned" data-index="4" tabindex="-1" aria-hidden="true" style="width: 561px;"></div> event
    <div class="slick-slide slick-cloned" data-index="5" tabindex="-1" aria-hidden="true" style="width: 561px;"></div> event
    <div class="slick-slide slick-cloned" data-index="6" tabindex="-1" aria-hidden="true" style="width: 561px;"></div> event
    <div class="slick-slide slick-cloned" data-index="7" tabindex="-1" aria-hidden="true" style="width: 561px;"></div> event
    ::after
  </div>
</div>
<div class="jsx-3961344686 slick-arrow slick-next" role="button" aria-label="right photo"></div>
event
</div>
<div class="leadform-wrapper"></div> flex
<div class="ad-wrapper"></div>
```



Learning Path - Data Acquisition & Preprocessing



DIFFBOT



	image_url	label	probability	Address	Price
0	https://ap.rdcpix.com/d7b0b0ef57213b43da91b2f1...	house_view	0.987999	38066-Stenhammer-Dr_Fremont_CA_94536_M26705-99955	875000
1	https://ap.rdcpix.com/d7b0b0ef57213b43da91b2f1...	yard	0.527620	38066-Stenhammer-Dr_Fremont_CA_94536_M26705-99955	875000
2	https://ap.rdcpix.com/d7b0b0ef57213b43da91b2f1...	porch	0.822153	38066-Stenhammer-Dr_Fremont_CA_94536_M26705-99955	875000
3	https://ap.rdcpix.com/d7b0b0ef57213b43da91b2f1...	living_room	0.992320	38066-Stenhammer-Dr_Fremont_CA_94536_M26705-99955	875000
4	https://ap.rdcpix.com/d7b0b0ef57213b43da91b2f1...	living_room	0.999825	38066-Stenhammer-Dr_Fremont_CA_94536_M26705-99955	875000



Learning Path - Model building

Team: *trying to do our best to improve test accuracy*

Model:





Learning Path - Model building

```
run_test_harness('all', epochs=25, test_size = 0.1, shape=(100, 100, 3), bat
```

```
Classes: ['unknown' 'kitchen' 'bathroom' 'bedroom' 'house_view' 'living_room'  
         'road_view' 'yard' 'porch' 'dining_room' 'porch_yard' 'garage']
```

```
Number of observations: 1336
```

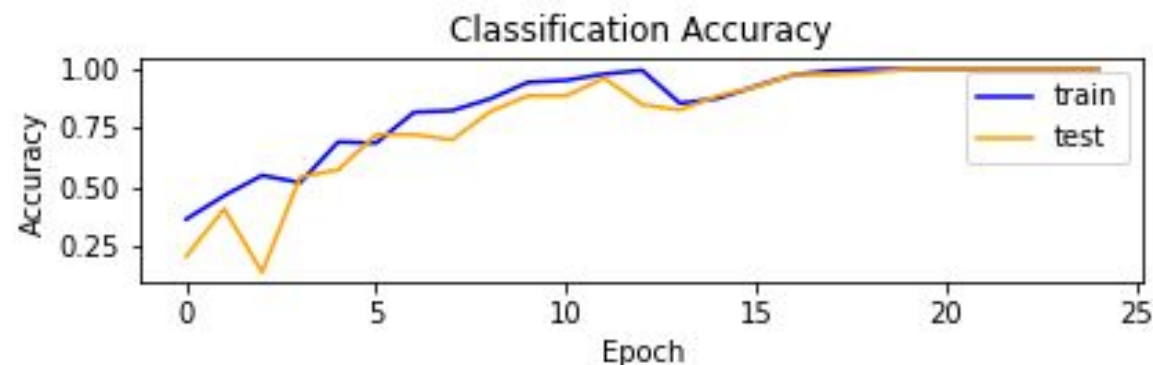
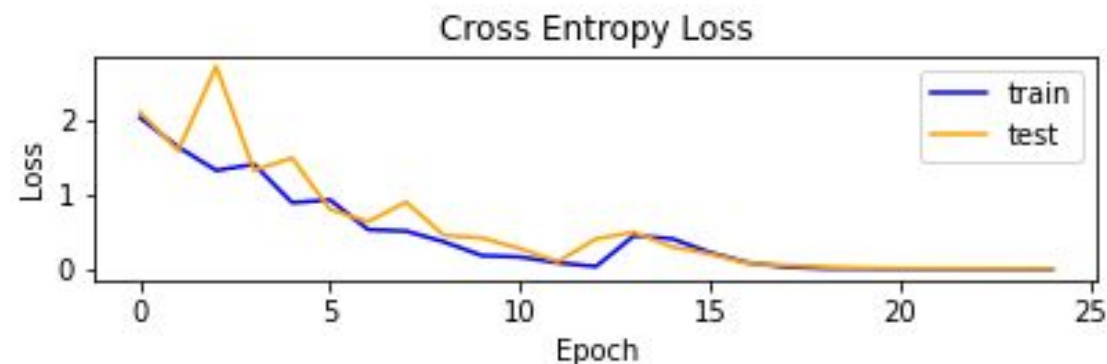
```
Dataset Loaded!
```

```
Dataset Preprocessing Turned Off!
```

```
Defining Model...
```

```
Model: "functional_3"
```

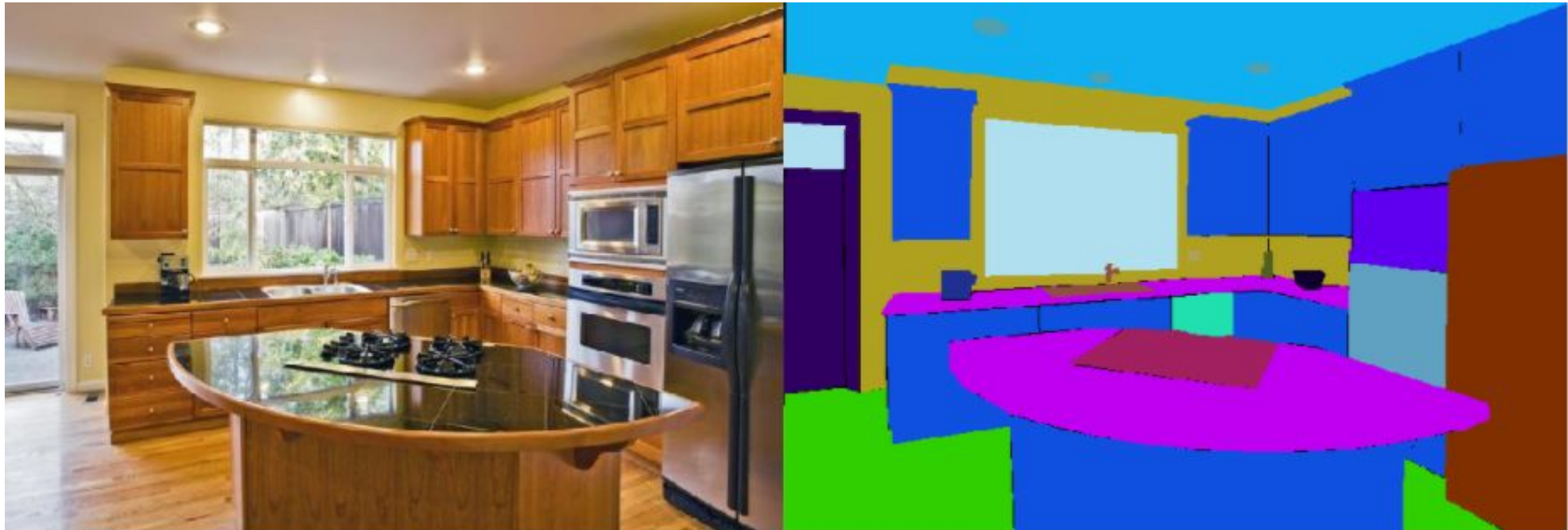
Layer (type)	Output Shape	Param #
=====		
input_2 (InputLayer)	[(None, 100, 100, 3)]	0
block1_conv1 (Conv2D)	(None, 100, 100, 64)	1792
block1_conv2 (Conv2D)	(None, 100, 100, 64)	36928
block1_pool (MaxPooling2D)	(None, 50, 50, 64)	0





Next Steps

- Use ADE20K dataset to segment features in the images
- Object detection to classify attributes in the room

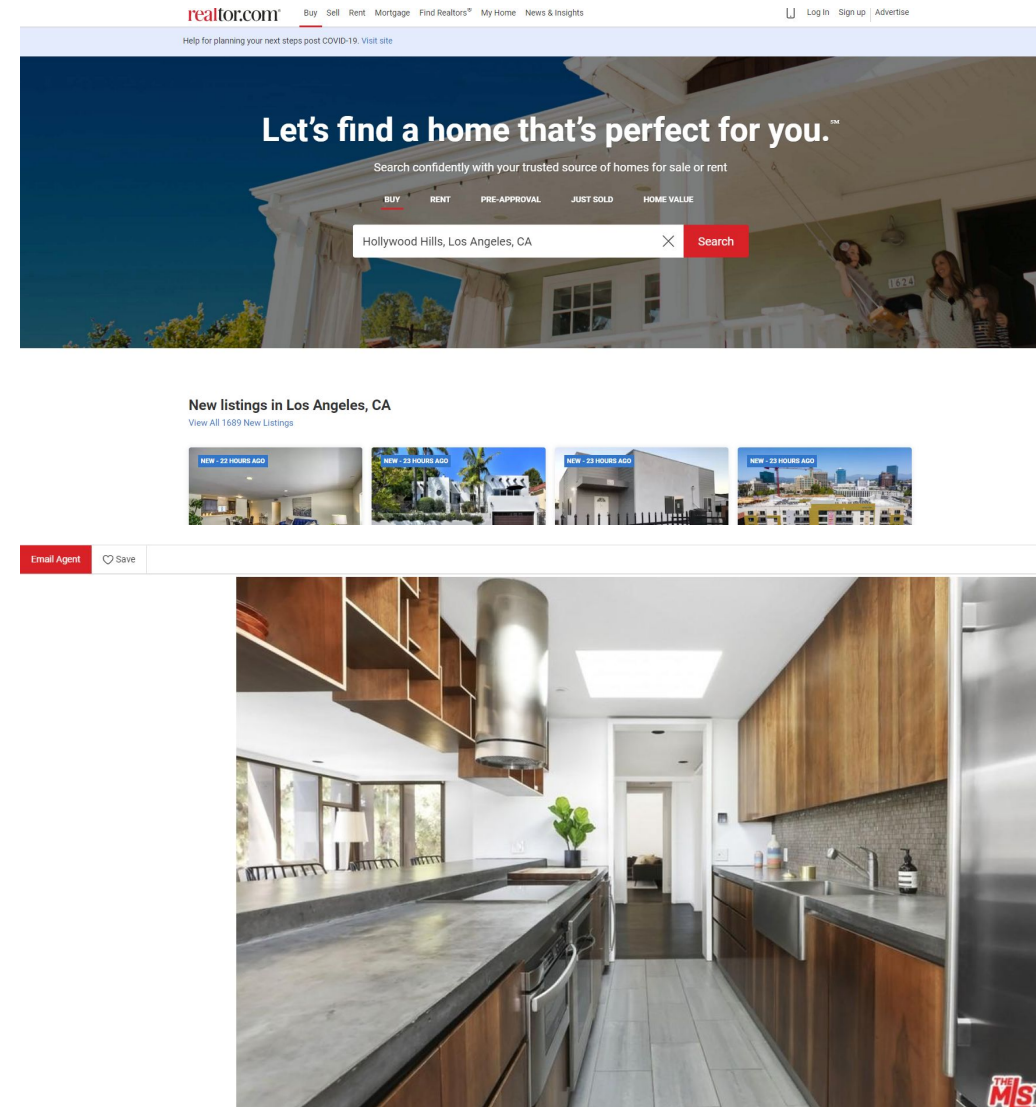


The background is a solid blue color. It is decorated with a pattern of white line-art icons. These icons include various sizes of cubes, some of which are stacked. Other icons include a hand with an index finger pointing, a magnifying glass with a plus sign, and a funnel. The icons are scattered across the entire background, with a higher density on the left side.

Thank you!

Giving Personalized Recommendations to Home-Buyers

- **Mission: Bringing science to the art of home buying.**
- Use data science to combine objective and subjective features to provide tailored recommendations for home buyers



Prof Sidhu recommendation:

9 Top Dream House Features Every Homeowner Needs

Open Floor Plan. ...

Tall Ceilings. ...

Outdoor Living Area. ...

Large Kitchen. ...

Plenty of Natural Light. ...

Lots of Storage/Closet Space. ...

Low-Maintenance Landscaping. ...

Garage. If you live in a cold-weather climate, a garage will prevent you from having to scrape ice your windshield every morning.

manually label images say like kitchen size - small-medium-large

could do feature by feature, realator.com pics, text might be rough compared to picture (might be vague descriptions and be from seller trying to sell the house)

or take pictures of house and map it to size of house / sq ft (?)

house may seem more upgraded because more \$\$ because of house/sqft (ooo)

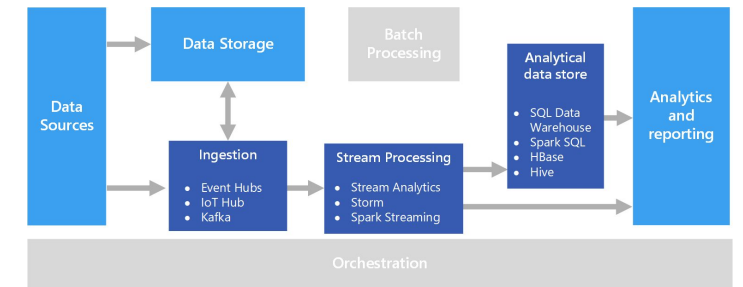
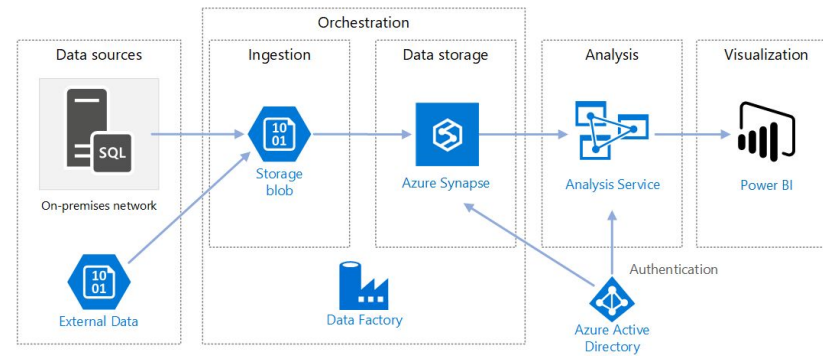
algorithm can pick up something we didn't see :thinking:

when we lay down the arch. brainstorm other things while the web crawler is crawling

make sure env running, make sure there is a smooth transition from crawler -> model

Sample System Architecture

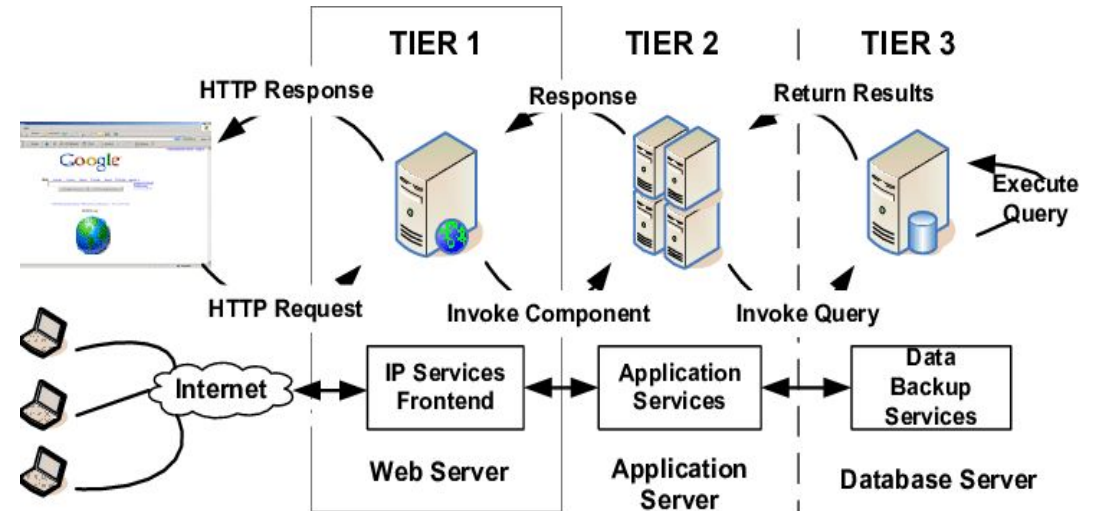
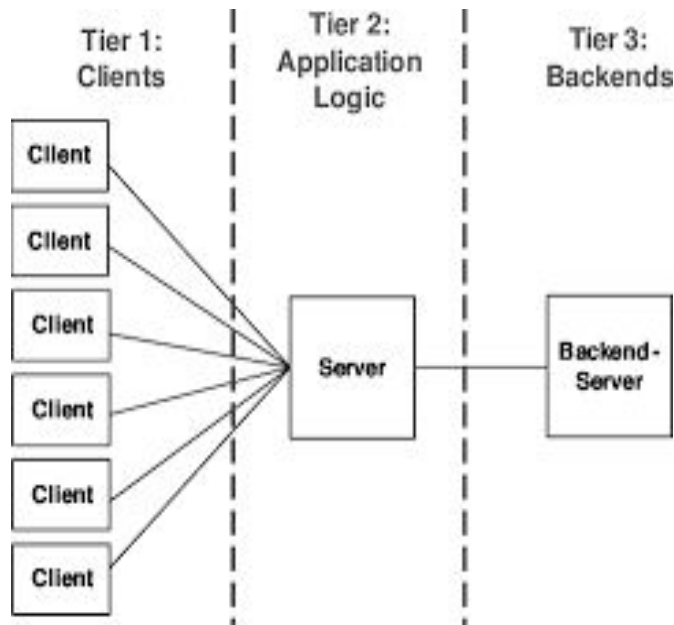
Replace the diagram below with your rough
System Architecture and/or Data Models
(it is okay to use two or more slides in this section)



Example: Data Pipelines from Microsoft

Sample System Architecture

Replace the diagram below with your rough
System Architecture and/or Data Models
(it is okay to use two or more slides in this section)

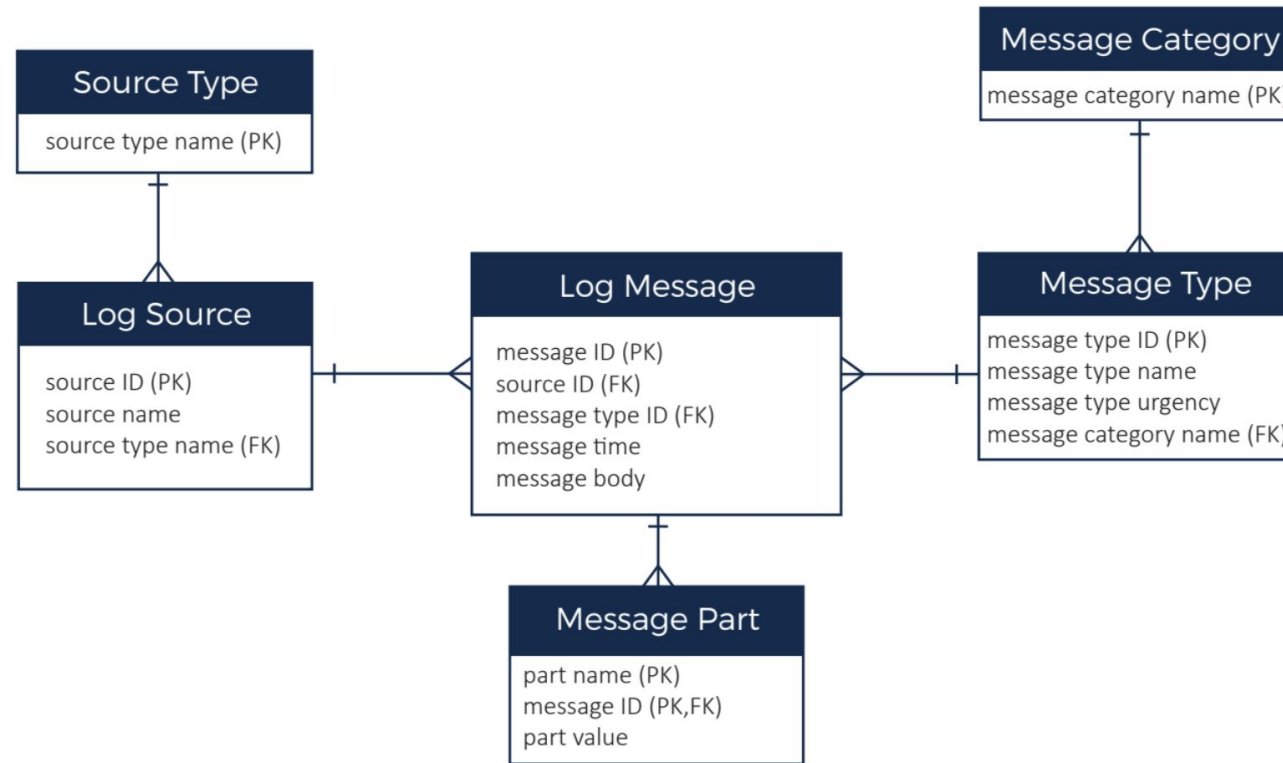


Trevor N. Mudge

Example: Client Server Architectures

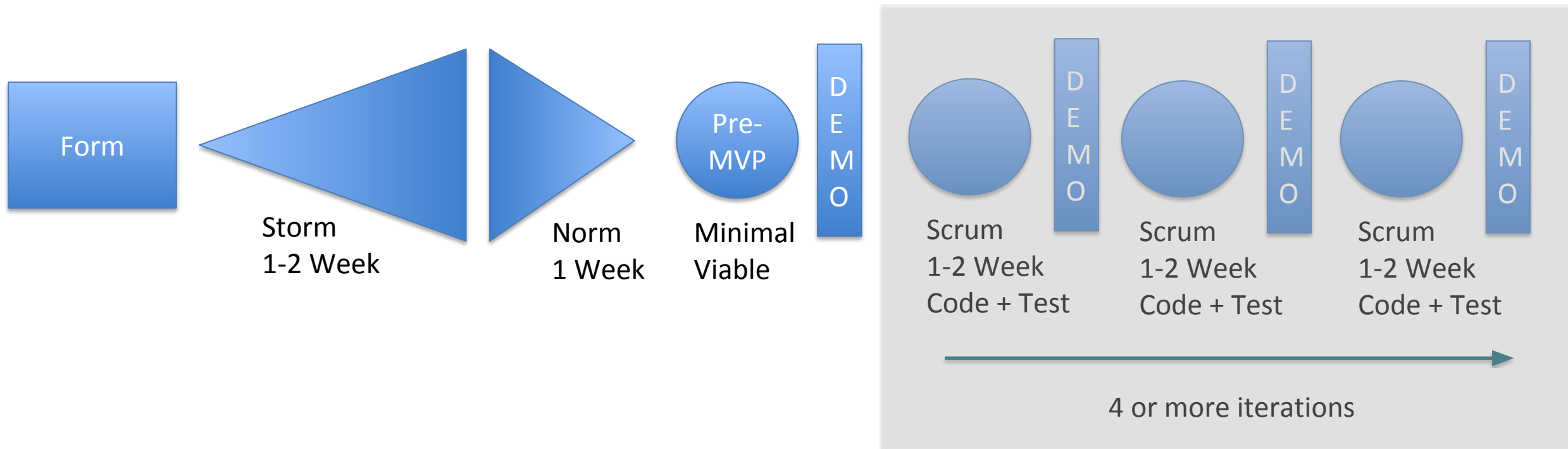
Sample System Architecture

Replace the diagram below with your rough
System Architecture and/or Data Models
(it is okay to use two or more slides in this section)



Example: Data Model for Social Media Data from Cassandra

What will you do next?



- What is the initial set of tasks (3-5) tasks
- Put initials or a name next to each
- Hint: start with the red items on your list of technical components
- Brainstorm/Research → Normalize Concepts → Simplest Minimal Demonstrable Version
- Consider using “Swim Lanes” to organize the team