

MICHIGAN STATE

UNIVERSITY

Beta Presentation Amazon Data Hub

The Capstone Experience

Team Amazon

Josh Barnett

Austin Cozzo

Dan Farat

Cameron Nejman

Robert Ramirez

Department of Computer Science and Engineering

Michigan State University

Spring 20



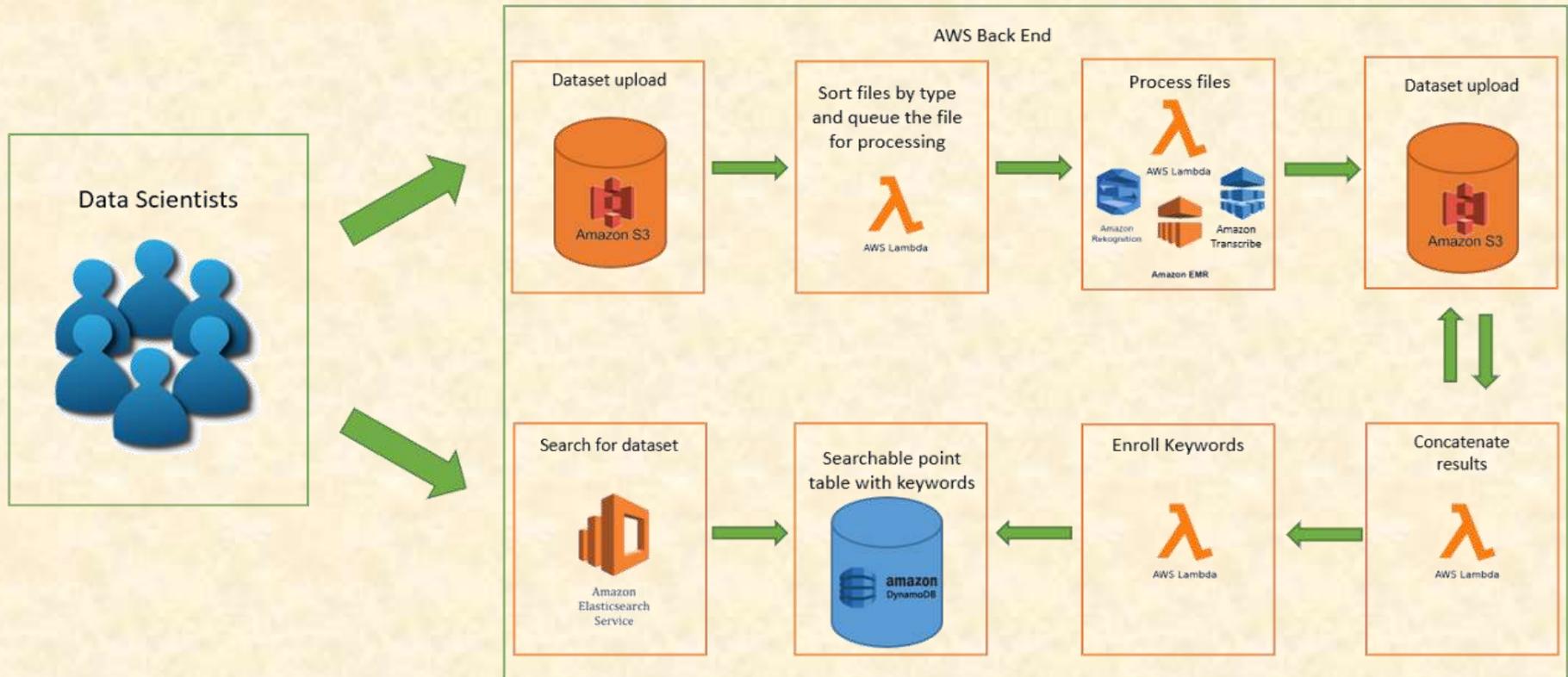
*From Students...
...to Professionals*

Project Overview

- Currently, Data Scientists waste a lot of time doing research on finding the “right” dataset
 - Dataset descriptions are often vague. The data in datasets can also be too old, too narrow, or too large
- Amazon Data Hub (ADH) assists in the process of finding useful datasets
 - ADH serves as a catalog of datasets. It also extracts metadata and generates keywords from datasets



System Architecture



Results page

Data Hub Login | Home

Results

Query

Count per page
10 ▾

Results: 3

Dataset Title: Bird test
Version: N/A
Uploader: ADHTest
Tags: class1: Animal, class0: Bird, file_type: .jpg, class3: Bluebird, class2: Robin, class4: Jay



Dataset Title: png birds
Version: 1.0
Uploader: ADHTest
Tags: class2: Person, class0: Bird, class3: Poster, class1: Bee Eater, class4: Dove, file_type: .png



Dataset Title: coco train
Version: 3.0
Uploader: ADHTest



© Team Amazon

Upload Edit page

Data Hub Login | Home

Upload More Info

Title

Version

Edit Dataset Tags:

Tag	Type	Remove?
Furniture	class1	<input type="button" value="X"/>
.jpg	file_type	<input type="button" value="X"/>
Chair	class0	<input type="button" value="X"/>

Edit Dataset Links:

Link **Remove?**

© Team Amazon



Dataset page (image dataset)

Data Hub Login | Home

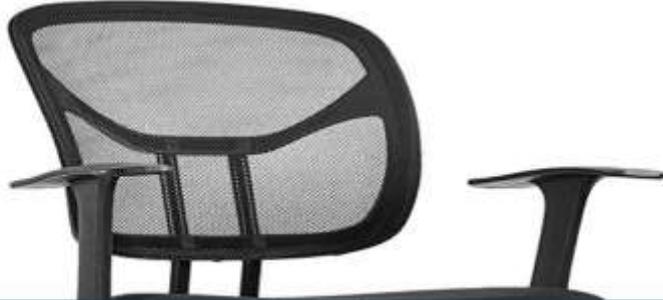
#148d371c76 - Annotated Chairs

Owner: ADHTest
Created: 3/4/2020
Version: 1.0
Similar Datasets:
• None



[Edit Dataset](#) [Delete Dataset](#)

148d371c76/91ula958mkl_ac_sx522_.jpg



© Team Amazon



Dataset page (video dataset)

Data Hub Login | Home

#6fa54c76c2 - animal video dataset

Owner: ADHTest
Created: 5/4/2020
Version: 1.0

Similar Datasets:

- None

[Edit Dataset](#) [Delete Dataset](#)

6fa54c76c2/animal_vid.mp4



TAGS

© Team Amazon



What's left to do?

- Web app is feature complete, but client has several valuable suggestions towards improvements
- Expand bootstrap usage
- Add stretch goals
 - favorite button
 - implicit dataset links
- Add JSON file support
- Download support



Questions?

?

?

?

?

?

?

?

?

?

