

**MICHIGAN STATE**  

---

**U N I V E R S I T Y**

# Alpha Presentation

## Amazon Data Hub

The Capstone Experience

Team Amazon

Joshua Barnett

Austin Cozzo

Dan Farat

Cameron Nejman

Robert Ramirez

Department of Computer Science and Engineering  
Michigan State University

Spring 2020



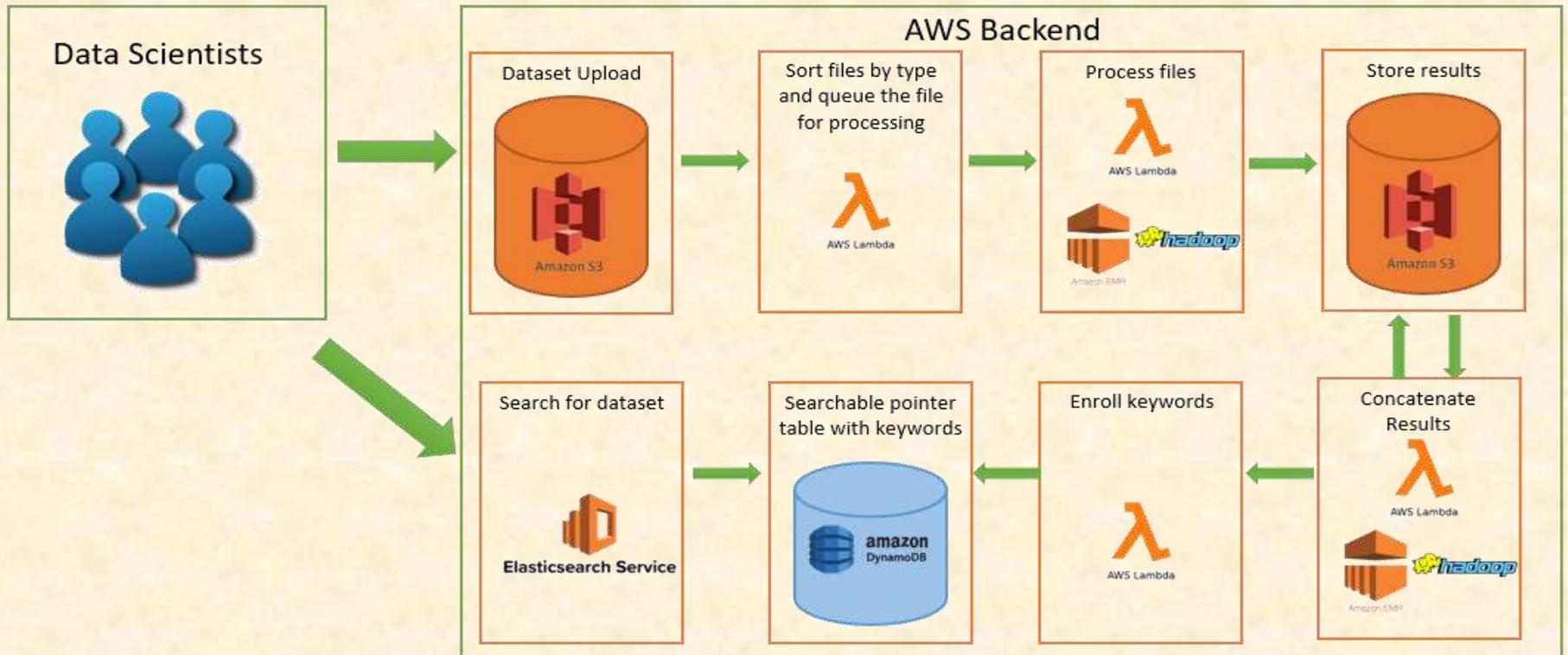
*From Students...  
...to Professionals*

# Project Overview

- Currently, Data Scientists waste a lot of time doing research on finding the “right” dataset
  - Datasets are often vague, old, too narrow, or too large
- Amazon Data Hub (ADH) will be used to assist in the process of finding useful datasets
  - Will be achieved through the catalog of datasets, the extraction of metadata, and the generation of keywords



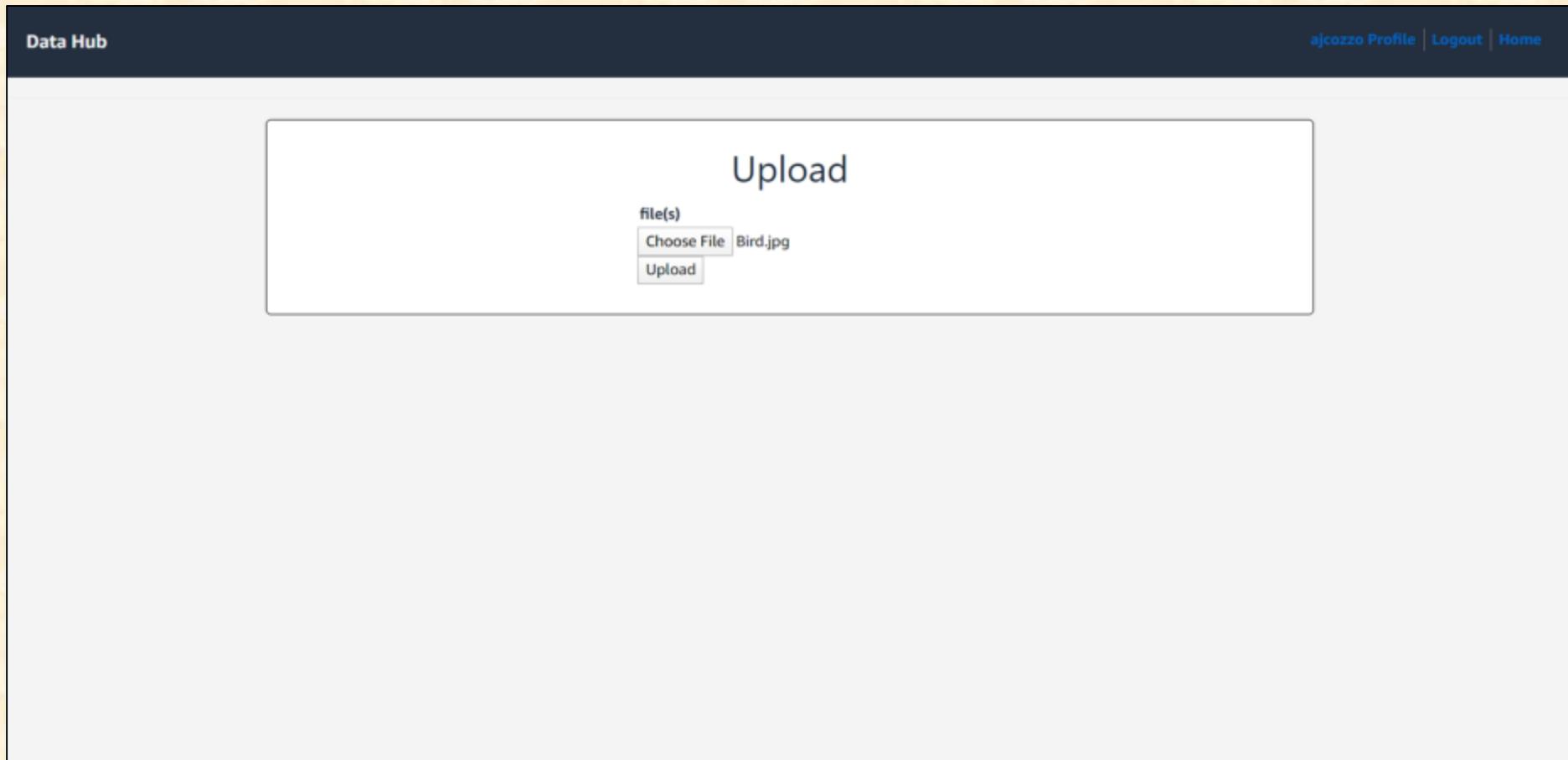
# System Architecture (we have to fix the thickness of the arrow)



# Search Page



# Upload Page



# Results Page Example 1

Data Hub [ajcaza Profile](#) | [Logout](#) | [Home](#)

## Results

Query

Count per page

**title: Annotated Road Ice**  
**version: V1.0**  
**file-type: .png**

**title: Annotated Road Ice**  
**version: V2.0**  
**file-type: .png**



# Results Page Example 2

Data Hub [ajcozzo Profile](#) | [Logout](#) | [Home](#)

## Results

Query

Count per page

**title:** Pictures of Birds  
**class0:** bird  
**class1:** bluebird

# What's left to do?

- Implement video datasets into ADH
- Expand into as many file types as possible
- Add in EMR functionality for larger datasets
- Complete Zip file upload and processing
- Allow for users to download Results/Data
- Visualization
- Parallelization
- Make web app feature complete



# Questions?

---

