

MICHIGAN STATE

U N I V E R S I T Y

Alpha Presentation

Document Management at Google Scale

The Capstone Experience
Team Technology Services Group

Ali Alaali
Joe Wan
Justin Newman
Luke Kline
Rohit Sen

Department of Computer Science and Engineering
Michigan State University
Fall 2019



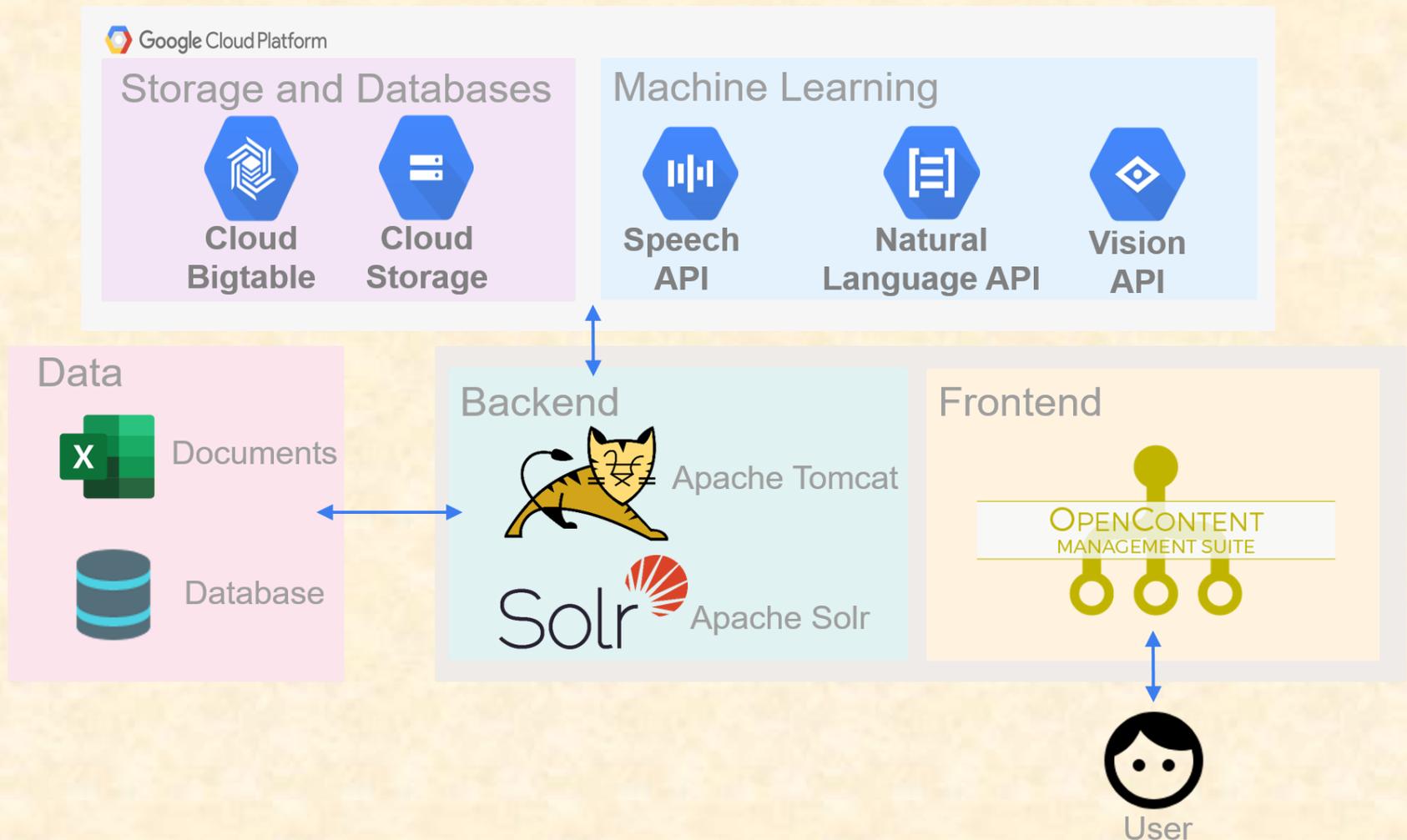
*From Students...
...to Professionals*

Project Overview

- Technology Services Group provides content management services
- Our project:
 - Research and development with Google Cloud Platform
 - Introduce speech-to-text
 - Implement image search
 - Provide document context search



System Architecture



Video with Transcription

The screenshot displays the OpenContent Management Suite (OCMS) interface. The browser address bar shows the URL: localhost:8080/ocms/Stage/controleddocs/d816823c-2947-43a4-9b03-41f7e5dcae6a-0.1%7C00f44294-8ff3-4ff6-8b1b-5008ffbe8874-0.1. The OCMS navigation bar includes options like Search, Collections, Dashboard, and a user profile for 'admin admin' with a 'Sign Out' button. Two document windows are open: 'justin-video-text' and 'justinmp4'. The 'justin-video-text' window shows a transcription of a video, with the following text:

5.0s: or something backed into me can
7.6s: see it's dented their in the
8.9s: front and the impact also
11.9s: seemed to raise the hood of my
14.3s: car is about in an inch Gap
17.0s: there now also scraped up my
21.3s: front bumper here and the
24.0s: other side is also pretty
25.3s: raise then now I know it's

The 'justinmp4' window shows a video player with a close-up image of a car's front end, including the hood and headlight. The video player has a progress bar at the bottom showing a time of -0.23.



Searching in a PDF

The screenshot shows a web browser window with the address bar displaying `localhost:8080/ocms/Stage/controlleddocs/d816823c-2947-43a4-9b03-41f7e5dcae6a-0.1%7CF704f8f1-907d-4baf-aa49-1a03d422a20b-0.1`. The browser tabs include "OpenContent Management Suite" and "phoneCall.pdf". The browser's address bar also shows "Apps", "phoneCall", and "justinVideo".

Below the browser window is a navigation bar with the following items: "Search", "Collections", "Dashboard", "Create New Folder", "Add Document(s)", "Create Placeholder Document(s)", "Hello, admin admin", and a "Sign Out" button.

The main content area is split into two panels. The left panel, titled "justin-phone-call-text", shows a PDF viewer with a search overlay. The search overlay has a "TEXT SEARCH" section with a search box containing "payout". Below the search box, it says "Match 1 of 1, matches across 1 page(s)". The search results show "Page #: 1" and "28.2s: \$600 payout and a courtesy car". The PDF viewer also has a "Summary" and "Search" tab, and a "Mode" dropdown menu.

The right panel, titled "justin-phone-call", shows a video player with a progress bar at the bottom indicating a duration of -0:52.



Searching Document Content

The screenshot displays the OpenContent Management Suite search interface. The browser address bar shows the URL: localhost:8080/ocms/search/controleddocs/Controlled%20Document?paramName=allproperties¶mValue=hood¶mType=allproper... The search results page shows 4 results for the search term 'hood'. The results table includes columns for Document Number, Version Label, Status, Approved Date, and Effective Date. The search term 'hood' is entered in the search box, and the search button is visible. The interface also includes navigation links like 'Prev', 'Next', and '50/Page', and a 'Quick Filter' button.

Search returned 4 results.

<input type="checkbox"/>	Document Number	Version Label	Status	Approved Date	Effective Date
<input type="checkbox"/>	justin text	0.1	Draft		
<input type="checkbox"/>	justin-under-the-h...	0.1	Draft		
<input type="checkbox"/>	justin-under-the-h...	0.1	Draft		
<input type="checkbox"/>	justin-video-text	0.1	Draft		



Result From Content Search

The screenshot displays the OpenContent Management Suite (OCMS) interface. The browser address bar shows the URL: localhost:8080/ocms/Stage/controleddocs/d816823c-2947-43a4-9b03-41f7e5dcae6a-0.1%7C00f44294-8ff3-4ff6-8b1b-5008ffbe8874-0.1. The OCMS navigation bar includes links for Search, Collections, Dashboard, Create New Folder, Add Document(s), and Create Placeholder Document(s). The user is logged in as 'admin admin' and can click 'Sign Out'.

The main content area shows the 'Controlled Docs' section for the document 'justin-video-text'. The 'Document Properties' table is as follows:

Document Properties	
Name	justin-video-text
Version Label	0.1
Status	Draft

The 'Actions' section for this document includes:

- View Properties
- Create New Folder
- Add Documents
- Delete Folder

The 'Related Objects' section is currently loading.

The document content is displayed in a viewer window titled 'justin-video-text'. The text is as follows:

7.6s: see it's dented their in the
8.9s: front and the impact also
11.9s: seemed to raise the hood of my
14.3s: car is about in an inch Gap
17.0s: there now also scraped up my
21.3s: front bumper here and the



What's Left to Do?

- Add relationship from transcription with parent audio file
- Option to transcribe on user's discretion
- Vision API – image search
- Natural Language API – document classification



Questions?

?

?

?

?

?

?

?

?

?

