

**MICHIGAN STATE**  
**UNIVERSITY**

# Alpha Presentation

## Document Management at Google Scale

**The Capstone Experience**  
Team Technology Services Group

Ali Alaali

Joe Wan

Justin Newman

Luke Kline

Rohit Sen



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

Department of Computer Science and Engineering  
Michigan State University  
Fall 2019

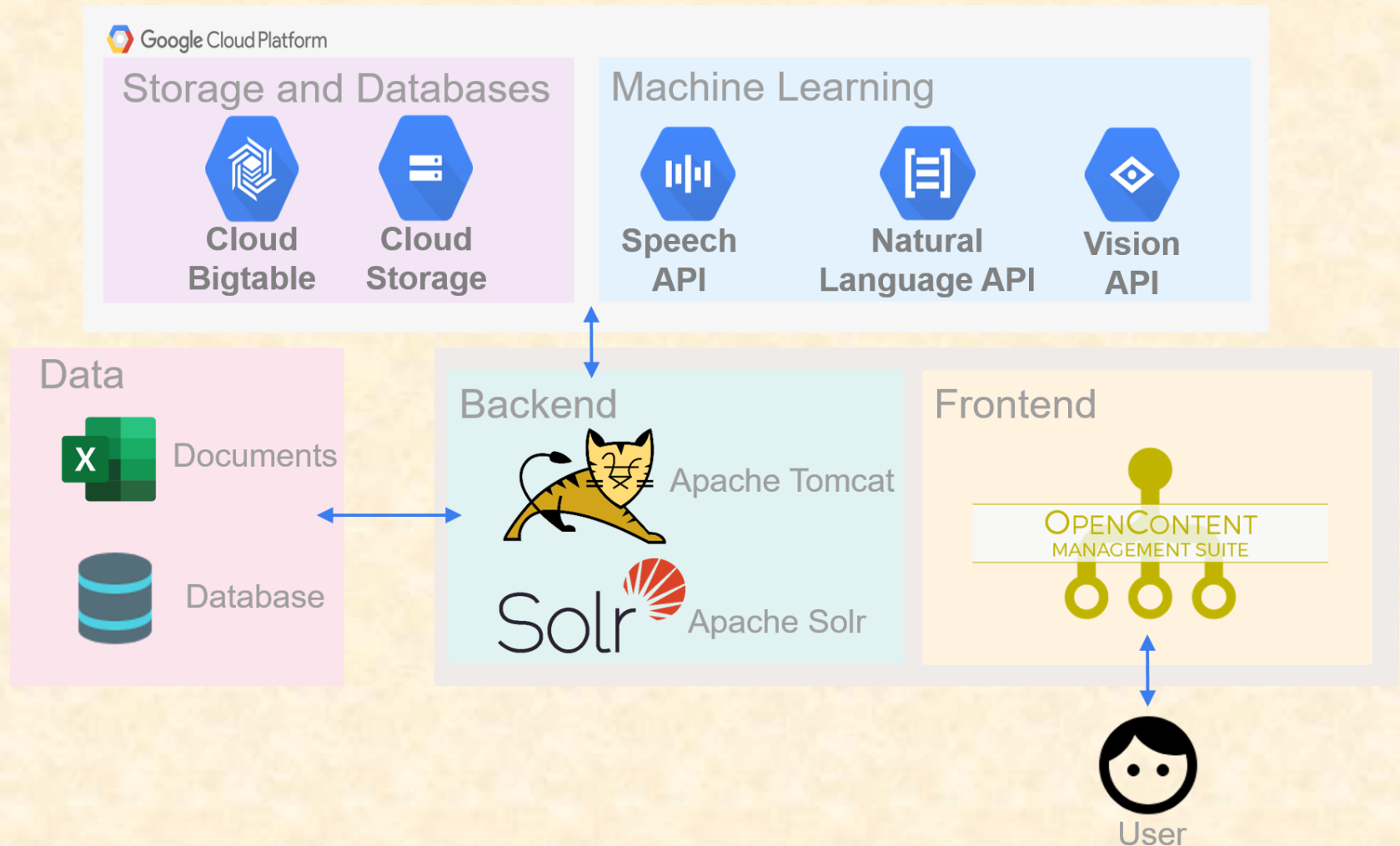
# Project Overview

---

- 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 a web browser window with the OpenContent Management Suite interface. The address bar shows the URL: `localhost:8080/ocms/Stage/controlddocs/d816823c-2947-43a4-9b03-41f7e5dcae6a-0.1%7C00f44294-8ff3-4ff6-8b1b-5008ffbe8874-0.1`. The browser tabs include "OpenContent Management Suite", "phoneCall.pdf", and "justinVideo.pdf". The OCMS navigation bar shows "Search", "Collections", "Dashboard", "Create New Folder", "Add Document(s)", "Create Placeholder Document(s)", and a user profile for "admin admin" with a "Sign Out" button.

Below the navigation bar, two document windows are open:

- justin-video-text**: This window displays a transcription of a video. The text is as follows:  
**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**
- justinmp4**: This window displays a video player showing a close-up of a car's front end, specifically the hood and headlight area. The video player includes a progress bar at the bottom with a timestamp of -0.23.



# Searching in a PDF

The screenshot displays a web application interface for searching within a PDF document. The browser window at the top shows the URL `localhost:8080/ocms/Stage/controledldocs/d816823c-2947-43a4-9b03-41f7e5dcae6a-0.1%7Cf704f8f1-907d-4baf-aa49-1a03d422a20b-0.1`. The application's top navigation bar includes links for Search, Collections, Dashboard, and document management actions. Two side panels are open:

- justin-phone-call-text**: This panel shows a text search interface. The search term "payout" is entered in the search bar. The results show "Match 1 of 1, matches across 1 page(s)". The specific match is "Page #: 1" with the text "28.2s: \$600 payout and a courtesy car".
- justin-phone-call**: This panel shows a video player interface with a progress bar at the bottom indicating a duration of -0:52.



# Searching Document Content

The screenshot displays the OpenContent Management Suite (OCMS) search interface. The browser address bar shows the URL: `localhost:8080/ocms/search/controlddocs/Controlled%20Document?paramName=allproperties&paramValue=hood&paramType=allproper...`. The OCMS navigation bar includes links for Search, Collections, Dashboard, Create New Folder, Add Document(s), Create Placeholder Document(s), and a user profile for 'admin admin' with a 'Sign Out' button.

The search results page shows a search for 'hood' with 4 results. The search bar on the left contains the text 'hood' and a 'Search' button. Below the search bar, there are options for 'Controlled Docs' and 'Controlled Document'. A 'Saved Search' section shows 'No saved searches'. An 'Attribute Search' section includes checkboxes for 'Search All Versions' (Yes/No), and input fields for 'Document Number', 'Title', and 'Status'.

The search results table displays the following data:

Document Number	Version Label	Status	Approved Date	Effective Date
justin text	0.1	Draft		
justin-under-the-h...	0.1	Draft		
justin-under-the-h...	0.1	Draft		
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/controlddocs/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 is titled 'Controlled Docs' and shows the 'Document Properties' for 'justin-video-text'. The properties are:

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

Below the properties, the 'Actions' section includes buttons for 'View Properties', 'Create New Folder', 'Add Documents', and 'Delete Folder'. The 'Related Objects' section is currently loading.

The document content, titled 'justin-video-text', is displayed in a large text area. 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?

---

?

?

?

?

?

?

?

?

?

