

---

# ISA 414 – Managing Big Data

## Lecture 20 – Cloud Computing and Storage (*Part II*)

Dr. Arthur Carvalho

[arthur.carvalho@miamioh.edu](mailto:arthur.carvalho@miamioh.edu)



MIAMI UNIVERSITY

Copyright © 2021 Arthur Carvalho

---

# Lecture Objectives

---

- Understand some business issues related to the use of cloud technologies
- Hands-on experience with some cloud services
  - Software as a Service
    - IBM Natural Language Understanding
    - Microsoft Azure Computer Vision

# Lecture Instructions

---

- Download the notebook “*Lecture 20.ipynb*” available on Canvas
  - Open that file with VS Code
- Download the file *iamaperson.jpg* from Canvas
- Make sure to add all the files to the same folder

# Lecture Instructions

---

- (Optional) Go to <https://cloud.ibm.com/login>
  - Click on “Create an account”
  - Fill in the application and create an account
    - To avoid fake accounts, you will be asked for your credit card info
      - That is why creating an account is optional
  - Check your email and confirm the registration
  - Sign in
  - After login, click on *Catalog* (top right)



# Cloud Service Models

---

- Cloud computing and storage
  - Crucial big-data analytics enabler
- Cloud = rental agreement (on demand)
  - Different business models
    - Different levels of engagement
    - Different levels of servicing
- Traditional cloud service models include:
  - Application (SaaS)
  - Platform (PaaS)
  - Infrastructure (IaaS)

# Cloud Service Models

---

## ➤ How do cloud technologies enable big-data analytics?

- Infrastructure (IaaS)
  - Mitigate the issue of purchasing/maintaining/updating hardware
    - This is “outsourced” to the cloud provider
  - An organization can easily expand/reduce the IT infrastructure
    - Resources are rented as needed
    - No waste of computational resources
- Platform (PaaS)
  - Focus on managing and using data-analytics platforms/environments
  - No time wasted on maintaining the physical devices running the platform
- Application (SaaS)
  - Has the potential to speed up the development of data-analytics solutions
  - Focus on using an application

# Cloud Service Models

---

- We will not have hands-on experience with IaaS in this course
  - Too many hardware/IT details that are beyond the scope of this course
- We experienced PaaS before with MongoDB cloud
- We shall now focus on SaaS tailored to (big) data analytics
  - AI services on IBM Cloud and Microsoft Azure

# Cloud Service Models

---

- Important business discussion
  - In-house vs cloud infrastructure
    - Considerations
      - Costs
      - Demand
      - Skills
      - Security



# Cloud Service Models

---

## ➤ Cost: in-house (on-premise) infrastructure

- Initial purchase and future upgrades (3-, 5-, 10-year plan)
  - Tough question: how to estimate the hardware needs?
    - Common approach: peak analysis
    - Example: previous ISA 414 cluster
      - 35 students working on 10 GB data sets simultaneously = 350 GB of RAM
      - 2 cores per student = 70 cores
- Real-state costs (machine room)
- Electricity bill (keep the machines and the cooling system on)
- Staff costs

# Cloud Service Models

---

## ➤ Cost: cloud

- Usually, cost per usage time
  - Cloud-service providers often have great cost calculators
    - Remove most of the cost-related uncertainty
- How to estimate the hardware needs?
  - One does not need to when using the cloud
  - Infrastructure is easily (and often automatically) adjustable
    - *E.g.*, one can often increase storage and computational power in a few clicks (or even have this automatically adjusted as needed)

# Cloud Service Models

---

## ➤ Demand

- Key question: will the infrastructure be heavily utilized?
  - Example: does it make sense to buy powerful machines to be used by ISA 414/514 students?
    - ISA 414/514 is taught twice a year (8 months)
- Underutilization
  - Often (not always) means unjustified investments
  - Utilization is not so easy to estimate
- Cloud infrastructure is likely a better option if the infrastructure will be underutilized now or in the future
  - Remember that cloud = pay as you need

# Cloud Service Models

---

## ➤ Skills

- Is there an in-house team capable of setting up and maintaining the infrastructure?
  - IT, security, database, network professionals
- If not, how much will hiring these professionals cost?
  - See the cost discussion
- Benefits of having an in-house, talented team:
  - Exclusivity
- Cloud also means “outsourcing” the maintenance team
  - Do you trust the cloud provider?

# Cloud Service Models

---

## ➤ Security

- How will the cloud provider secure the infrastructure (including stored data)?
  - Not only about protection against internal/external hackers
  - Where is the server located?
    - Where is the *cloud*?
      - Would you be comfortable with your bank storing data about you in a 3<sup>rd</sup> party computer?
    - Geopolitics matter
      - Some of the cloud platforms allow users to select servers' locations

# Cloud Service Models

---

## ➤ In-house vs cloud infrastructure

- It really is context dependent
  - There is no bullet-proof solution
- Good news: one does not have to choose one over another
  - Hybrid of cloud and in-house infrastructure is possible
    - Keep the crucial, sensitive data in an in-house database; send the rest to the cloud
    - Example: Best Buy buys extra storage capacity and processing power during Black Friday and Cyber Monday

---

# SaaS Case Study #1: Text Analytics

# Cloud Services

---

- Question: can we automate the analysis of sentiment and emotions in textual data?
  - Yes, we did it before!
    - We learned how to preprocess textual data using the bag-of-words approach
    - We learned how to calculate sentiment scores based on counts of positive and negative words
      - Very technical, yet naïve approach
  - Let's do it differently now
    - SaaS: IBM Cloud's Natural Language Understanding
    - <https://www.ibm.com/demos/live/natural-language-understanding/self-service/home>



# Lecture Instructions

- Let's experience some SaaS now
  - Search for the term “Natural Language Understanding”


The screenshot shows the IBM Cloud catalog interface. On the left is a dark sidebar with navigation links: 'Catalog', 'IBM Cloud catalog', 'Featured', 'Services', 'Software', and 'Consulting'. The 'Featured' link is highlighted. The main content area has a search bar at the top containing the text 'Natural language Understanding'. Below the search bar, there are two sections: 'Resource Results' and 'Catalog Results'. In the 'Resource Results' section, there is a result for 'Natural Language Understanding-fy Service'. In the 'Catalog Results' section, there are two results: 'Knowledge Studio' and 'Natural Language Understanding Node.js App'. The search bar and the 'Natural Language Understanding' result in the 'Catalog Results' section are highlighted with red boxes.

IBM Cloud

Natural language Understanding


Resource Results


[View all resource results](#)


 **Natural Language Understanding-fy**  
Service

Catalog Results

[View all catalog results](#)

 Knowledge Studio

 **Natural Language Understanding**

 Natural Language Understanding Node.js App

# Lecture Instructions

- Click on *Natural Language Understanding*
  - Select the region closest to where you are
  - Select Lite and click on “*Create*”

Create

About

Select a location

Select a location

Dallas (us-south) ▼

Select a pricing plan

Displayed prices do not include tax. Monthly prices shown are for country or location: United States

Plan	Features	Pricing
Lite	<p><b>30,000 NLU Items Per Month</b></p> <p>1 Custom Model</p> <p>Fixed API Rate Limit. See Standard plan for higher API Rate Limit</p> <p>NOTE: A NLU item is based on the number of data units enriched and the number of enrichment features applied. A data unit is 10,000 characters or less. For example: extracting Entities and Sentiment from 15,000 characters of text is (2 Data Units * 2 Enrichment Features) = 4 NLU Items. A custom model refers to an annotation model developed with Watson Knowledge Studio.</p>	Free

The Lite plan gets you started with 30,000 NLU Items per month at no cost. This plan also enables use of one custom model published through Watson Knowledge Studio.

Lite plan services are deleted after 30 days of inactivity.

# Lecture Instructions

- Click on *Manage*
- Copy the API Key and URL to a safe place

The screenshot shows the IBM Watson Natural Language Understanding console. The title is "Natural Language Understanding-fy" with a green "Active" status and an "Add tags" link. On the left, a sidebar menu has "Manage" selected, with other options: "Getting started", "Service credentials", "Plan", and "Connections". A blue information banner at the top right states: "The watsonplatform.net endpoint URLs are deprecated. Click 'Service credentials' to create new credentials, and then update your API calls. See [Update endpoint URLs from watsonplatform.net](\"#\")." Below this, a section titled "Start by viewing the tutorial" contains two links: "Getting started tutorial" (highlighted with a red box) and "API reference". The "Credentials" section shows the "API key:" field with a red box around the copy icon, and the "URL:" field with a red box around the copy icon. The URL is `https://api.us-south.natural-language-understanding.watson.cloud.ibm.com/instances/d60`. There are also "Download" and "Show credentials" links.

# Natural Language Understanding

- Copy your credentials from IBM Cloud to the notebook “*Lecture 20.ipynb*”

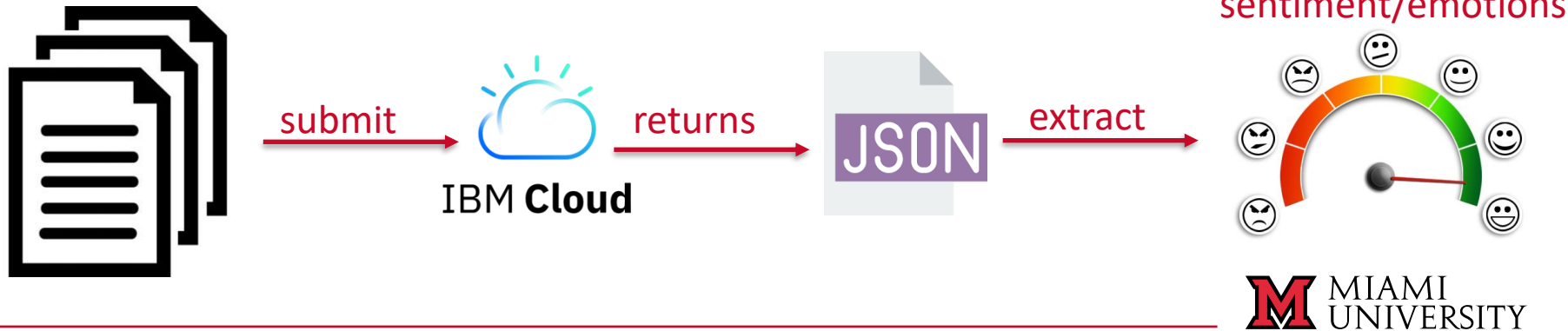


```
base_url =  
api_key =
```

# Cloud Services

- Let's experience IBM's Natural Language Understanding from inside Python
  - Illustrative problem: how did consumers' sentiment towards Starbucks after its pledge to hire refugees compare to consumers' sentiment before the pledge?

Facebook posts



# Cloud Services

---

- Natural Language Understanding API
  - Request: REST protocol (GET operation)
    - The text to be analyzed goes together with the base URL
  - Response: JSON
  - Documentation:  
<https://cloud.ibm.com/apidocs/natural-language-understanding>

# Cloud Services

## ➤ Natural Language Understanding API

- We have just signed for the IBM Natural Language service to get an API key
- Pricing scheme

Plan	Features	Pricing
Lite	30,000 NLU Items Per Month 1 Custom Model Fixed API Rate Limit. See Standard plan for higher API Rate Limit NOTE: A NLU item is based on the number of data units enriched and the number of enrichment features applied. A data unit is 10,000 characters or less. For example: extracting Entities and Sentiment from 15,000 characters of text is (2 Data Units * 2 Enrichment Features) = 4 NLU Items. A custom model refers to an annotation model developed with Watson Knowledge Studio.	Free
Standard	Unlimited NLU Items Per Month You will be charged per NLU Item & per Custom Model Increased API Rate Limit can be configured upon request NOTE: A NLU item is based on the number of data units enriched and the number of enrichment features applied. A data unit is 10,000 characters or less. For example: extracting Entities and Sentiment from 15,000 characters of text is (2 Data Units * 2 Enrichment Features) = 4 NLU Items. A custom model refers to an annotation model developed with Watson Knowledge Studio.	\$800.00 USD/Custom Model Instance per Month \$0.003 USD/NLU Item \$0.001 USD/NLU Item \$0.0002 USD/NLU Item
Premium Tier1	Everything in Standard plus... Usage and Training Data is Private + Stored in an Isolated Single Tenant Environment Transaction logging for service improvement is disabled by default High Availability and Service Level Agreements on Uptime IBM Cloud Service Endpoints HIPAA - Washington DC Only	

# Cloud Services

---

## ➤ Request & Response

```
import requests
```

```
content = "NEVER STARBUCKS EVER AGAIN"
```

```
key_values = {'version': '2021-03-25', 'text': content, 'features': 'sentiment'}
```

```
response = requests.get(base_url+"/v1/analyze",  
                        key_values,  
                        auth = ('apikey', api_key))
```

```
response.json()
```



# Cloud Services

---

## ➤ Reflection

- IBM Cloud's Natural Language Understanding allows one to easily analyze textual data
  - No need to understand/perform cumbersome preprocessing techniques
  - We can program the whole script using less than 20 lines of code
- Do you trust the obtained results?
  - Can we do better?
    - Training costs: OpenAI GPT-3 (175 billion parameters) = \$4.6 M
- How does NLU work?
  - Black-box model (proprietary technology)
  - How much do you trust IBM?

# Cloud Services

## ➤ Reflection: accuracy study

- Off-the-Shelf Technologies for Sentiment Analysis of Social Media Data: Two Empirical Studies (By Arthur Carvalho and Lucas Harris)

### Tweets

Technique/ Technology	Accuracy (Negative)	Accuracy (Neutral)	Accuracy (Positive)	Accuracy (Overall)
IBM NLU	6,696/7,341 [91.2%]	796/1,531 [52.0%]	1,369/1,508 [90.8%]	8,861/10,380 [85.4%]
Amazon Comprehend	4,902/7,341 [66.8%]	1,251/1,531 [81.7%]	1,391/1,508 [92.2%]	7,544/10,380 [72.7%]
Microsoft Text Analytics	5,036/7,341 [68.6%]	479/1,531 [31.3%]	1,361/1,508 [90.3%]	6,876/10,380 [66.2%]
Google Natural Language	5,705/7,341 [77.7%]	603/1,531 [39.4%]	1,384/1,508 [91.8%]	7,692/10,380 [74.1%]
Bag of Words	3,413/7,341 [46.5%]	1,073/1,531 [70.1%]	1,128/1,508 [74.8%]	5,614/10,380 [54.1%]

### Facebook Posts

Technique/ Technology	Accuracy (Negative)	Accuracy (Neutral)	Accuracy (Positive)	Accuracy (Overall)
IBM NLU	1174/1770 [66.3%]	39/133 [29.3%]	1128/1264 [89.2%]	2341/3167 [73.9%]
Amazon Comprehend	1258/1770 [71.1%]	63/133 [47.4%]	1096/1264 [86.7%]	2417/3167 [76.3%]
Microsoft Text Analytics	994/1770 [56.2%]	32/133 [24.1%]	1058/1264 [83.7%]	2084/3167 [65.8%]
Google Natural Language	1188/1770 [67.1%]	38/133 [28.6%]	894/1264 [70.7%]	2120/3167 [66.9%]
Bag of Words	612/1770 [34.6%]	57/133 [42.9%]	1005/1264 [79.5%]	1674/3167 [52.9%]

---

## SaaS Case Study #2: Image Analytics

# Cloud Services

---

- Question: can we automatically recognize objects (e.g., human beings) in an image?
- (Business) Applications
  - Surveillance
  - Autonomous cars
  - Aiding visually impaired people
    - Toy electric car + high resolution cams + speakers + Raspberry PI running an image-recognition software = a toy that can see and talk
- SaaS: Microsoft Azure's Computer Vision
  - <https://azure.microsoft.com/en-us/services/cognitive-services/computer-vision/>

# Cloud Services

---

- Let's experience Microsoft Azure's Computer Vision
- Illustrative story: you created a new dating website `iwannabeamiamimerger.com`
  - Constraint: each profile picture must have a person in it



# Cloud Services

---

- We shall focus on the last two steps
  - Request: REST protocol (POST operation)
    - We need to send (“post”) an image to Azure
    - Illustrative file: iamaperson.jpeg (available on Canvas)
  - Response: JSON
- I have already signed for the Computer Vision service to get an API key
  - You can use my API key today
  - The service is free, but unfortunately Azure requires credit card information to create an account

# Cloud Services

## ➤ Microsoft Azure's Computer Vision

- Documentation: <https://westcentralus.dev.cognitive.microsoft.com/docs/services/computer-vision-v3-1-ga>

- Pricing scheme:

INSTANCE	TRANSACTIONS PER SECOND (TPS)**	FEATURES	PRICE
Free - Web/Container	20 per minute		5,000 transactions free per month
S1 - Web/Container	10 TPS	Tag Face GetThumbnail Color Image Type GetAreaOfInterest	0-1M transactions — \$1 per 1,000 transactions 1M-10M transactions — \$0.65 per 1,000 transactions 10M-100M transactions — \$0.60 per 1,000 transactions 100M+ transactions — \$0.40 per 1,000 transactions
		OCR Adult Celebrity Landmark Detect, Objects Brand	0-1M transactions — \$1 per 1,000 transactions 1M-10M transactions — \$0.65 per 1,000 transactions 10M-100M transactions — \$0.60 per 1,000 transactions 100M+ transactions — \$0.40 per 1,000 transactions
		Describe+ Read	0-1M transactions — \$1.50 per 1,000 transactions 1M+ transactions — \$0.60 per 1,000 transactions

# Cloud Services

---

## ➤ Request & Response

- We must “post” the image to get a response

import requests

```
base_url = "https://arthur-carvalho-computer-vision.cognitiveservices.azure.com/vision/v3.0/analyze?visualFeatures=Tags"  
api_key = "5b6645a4c53c44498c66f2f4004e7e48"
```

```
image = {'media': open('iamaperson.jpg', 'rb')}  
authentication = {"Ocp-Apim-Subscription-Key":api_key}
```

```
response = requests.post(base_url, files= image, headers=authentication)
```

```
response.json()
```



# Cloud Services

---

## ➤ Reflection

- Microsoft Azure's Computer Vision allows one to analyze images
  - No need to understand cumbersome preprocessing techniques used in image analytics
  - One can build entire businesses/products around this service
  - Other possible services:
    - Detect gory or sexually suggestive content
    - Detect brands, celebrities, landmarks, handwritten text...
- Would you trust the obtained results?
  - Can we do better? (do we have the expertise?)

# Homework #9

---

- Time to implement the `iwannabeami.com` feature
  - Code to answer the question: is there a person in a picture?
    - Use the code in Slide 32 as a starting point
    - Save the JSON response to a variable (say, `response`)
    - Create a for-loop that iterates over the objects in `response["tags"]`
      - For each object, check whether there is a key-value pair `name:person`
        - Hint: `object["name"] == "person"`
      - If there is a `name:person` key-value pair, then check whether the confidence associated with the found key-value value is greater than 0.9
        - Hint: `object["confidence"] > 0.9`
      - If both conditions are true, then print *"Account Approved"*
    - Test your code using different images

# Summary

---

- There are several cloud technologies out there that can seriously improve big-data analytics process
  - From infrastructure to off-the-shelf software
- Many SaaS services are “black boxes”
  - We do not really know what is going on behind the scenes
  - This leads to the questions:
    - How much do you trust the service provider?
    - Do you feel comfortable using a technology without knowing what/how it analyzes data?
      - To a certain degree you do: have you ever checked the implementation of any R/SAS/Python function/package?
- Next lecture: Hadoop