
Personalized Twitter Audio

— Cloud Computing Final Project —

Haoran Ma, Chun-Yi Yang, Lizi Chen, Le Wang

Project Proposal

An Application enables users to listen to feeds from friends and channels they followed on social networks (Twitter).

Application Features

- Transfer tweets to audios
 - Elaborated audio that integrates information from twitter timeline
- Update periodically
 - Regular monitoring and scraping in a specific period
- Classify audio
 - A CNN support method to classify content into different category. Ex. sport, business, etc

Presentation Process

Cloud Architecture Design

- Tools
- Services Introduction

Machine Learning

- CNN learning on text classification
- Model implementation

Application

- Application deployment
- Application showcase

Cloud Service - AWS



Amazon EC2

- Virtual computing environments
- Various configurations of CPU, memory, storage, and networking capacity
- Static IPv4 addresses for dynamic cloud computing



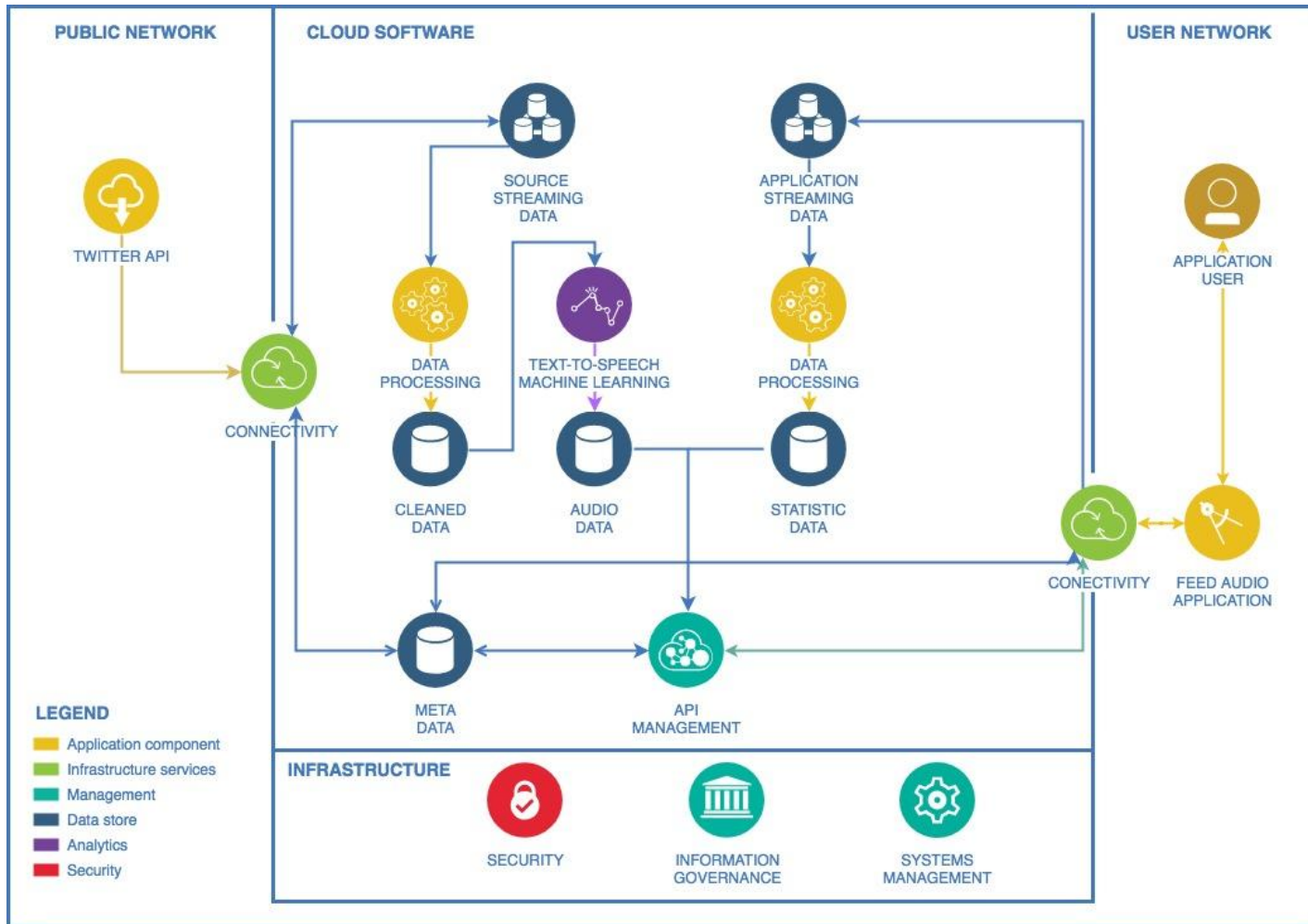
Amazon DynamoDB

- A NoSQL database service,
- Automatic data replication over three zones
- Designed for massive scalability
- DynamoDB delivers highly predictable performance



Amazon S3

- An online storage service
- Durability, availability and scalability
- Comprehensive security and compliance capability
- Flexible management, easy data transfer



EC2

Ubuntu 16.04-amd64-server

t2.micro

Launch Instance ▾

Connect

Actions ▾

Filter by tags and attributes or search by keyword

1 to 1 of 1

Name	Instance ID	Instance Type	Availability Zone	Instance State	Status Checks	Alarm Status	Public DNS (IPv4)	IPv4 Pu
	i-0da63a238c6e5bd87	t2.micro	us-east-1c	running	2/2 checks ...	None	ec2-54-90-73-104.com...	54.90.7...

Instance: i-0da63a238c6e5bd87

Public DNS: ec2-54-90-73-104.compute-1.amazonaws.com

Description

Status Checks

Monitoring

Tags

Instance ID	i-0da63a238c6e5bd87	Public DNS (IPv4)	ec2-54-90-73-104.compute-1.amazonaws.com
Instance state	running	IPv4 Public IP	54.90.73.104
Instance type	t2.micro	IPv6 IPs	-
Elastic IPs		Private DNS	ip-172-31-81-146.ec2.internal
Availability zone	us-east-1c	Private IPs	172.31.81.146
Security groups	launch-wizard-1 . view inbound rules	Secondary private IPs	
Scheduled events	No scheduled events	VPC ID	vpc-ea4f8a91
AMI ID	ubuntu/images/hvm-ssd/ubuntu-xenial-16.04-amd64-server-20180306 (ami-43a15f3e)	Subnet ID	subnet-60743a4f
Platform	-	Network interfaces	eth0
IAM role	-	Source/dest. check	True
Key pair name	ccProject	T2 Unlimited	Disabled
EBS-optimized	False	Owner	678663023884
Root device type	ebs	Launch time	April 27, 2018 at 12:31:24 PM UTC-4 (326 hours)
Root device	/dev/sda1	Termination protection	False
Block devices	/dev/sda1	Lifecycle	normal
Elastic GPU	-	Monitoring	basic
Elastic GPU type	-	Alarm status	None
Elastic GPU status	-	Kernel ID	-
		RAM disk ID	-
		Placement group	-
		Virtualization	hvm

Dynamodb

Twitter Schema

```
[
  "user_id": "andy",
  "datetime": "051002" // MONTH/DAY/HOUR
  "tweets": [
    {
      "category": "0",
      "created_at": "Thu May 10 02:45:02 +0000 2018",
      "favorite_count": "3",
      "quote_count": "6",
      "reply_count": "2",
      "retweet_count": "10",
      "screenName": "The Wall Street Journal",
      "text": "David Mayman has helped make sci-fi a",
      "tweet_id": "994407939959148546"
    },
    ...
  ] <!--END OF TWEET LIST-->
  ...
] <!--END OF USER LIST-->
]
```

TWTR [Close](#)

[Overview](#) **[Items](#)** [Metrics](#) [Alarms](#) [Capacity](#) [Indexes](#) [Global Tables](#) [Backups](#) [Triggers](#) [More](#) ▾

[Create item](#) [Actions](#) ▾

[Scan](#) ▾ [Table] TWTR: user_id, datetime ▾ ^

[+ Add filter](#)

[Start search](#)

Viewing 1 to 31 items

<input type="checkbox"/>	user_id	datetime	tweets
<input checked="" type="checkbox"/>	andy	051002	[{"M": {"cat...
<input type="checkbox"/>	andy	051004	[{"M": {"cat...
<input type="checkbox"/>	andy	051005	[{"M": {"cat...
<input type="checkbox"/>	andy	051014	[{"M": {"cat...
<input type="checkbox"/>	andy	051018	[{"M": {"cat...
<input type="checkbox"/>	andy	051020	[{"M": {"cat...
<input type="checkbox"/>	andy	051021	[{"M": {"cat...
<input type="checkbox"/>	andy	051022	[{"M": {"cat...
<input type="checkbox"/>	andy	051023	[{"M": {"cat...
<input type="checkbox"/>	andy	051100	[{"M": {"cat...
<input type="checkbox"/>	andy	051101	[{"M": {"cat...
<input type="checkbox"/>	andy	051102	[{"M": {"cat...
<input type="checkbox"/>	andy	051103	[{"M": {"cat...

S3

Bucket: cc-project-s3

Folder: user_id / time_category

Amazon S3 > cc-project-s3 / andy

Overview

Q

Type a prefix and press Enter to search. Press ESC to clear.

Upload

Create folder

More

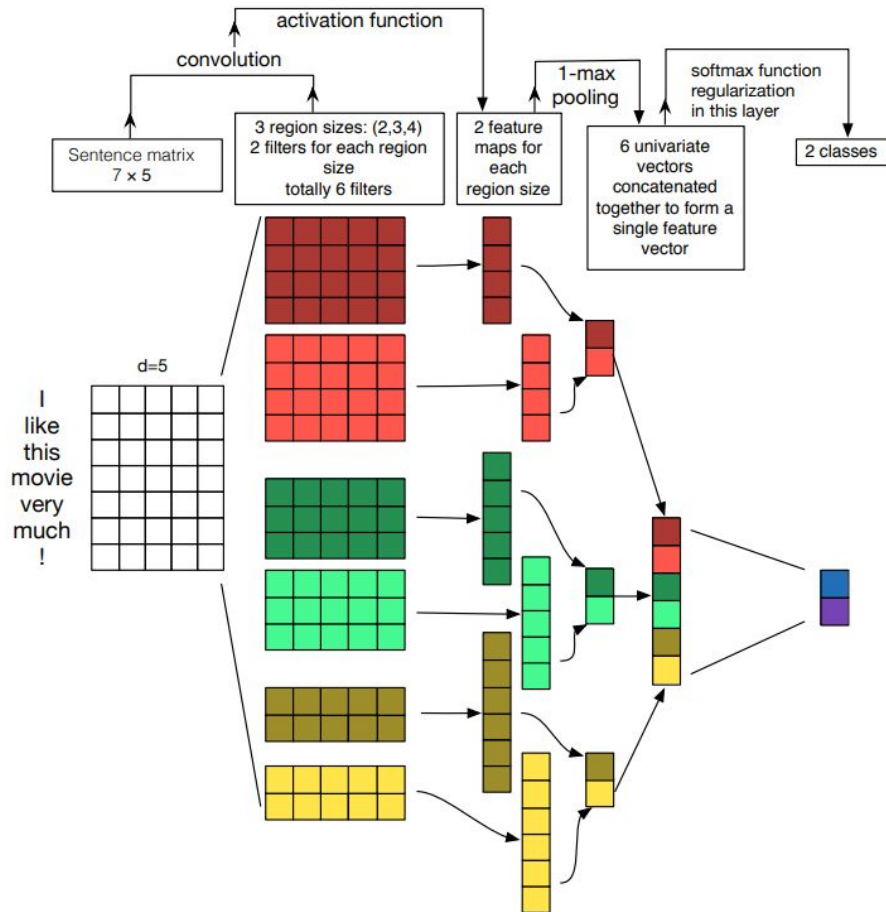
<input type="checkbox"/>	Name	Last modified	Size
<input type="checkbox"/>	0_business.mp3	May 10, 2018 8:00:29 PM GMT-0400	56.7 KB
<input type="checkbox"/>	0_science.mp3	May 10, 2018 8:00:28 PM GMT-0400	373.1 KB
<input type="checkbox"/>	0_summary.mp3	May 10, 2018 8:00:24 PM GMT-0400	843.6 KB
<input type="checkbox"/>	0_world.mp3	May 10, 2018 8:00:33 PM GMT-0400	415.7 KB
<input type="checkbox"/>	12_business.mp3	May 10, 2018 8:27:52 AM GMT-0400	50.9 KB
<input type="checkbox"/>	12_science.mp3	May 10, 2018 8:27:51 AM GMT-0400	126.7 KB
<input type="checkbox"/>	12_summary.mp3	May 10, 2018 8:27:50 AM GMT-0400	680.6 KB
<input type="checkbox"/>	12_world.mp3	May 10, 2018 8:27:58 AM GMT-0400	501.8 KB
<input type="checkbox"/>	14_business.mp3	May 10, 2018 10:10:27 AM GMT-0400	54.3 KB
<input type="checkbox"/>	14_science.mp3	May 10, 2018 10:15:24 AM GMT-0400	281.4 KB
<input type="checkbox"/>	14_sport.mp3	May 10, 2018 10:15:25 AM GMT-0400	84.1 KB
<input type="checkbox"/>	14_summary.mp3	May 10, 2018 10:15:21 AM GMT-0400	819.6 KB



News Classifier

- World
- Sports
- Business
- Sci/Tech
- 300,000 Training set
- 100,000 Testing set
- Precision: 92%
- Recall: 91%

Further Improvement:
Use Tensorflow Serving as a constant running service.



Deployment on AWS EC2

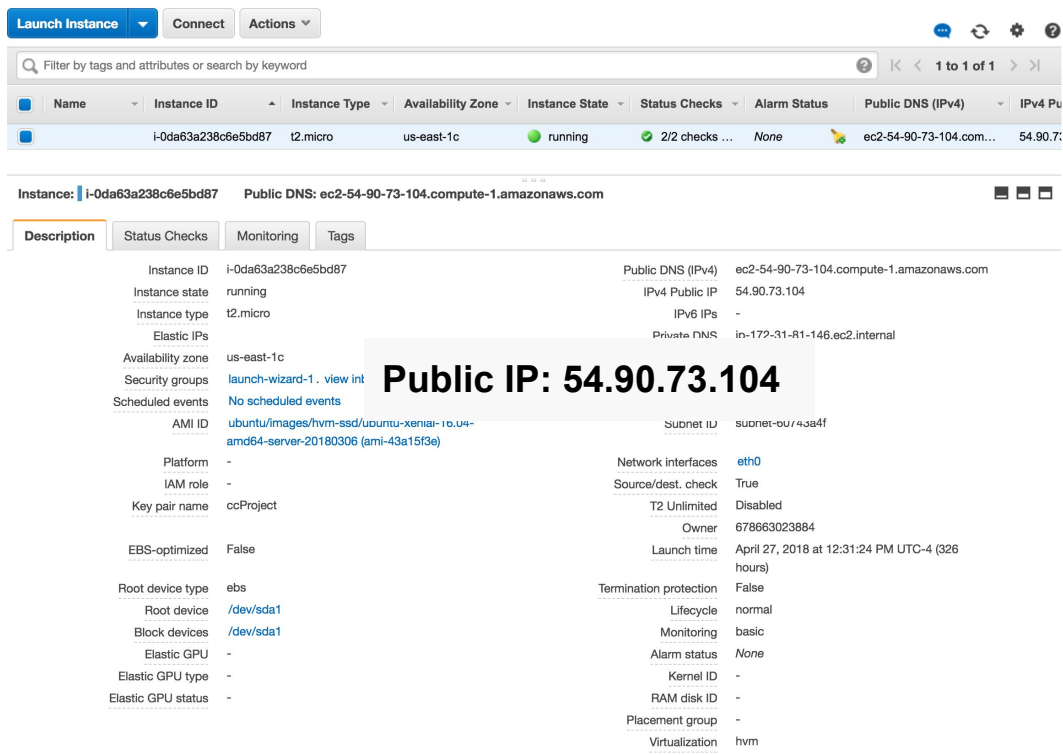
1. Setup EC2 on AWS
2. Connect to EC2

Code sample for flask deployment

```
from flask import Flask
app = Flask(__name__)

@app.route('/')
def hello_world():
    return 'Hello, World!'

if __name__ == "__main__":
    app.run(host="0.0.0.0", port=80)
```



The screenshot displays the AWS Management Console interface for an EC2 instance. At the top, there are buttons for 'Launch Instance', 'Connect', and 'Actions'. Below these is a search bar and a table of instances. The instance 'i-0da63a238c6e5bd87' is highlighted, showing it is a 't2.micro' instance in the 'us-east-1c' availability zone, currently in a 'running' state. Below the table, the 'Description' tab is selected, showing various instance details. A callout box highlights the 'Public IP: 54.90.73.104'.

Name	Instance ID	Instance Type	Availability Zone	Instance State	Status Checks	Alarm Status	Public DNS (IPv4)	IPv4 Pu
	i-0da63a238c6e5bd87	t2.micro	us-east-1c	running	2/2 checks ...	None	ec2-54-90-73-104.com...	54.90.7...

Instance: i-0da63a238c6e5bd87 Public DNS: ec2-54-90-73-104.compute-1.amazonaws.com

Public IP: 54.90.73.104

Description	Status Checks	Monitoring	Tags
Instance ID	i-0da63a238c6e5bd87	Public DNS (IPv4)	ec2-54-90-73-104.compute-1.amazonaws.com
Instance state	running	IPv4 Public IP	54.90.73.104
Instance type	t2.micro	IPv6 IPs	-
Elastic IPs		Private DNS	in-172-31-81-146.ec2.internal
Availability zone	us-east-1c		
Security groups	launch-wizard-1 · view int		
Scheduled events	No scheduled events		
AMI ID	ubuntu/images/hvm-ssd/ubuntu-xenial-16.04-amd64-server-20180306 (ami-43a15f3e)	Subnet ID	subnet-bu743e4f
Platform	-	Network interfaces	eth0
IAM role	-	Source/dest. check	True
Key pair name	ccProject	T2 Unlimited	Disabled
		Owner	678663023884
EBS-optimized	False	Launch time	April 27, 2018 at 12:31:24 PM UTC-4 (326 hours)
Root device type	efs	Termination protection	False
Root device	/dev/sda1	Lifecycle	normal
Block devices	/dev/sda1	Monitoring	basic
Elastic GPU	-	Alarm status	None
Elastic GPU type	-	Kernel ID	-
Elastic GPU status	-	RAM disk ID	-
		Placement group	-
		Virtualization	hvm