

## Sujet d'épreuves de la 48<sup>e</sup> Compétition Nationale des Métiers

# MÉTIER N°53 CLOUD TECHNOLOGIES

Soumis par :

Jade Marquette Experte Nationale WorldSkills France

Jonathan Bruneau Expert adjoint WorldSkills France

# DESIGN AND DEPLOYMENT OF A SECURE, SCALABLE IMAGE PROCESSING PIPELINE ON AWS

DUREE TOTALE DE L'ÉPREUVE	6 heures
---------------------------	----------

DIFFUSION DU SUJET	<i>Découvert le jour de la compétition</i>
--------------------	--

**Grading :** The grading for this part is 66 points.

You have the right to the Internet but no AI, whether in a code editor or online or in any other form. Should you be caught using this technology, the sub-section in question will be retrieved and the comment “AI cheats” with a grade of 0 will be awarded. If you're caught a second time, you'll be eliminated from the day's event or even from the competition.

If you're really stuck, or need a break, just raise your hand.

To complete this project, you have access to an AWS account. To demonstrate your ability to properly use this cloud provider, we require that you follow all best practices defined in the Well-Architected Framework.

# PART 1 - INTEGRATION AND DEPLOYMENT OF IMAGE PROCESSING PIPELINE USING AWS SERVICES

## CONTEXT

The company "Ctrl+Z", which specializes in image processing, continuously generates a large number of photos. The division director has tasked the IT department with deploying a serverless application to process these photos

To meet this requirement, you have been hired to implement the solution. As part of this, the development team has provided you with a script designed to meet the objectives set by your director.

## GOAL

Based on the technical details provided, you must design and deploy a cloud architecture to support the image processing pipeline. Your work will be assessed on how well it follows the specifications given and how closely it aligns with AWS best practices, including the Well-Architected Framework.

## DIAGRAM

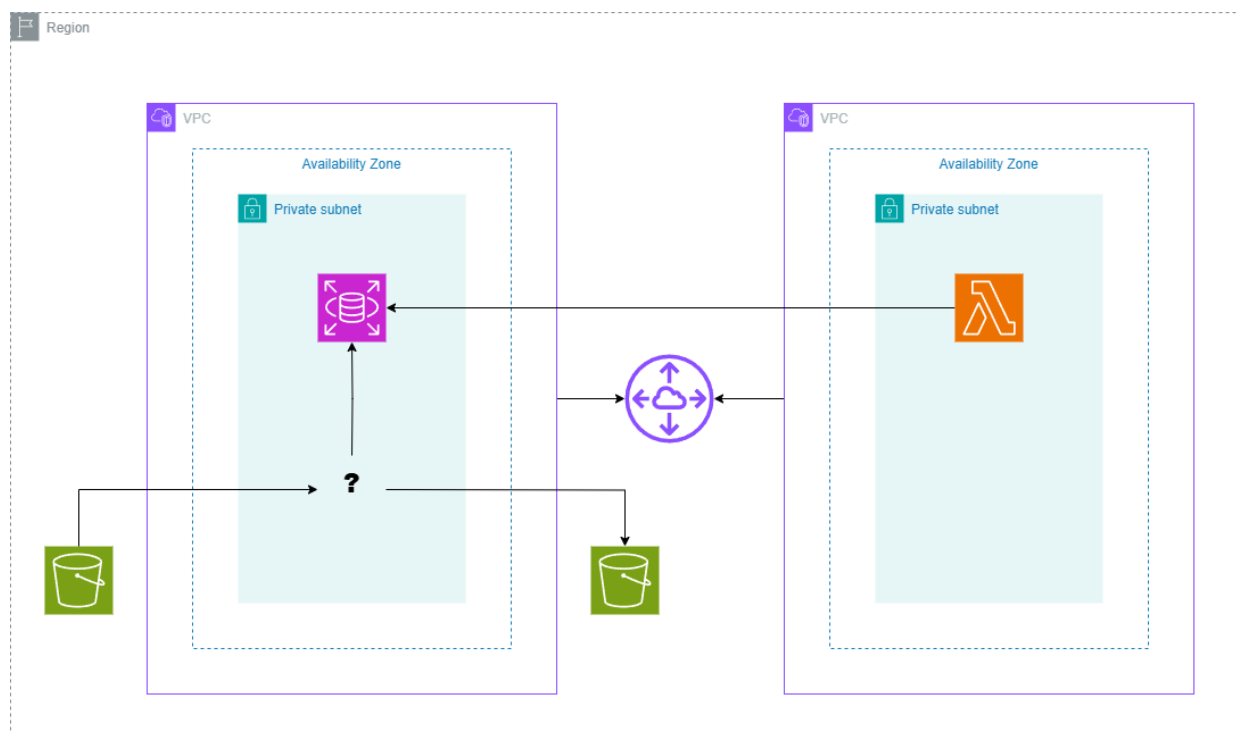


Figure 1 Non-detailed diagram – Step 1

## TECHNICAL DETAILS

### Storage

The infrastructure branch of the IT department will deploy an S3 bucket for you to receive the photos in your environment, as well as another S3 bucket where the director expects to retrieve the results.

To ensure proper integration with your colleagues' software, you are asked not to delete these shared storage spaces.

## Processing

The development branch of the IT department has provided you with the script to process the images. It runs using Python 3.12 and relies on well-known libraries, which are not included. Unfortunately for you, your contacts in that branch have gone on vacation...

From here, deploy the development team's program on the service you would recommend. Don't forget to configure the script to interface with the S3 buckets as well as with an SQL database.

Description	Valeur
Computing resource name and/or tag name	app-processing
Runtime	Python 3.12
Time estimated	>=10s
Memory	1 Gb

## Database

The image processing script relies on a MariaDB database to log the image processing tasks performed. You may choose any service you prefer to deploy this database.

The SQL scheme expected by the script is as follows:

```
CREATE DATABASE IF NOT EXISTS app_data
USE app_data;
CREATE TABLE IF NOT EXISTS file
(
  id INT PRIMARY KEY NOT NULL,
  content_key VARCHAR(260) NOT NULL,
  content_hash VARCHAR(256) NOT NULL
);
```

Description	Valeur
Database name and/or tag name	app_data
Engine	MariaDB
Version	11.4.5
Instance class	db.t3.micro
Allocated storage	10 Gb
Username	Admin

## Extraction

Your director sent you a script before you began your work to list the processing tasks that have been completed. According to him, the script once again runs on Python 3.12 and uses well-known libraries that are not included. He, too, has gone on vacation...

You must deploy this script on the Lambda service and configure it to interface only with the previously deployed database.

Description	Valeur
Function name and/or tag name	app-extraction
Runtime	Python 3.12
Memory	128 Mb

## Network

For this infrastructure, you need to deploy two VPCs.

- The first will be for the database and the service you choose to run the image processing script.
- The second will be for your director's script.

Considering that other services on different VPCs within the company may need to access the database, choose wisely how to interconnect the VPCs.

Description	Valeur
VPC name and/or tag name	app_vpc_a
VPC CIDR	10.0.0.0/16
Subnet1 name and/or tag name	subnet_a1
Subnet1 CIDR	10.0.1.0/24
Subnet2 name and/or tag name	subnet_a2
Subnet2 CIDR	10.0.2.0/24
Route table name and/or tag name	app_route_a

Table 1 VPC A

Description	Valeur
VPC name and/or tag name	app_vpc_b
VPC CIDR	10.1.0.0/16
Subnet1 name and/or tag name	subnet_b1
Subnet1 CIDR	10.1.1.0/24
Subnet2 name and/or tag name	subnet_b2
Subnet2 CIDR	10.1.2.0/24
Route table name and/or tag name	app_route_b

Table 2 VPC B

# PART 2 - ENHANCING SECURITY AND OBSERVABILITY IN AWS-BASED IMAGE PROCESSING WITH DYNAMODB INTEGRATION

## CONTEXT

The security branch has detected a violation of the company's security policy in the script you just implemented. To address this, they have modified the processing script and would like you to integrate a DynamoDB database into the infrastructure so they can log actions.

## GOAL

Following updated security requirements, you are required to modify the existing architecture to include a DynamoDB table for logging processing activity. Using the details provided, update the system to ensure scalability and reliability. Evaluation will be based on your ability to meet the scenario's technical constraints and apply AWS best practices appropriately.

## DIAGRAM

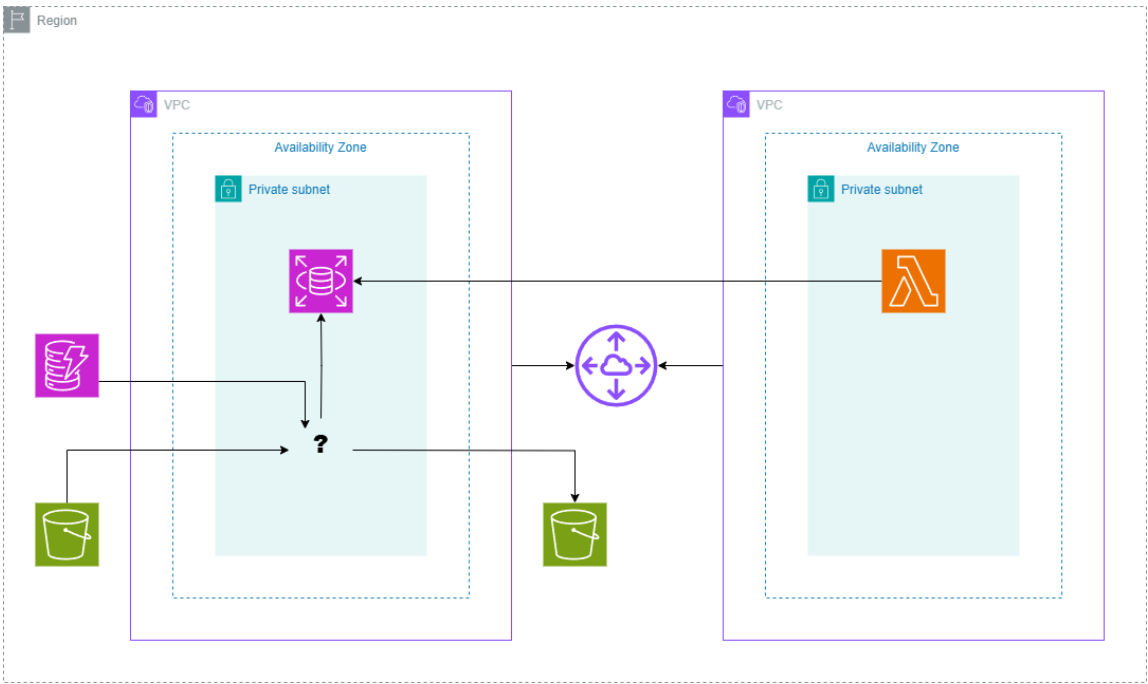


Figure 1 Non-detailed diagram – Step 2

## TECHNICAL DETAILS

### DynamoDB

The DynamoDB database will need to be deployed on your infrastructure account. The script will need to be configured once again to accommodate the new requirements. The DynamoDB database will have two keys:

- "key": The name of the processed image
- "hash": The encrypted value derived from the processed image

You are unsure how many requests will be made to this database, so deploy it with scalability in mind.

Description	Valeur
Table name and/or tag name	app-log
Partition key	key

# PART 3 - ECURING CREDENTIAL MANAGEMENT WITH AWS SECRETS MANAGER IN IMAGE PROCESSING PIPELINES

## CONTEXT

The security branch wants the database password to **not** be stored in the environment variables of the scripts you just deployed. They have therefore modified the scripts once again to introduce **Secret Manager** into your infrastructure.

## GOAL

To enhance credential security, adapt the infrastructure by integrating AWS Secrets Manager. Configure it to store and rotate the database password securely, and modify the script to use this new setup. Your solution will be reviewed based on how well it respects the given requirements and incorporates security best practices from the AWS Well-Architected Framework.

## DIAGRAM

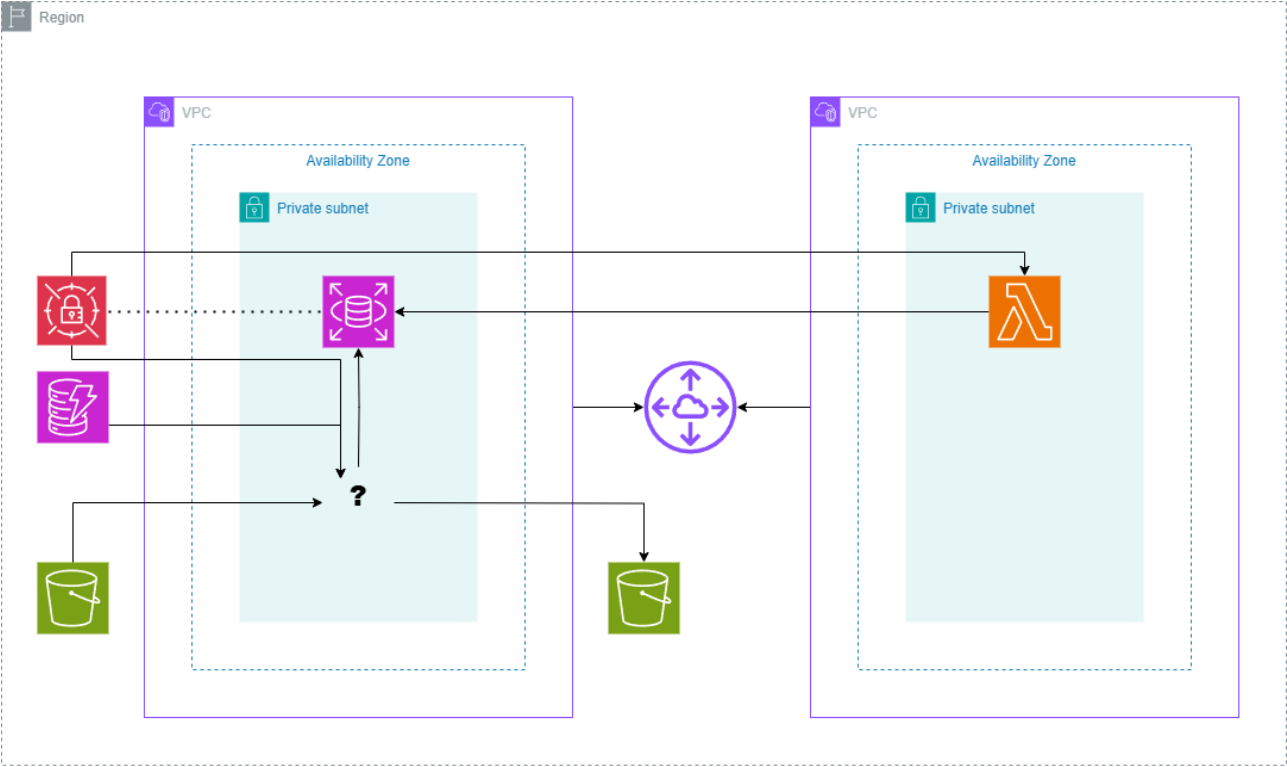


Figure 1 Non-detailed diagram – Step 3

## TECHNICAL DETAILS

### Secret Manager

Deploy a secret on Secret Manager and modify the script to account for it. The script will need to be configured once again to meet the new requirements.

Description	Valeur
Secret name and/or tag name	app-secret
Rotation	Every 14 days

# CLOUD DATA PROCESSING AND ANALYTICS WITH SERVERLESS AND STREAMING TECHNOLOGIES

DUREE TOTALE DE L'ÉPREUVE

5 heures

DIFFUSION DU SUJET

C2

*Découvert le jour de la compétition*

## PART 1 - SEMI-SYNCHRONOUS, IOT-STYLE DATA PROCESSING PIPELINE USING SERVERLESS SERVICES

<i>Do I keep this document</i>	<i>This document can be kept throughout the competition.</i>
<i>Allocated time</i>	<i>2h45 (3h00 with the break)</i>
<i>Grading</i>	<i>The grading for this first part is out of 36 points, but for greater readability, it will be expressed out of 100 in the subject.</i>
<i>Goals of this part</i>	<i>The goal for this morning is to set up a semi-synchronous, IoT-style data processing pipeline using serverless services. The objective is not just to deploy for the sake of it, but to reason through the architecture, understand the services, write code, and approach the task the way a cloud engineer would in real-world scenarios.</i>
<i>Context</i>	<p><i>On April 9, 2026, a Cloud Summit will bring together numerous professionals and experts in the field. To ensure optimal monitoring of the event and enhance the participant experience, the organizers aim to implement a real-time data flow analysis system.</i></p> <p><i>The goal is to collect and process data on attendee traffic and engagement, optimizing event organization and providing better visibility into session attendance.</i></p> <p><i>Conference attendance tracking: Using entry checkpoints placed at the entrance of each conference room. These checkpoints record each entry and transmit their data every 5 minutes.</i></p> <p><i>Participant interactions: Each attendee has a tablet allowing them to view the event schedule, scan badges of other participants to exchange information, and navigate an interactive map. These interactions are also sent every 5 minutes.</i></p> <p><i>To minimize costs and ensure efficient data processing, data will be aggregated and stored in a database chosen by the organizers. The processing will occur in</i></p>



	<p><i>batches every 15 minutes, generating analyzable reports.</i></p> <p><i>This system will provide organizers with a semi-synchronous view of attendance and participant interactions while optimizing cloud resources used for data processing.</i></p>
--	---

### Breakdown of the session

- € 1.1 Discovering the deployed resources (0h30)
  - ⌞ Submission of responses at the end of the allocated time
- € 1.2 Retrieving data from SNS topics - corrections and additions (0h50)
  - ⌞ Data will be sent after 20 minutes to allow storage to begin
  - ⌞ Submission of responses at the end of the allocated time

*BREAK (15min)*

- € 1.3 Creating the database chosen by the cloud team (0h30)
  - ⌞ Submission of responses after 22min
- € 1.4 Writing data into the database (0h55)
  - ⌞ Data will be sent after 20 minutes to allow storage to begin
  - ⌞ Submission of responses at the end of the allocated time

### Grading

You will only be assessed on the explicitly stated points: written answers to questions starting with Qx.x.x and practical tasks that will be checked starting with Px.x.x.

As you will see, today's goal is to understand the flow of data from the application to visualization using serverless tools. Most of the assessment will be based on answering the questions rather than practical implementation. However, answering the questions will require hands-on practice.

Sheets will be submitted regularly to ensure you are not blocked from progressing in the test.

Your work from this morning will not impact the afternoon session, as only Part 1.3 will be reused. However, we will review your progress during the midday break to ensure that no one will be stuck.

You have the right to the Internet but no AI, whether in a code editor or online or in any other form. Should you be caught using this technology, the sub-section in question will be retrieved and the comment "AI cheats" with a grade of 0 will be awarded. If you're caught a second time, you'll be eliminated from the day's event or even from the competition.

If you're really stuck, or need a break, just raise your hand.

Important: Time is a tiebreaker, so submit your work as soon as you complete it. Once submitted, you will not be allowed to make any further changes. There are many questions for the given time, so answer as quickly as possible and refine your answers later. No points will be deducted for incorrect responses.

## SUB-PART 1: DISCOVERING THE DEPLOYED RESOURCES

Time of submission	Grade	Comment

<i>Do I keep this document</i>	<i>This document will be collected at the end of the time limit mentioned below or before if the competitor has finished.</i>
<i>Allocated time</i>	<i>0h30</i>
<i>Goals of this sub-part</i>	<i>The cloud infrastructure team has already deployed several resources for the event. The goal of this first part is to understand these resources by answering 24 questions. Answer briefly without much justification, time is short.</i>
<i>Context</i>	<i>As stated in the context, the main goal of this project is to collect and process data on attendee traffic and engagement in order to optimize event organization and provide better visibility into session attendance. At this stage, the focus is on conference attendance tracking and participant interactions. Some components have already been deployed, and your task is to understand how they work. This part of the project does not require any AWS deployment; it is mainly based on reading and analysis. Taking the time to study these elements carefully will help you save time later on, as the objective is to build a solid foundation before diving into more complex tasks.</i>

### SNS

The applications send data to **SNS topics**.

	Q1.1.1	How many SNS topics have been created?	
	Q1.1.2	What type of topics are they ("Details" section)?	
	Q1.1.3	What other types of SNS topics could exist that are not present here?	
	Q1.1.4	Do you think a different type of SNS topic should have been used for the deployed SNS topics? Justify your answer.	

	Q1.1.5	Based on the title of this section, which service is responsible for processing the messages arriving in the topics?	
	Q1.1.6	One of these resources already exists. What is its name?	

#### Consumer service analysis

	Q1.1.7	By looking only at the code of the service Q.1.1.6, what do you think is the name of the function called when the service is triggered?	
	Q1.1.8	In which section of the AWS console can you find information about the function being called?	Tab name above the code:  Subsection name:
	Q1.1.9	What is the programming language used?	
	Q1.1.10	What is the role of the <code>lambda_handler</code> function?	
	Q1.1.11	Where can you find the logs of this function? Which AWS service is specifically designed for this purpose?	
	Q1.1.12	What will be the name of the "directory" where the logs will be stored in the service mentioned in this section ?	
	Q1.1.1	Does the function require	

	3	environment variables? If yes, which ones?	
	Q1.1.1 4	Where can you add environment variables?	Tab name above the code:  Subsection name:
	Q1.1.1 5	What IAM role is assigned to the service?	
	Q1.1.1 6	What permissions does this role provide?	

### SNS and S3 Integration

	Q1.1.17	Based on your answers, what is the name of the S3 bucket where the data is written?	
	Q1.1.18	What happens when the SNS topic "networking interactions" receives data with the currently deployed infrastructure?	
	Q1.1.19	How would you implement a similar logic to process the messages?	
	Q1.1.20	What IAM role would you assign to your function? (General purpose not necessarily the exact name of the role/policy AWS)	

### DynamoDB integration

The chosen database for the project is **DynamoDB**.

	Q1.1.21	Does the database already exist?	
--	---------	----------------------------------	--

	Q1.1.22	What would be the best serverless solution to write data into DynamoDB every 15 or 30 minutes by collecting the data from the bucket mentioned, following the existing logic?	
	Q1.1.23	How would your solution be triggered? List the AWS service or services involved.	

## SUB-PART 2 - RETRIEVING DATA FROM SNS TOPICS - CORRECTIONS AND ADDITIONS

Time of submission	Grade	Comment

<i>Do I keep this document</i>	<i>This document will be collected at the end of the time limit mentioned below or before if the competitor has finished.</i>
<i>Allocated time</i>	<i>0h50</i>
<i>Goals of this sub-part</i>	<p><i>Ensure the first part of the target architecture (as shown in DIAG 2) is fully operational.</i></p> <p><i>Identify and correct issues in the existing Lambda function <code>la-atd-track-sns-to-s3</code>, and implement a second Lambda function to process data from the second topic similarly.</i></p>
<i>Context</i>	<i>In this part of the project, you will refine and extend the existing architecture by focusing on the initial stage of the data pipeline. Based on your previous analysis, you may have identified issues with the current implementation of the <code>la-atd-track-sns-to-s3</code> Lambda function—this is your chance to fix them. You'll also develop an additional Lambda function to handle another data stream, following the same processing pattern. Note that while some questions can be addressed without live data, others will require data to start flowing—which will happen approximately 20 minutes after the beginning of this phase.</i>

*Questions that do not require data reception*

	P1.2.1	The lambda <code>la-atd-track-sns-to-s3</code> currently lacks the necessary permissions to write logs to CloudWatch. Grant it the minimum required permissions to enable logging. The role name should still be the one displayed on diagram 2.
	Q1.2.1	What issue does the function have? Why won't it work when data starts being sent? How do you fix it ?

Fix the issue.

As noted in Sub-Part 1, data arriving on the SNS topic **Networking-Interactions** is lost because no service is processing it.

	P1.2.2	Deploy a new Lambda function, using the existing one as a reference, to handle data from the Networking-Interactions topic while following the structure defined in DIAG 2.
--	--------	---

*Questions that require data reception*

Once your Lambdas are properly configured, files should start arriving in the S3 bucket. What do you observe about the data? Read the five following questions before answering them.

	Q1.2.2	Is there a pattern in the file naming? fix it. Quickly justify.	
	Q1.2.3	What is the ID for data related to attendance tracking?	
	Q1.2.4	What is the ID for data related to networking interaction?	
	Q1.2.5	How is the data structured inside the S3 objects? If you were to create a database table for each SNS topic, what would be the column names or attributes?	

Debugging an issue with a conference checkpoint

A bug has appeared on one of the conference entry checkpoints.

	Q1.2.6	After reviewing the logs, what is the issue with this checkpoint? What is it doing that is causing errors?	
--	--------	--	--

## SUB-PART 3 - CREATING THE DATABASE CHOSEN BY THE CLOUD TEAM

Time of submission	Grade	Comment

<i>Do I keep this document</i>	<i>This document will be collected <b>before the end of the allotted time, after 22 minutes</b> or earlier if the competitor has finished before. A second document will then be given in exchange, and this one can be kept.</i>
<i>Allocated time</i>	<i>0h22</i>
<i>Goals of this sub-part</i>	<i>Understand the purpose of DynamoDB - learn and apply key concepts and terminology related to DynamoDB</i>  <i>Design the most appropriate data schema based on the provided context</i>
<i>Context</i>	<i>In this section, your objective is to explore how DynamoDB fits into the overall system and to develop a solid understanding of its key principles—such as partition keys, sort keys, access patterns, and data modeling strategies. Based on your analysis, you will propose a schema that best matches the requirements of the use case. Once your schema is submitted, a reference version will be shared with everyone to ensure alignment before moving forward with implementation.</i>

### Understanding DynamoDB

These questions will not be marked. However, it is advisable to know how to answer them in order to answer this sub-section correctly.

- What type of database is DynamoDB? Is it relational or non-relational? Which subtype ?
- Define an attribute in DynamoDB.
- Define a Partition Key in DynamoDB.
- Define a Sort Key in DynamoDB.
- What is a composite primary key in DynamoDB? What are its advantages?
- What are the risks of choosing an inappropriate Partition Key when dealing with a large volume of data?

### Design the most appropriate data schema

the following table shows the data formats processed:

Attendance tracking	<code>atd-track;user_id;room_id;timestamp</code>
Conference Feedback	<code>fdb-conf;user_id;timestamp;conf_id;conf_grade;comment</code>
Food Feedback	<code>fdb-food;user_id;timestamp;food_corner_id;dish_type;dish_grade;waiting_time_grade;comment</code>
Network Interaction	<code>ntw-inte;user_id;contact_id;timestamp</code>



For now, only two formats will be used in Part 1. The other two will be used in Part 2 later this afternoon. However, we will create all four tables in this section.

Here's what the organizers want to do with each data table:

Table name	Important information for defining the schema and filling in the table below
atd-track	<ul style="list-style-type: none"> <li>– The main query will be to retrieve all users for a given conference.</li> <li>– Each user can be unique in the context of each conference.</li> </ul>
fdb-conf	<ul style="list-style-type: none"> <li>– Each user can submit only one review per conference.</li> <li>– Queries will retrieve all reviews for a given conference.</li> </ul>
fdb-food	<ul style="list-style-type: none"> <li>– Each user can submit only one review per food corner.</li> <li>– Queries will retrieve all reviews for a given food corner.</li> </ul>
ntw-inte	<ul style="list-style-type: none"> <li>– Each entry represents an interaction a user had at a specific timestamp.</li> <li>– Queries will retrieve all networking interactions for a user.</li> </ul>

	Q1.3.1	Based on your answers to the previous questions, and knowing the following information complete the following table. Be careful, the exercise is more tricky than it looks and is the most important question of this sub-part.
--	--------	---

Table name	Partition Key	Sort key
atd-track		
fdb-conf		
fdb-food		
ntw-inte		

	Q1.3.8	optional: Justify your choice of keys. If the table is correct, the answers will not be read. If the answers are wrong, the justification will be read and half the points may be awarded.	
--	--------	--	--

--	--	--	--

## SUB-PART 3 CONTINUED - CREATING THE DATABASE CHOSEN BY THE CLOUD TEAM

Time of submission	Grade	Comment

<i>Do I keep this document</i>	<i>This document can be kept throughout the competition.</i>
<i>Allocated time</i>	<i>minimum 8min or a little more if the competitor has handed in his question sheet early</i>
<i>Goals of this sub-part</i>	<i>Create the four DynamoDB tables using the information and schema provided.</i>
<i>Context</i>	<i>Now that the expected schema is clear and shared, your task is to implement it by creating the four DynamoDB tables accordingly. Pay close attention to the defined keys and structure, as these will directly impact how the data is queried and processed later in the pipeline.</i>

All tables will be in “On-demand” mode. Default values will be retained.

Table name	Primary key	Sort key
fdb-food	<i>food_corner_id</i>	<i>user_id</i>
atd-track	<i>conf_id</i>	<i>user_id</i>
fdb-conf	<i>conf_id</i>	<i>user_id</i>
ntw-inte	<i>user_id</i>	<i>timestamp</i>

	P1.3.1	Create the 4 DynamoDB tables described in the previous table with this information. The naming must be respected or you won't get the points.
--	--------	---

## SUB-PART 4 - WRITING DATA INTO THE DATABASE

Time of submission	Grade	Comment

<i>Do I keep this document</i>	<i>This document will be collected at the end of the time limit mentioned below or before if the competitor has finished.</i>
<i>Allocated time</i>	<i>0h50</i>
<i>Goals of this sub-part</i>	<p><i>Complete the architecture by creating the Lambda function <code>la-common-s3-to-dynamo</code>, a scheduler, and the IAM role <code>scheduler-role-invoke-lambda</code>.</i></p> <p><i>Gain familiarity with AWS SDK documentation, especially for understanding how to interact with services programmatically.</i></p> <p><i>Deepen your understanding of CRON expressions and REGEX, which are both essential tools you'll encounter regularly in a cloud engineering career.</i></p>
<i>Context</i>	<p><i>In this final subsection, your objective is to finalize the data pipeline by adding the last missing components. You will create the Lambda function <code>la-common-s3-to-dynamo</code>, which retrieves data from an S3 bucket, processes it, stores it in DynamoDB, and then deletes the source file to prevent unnecessary accumulation in the bucket. You'll also set up a scheduler and assign the appropriate IAM role (<code>scheduler-role-invoke-lambda</code>) to enable periodic execution.</i></p> <p><i>The Lambda code is available in the S3 bucket named <code>code-for-part-1-4</code>. Although the next batch of data will arrive 20 minutes after the start of this phase, the morning batch remains available for testing.</i></p> <p><i>This exercise is also an opportunity to practice reading and understanding the official AWS SDK documentation—an essential skill for any cloud engineer. Finally, you will encounter questions related to CRON expressions and regular expressions (REGEX), both of which are critical concepts in cloud automation and data processing workflows.</i></p>

	Q1.4.1	What roles does your Lambda function need? Justify your answer and keep in mind every action it takes.	
	P1.4.1	Create the IAM role for the Lambda function with the correct naming convention.	

	P1.4.2	Create the Lambda function using the zipped code and the created IAM role.
--	--------	--

In the "Code" tab, review the provided code.

	Q1.4.2	What is the purpose of the regex in the lambda_handler function? Would the code work without this line?	
	Q1.4.3	Study the regex in the lambda_handler function, ticking the 2 valid regexes and circling the errors in the other two. Don't be intimidated, it's simpler than it looks, there are no traps.	<p>⊘ ABC-123-XYZ-2024-04-04_12-30_2024-04-04_14-30.txt</p> <p>⊘ ABC-123-XYZ-2024-04-04_12-30_to_2024-04-04_14-30.txt</p> <p>⊘ file-5678-2023-12-31_23-59_to_2024-01-01_00-30.txt</p> <p>⊘ file_5678-2023-12-31_23-59_to_2024-01-01_00-30.txt</p>
	Q1.4.4	Briefly explain the purpose of the lambda_handler function. (TO CHANGE FOR A MCQ)	
	P1.4.3	Complete the delete_file_from_s3 function. Hint: only one line needs to be added.	
	P.1.4.4	Based on the save_to_dynamodb_batch_atd_track function, replace TO_REPLACE in the save_to_dynamodb_batch_ntw_inte function.	

Test your function to ensure no errors occur, that you have access to logs, and that the function runs correctly, even if no data has been sent yet.

	Q1.4.5	Which attributes does the <code>s3.get_object()</code> function return? Please refer to the documentation.	<p>✗ <code>DeleteMarker</code></p> <p>✗ <code>ContentType</code></p> <p>✗ <code>ChecksumSHA256</code></p> <p>✗ <code>ResponseExpires</code></p> <p>✗ <code>CacheControl</code></p> <p>✗ <code>ChecksumMode</code></p>
	P1.4.5	Create the IAM role for the scheduler.	
	Q1.4.6	What is the correct CRON expression for scheduling every 15 minutes from Monday to Friday between 08:00 and 20:00?	
	P1.4.6	Create the scheduler "scheduler-la-common-s3-to-dynamo" with the CRON expression from above. The execution time must be precise, as the organizers require execution every 15 minutes.	

### Conclusion

By now, your architecture should be fully aligned with the target diagram. If everything has been correctly implemented, data should be present in two of your DynamoDB tables, and the S3 bucket should be empty—indicating that the Lambda function successfully processed and cleaned up the files.

This marks the end of the morning session. Great job making it this far—you're now well on your way to mastering real-world cloud workflows!

## PART 2 - STREAMING DATA ANALYTICS WITH AWS

<i>Do I keep this document</i>	<i>This document can be kept throughout the competition.</i>
<i>Allocated time</i>	<i>2h15 (2h30 with the break)</i>
<i>Grading</i>	<i>The grading for this first part is out of 30 points, but for greater readability, it will be expressed out of 100 in the subject.</i>
<i>Goals of this sub-part</i>	<i>Collect and process feedback data directly from attendees' tablets in real time.</i>  <i>Write feedback data directly to the DynamoDB tables</i>  <i>Use serverless components to handle the ingestion, processing, and reporting pipelines.</i>  <i>Develop the Python logic for processing feedback and preparing the data for analysis.</i>  <i>Lay the groundwork for real-time visualization of the feedback data, which will be used by the organizers to adjust the event as it unfolds.</i>
<i>Context</i>	<i>As we continue with the Cloud Summit on April 9, 2026, we now shift to real-time data processing. The organizers are looking to gain immediate insights from conference feedback and food stand satisfaction, allowing them to adjust and optimize the event dynamically based on attendee experiences.</i>  <i>In this part of the project, we move away from using SNS for event-driven messaging. Instead, the goal is to directly capture and process feedback from attendees' tablets in real-time. This feedback will be written straight into the DynamoDB tables created earlier, enabling the organizers to generate real-time reports.</i>  <i>You'll explore how to build a real-time streaming pipeline to capture feedback as it happens, moving away from batch processing. The goal is to ensure the system can scale and handle multiple feedback streams simultaneously. You'll also be tasked with exploring how to process and visualize the data, making it actionable for the organizers.</i>  <i>You'll be introduced to new concepts and tools, such as capturing data from multiple sources simultaneously, processing it with low latency, and creating live reports. These tools are critical in building modern, highly responsive systems in the cloud, and understanding how to use them will be a key skill for future cloud engineers.</i>

### Breakdown of the session

€ 2.1 Create the Kinesis stream (0h20)

- ⌚ Submission of responses at the end of the allocated time.
- ⌚ 2.2 Deploy the lambda function (0h50)
  - ⌚ Data will start being sent 20 minutes after the beginning to allow storage to begin.
  - ⌚ Submission of responses at the end of the allocated time.

*BREAK (15min)*

- ⌚ 2.3 Visualize the data (0h50)
  - ⌚ Submission of responses at the end of the allocated time.
- ⌚ 2.4 Final Assessment Questionnaire (15 minutes)
  - ⌚ Submission of responses at the end of the allocated time.

### Grading

You will only be assessed on the explicitly specified points: written responses to questions beginning with Qx.x.x and practical tasks starting with Px.x.x, except for part 3, where the evaluation will be based on the presence and validity of the screenshots in the designated

Sheets will be submitted regularly to ensure you are not blocked from progressing in the test.

Your work from this morning will not impact the afternoon session, as only Part 1.3 will be reused. However, we will review your progress during the midday break to ensure that no one will be stuck.

You have the right to the Internet but no AI, whether in a code editor or online or in any other form. Should you be caught using this technology, the sub-section in question will be retrieved and the comment “AI cheats” with a grade of 0 will be awarded. If you're caught a second time, you'll be eliminated from the day's event or even from the competition.

If you're really stuck, or need a break, just raise your hand.

Important: Time is a tiebreaker, so submit your work as soon as you complete it. Once submitted, you will not be allowed to make any further changes. There are many questions for the given time, so answer as quickly as possible and refine your answers later. No points will be deducted for incorrect responses.



## SUB-PART 1: CREATE THE KINESIS STREAM

Time of submission	Grade	Comment

<i>Do I keep this document</i>	<i>This document will be collected before the end of the sub-part allotted time (20min), after 12 minutes or earlier if the competitor has finished before. A second document will then be given in exchange, and this one can be kept.</i>
<i>Allocated time</i>	<i>0h12</i>
<i>Goals of this sub-part</i>	<i>The objective of this section is to assess your understanding of the key differences between AWS SNS and Kinesis, as well as your ability to choose the appropriate service for a specific use case. Additionally, you will deepen your understanding of Kinesis streams, shards. This quiz will also test your ability to determine the appropriate configuration and scaling for a Kinesis stream based on data throughput requirements.</i>
	<i>In this section, you will study and create the Kinesis stream or streams, which will serve as the endpoint.s for receiving real-time data from various applications (such as event feedback or attendee interactions). The stream.s will act as a receptacle for incoming data, enabling efficient data processing downstream.</i>

	Q2.1.1	What is the main difference between SNS (Simple Notification Service) and Kinesis?	<p>A) SNS is used for one-time notifications to multiple subscribers, while Kinesis is used for real-time data streaming with continuous data processing.</p> <p>B) SNS and Kinesis are exactly the same; they are both used to send data to multiple systems.</p> <p>C) SNS is used for high-throughput data processing, while Kinesis is used for simple message delivery.</p> <p>D) Kinesis can only handle text-based data, while SNS handles both text and binary data.</p>
	Q2.1.2	Which of the following statements about Kinesis streams is true?	<p>A) Kinesis allows real-time processing of data by dividing it into multiple shards, each with a fixed throughput capacity.</p> <p>B) Kinesis can only handle data from SNS and cannot process data from other sources.</p> <p>C) Kinesis streams are limited to processing up to 10 records per second, regardless of the number of shards.</p> <p>D) Kinesis is primarily used for broadcasting notifications to multiple endpoints like SNS.</p>

	Q2.1.3	What is a shard in Kinesis?	<p>A) A shard is a fixed-size block of memory that holds data in Kinesis, allowing the service to scale its storage.</p> <p>B) A shard is a data partition in a Kinesis stream, responsible for handling a specific amount of data throughput (1MB/s for ingestion and 2MB/s for output).</p> <p>C) A shard is a message queue that holds data until the system is ready to process it.</p> <p>D) A shard is a type of compute resource used by Lambda functions to process Kinesis data.</p>
	Q2.1.4	If you need to process one line of data every 30 seconds in Kinesis, how many shards would you need?	<p>A) 1 shard, as each shard can handle 1000 records per second, which is far more than needed for one record every 30 seconds.</p> <p>B) 2 shards, since Kinesis requires at least two shards for any stream.</p> <p>C) 5 shards, as Kinesis requires additional shards for even the smallest amounts of data.</p> <p>D) 10 shards, because processing data every 30 seconds requires high throughput.</p>
	Q2.1.5	Should you create a separate Kinesis stream for each topic of data, or can you handle multiple topics within a single stream using Lambda?	<p>A) You must create a separate Kinesis stream for each topic to ensure data is handled correctly.</p> <p>B) You can handle multiple topics in a single stream by using Lambda to filter and process the data based on attributes (such as topic types or session IDs).</p> <p>C) You can only create a single Kinesis stream, as Lambda cannot process multiple topics.</p> <p>D) Lambda cannot process any data from Kinesis streams, so you must create separate streams for each topic.</p>

## SUB-PART 1 CONTINUED : CREATE THE KINESIS STREAM

Time of submission	Grade	Comment

<i>Do I keep this document</i>	<i>This document can be kept throughout the competition.</i>
<i>Allocated time</i>	<i>minimum 8min or a little more if the competitor has handed in his question sheet early</i>
<i>Goals of this sub-part</i>	<i>In this part, you will create a Kinesis stream</i>

	P2.1.1	Create the Kinesis stream named kin-stream-feedbacks with a retention period of 24 hours and set the shard count to 1.
--	--------	--

## SUB-PART 2: DEPLOY THE LAMBDA FUNCTION

Time of submission	Grade	Comment

<i>Do I keep this document</i>	<i>This document will be collected at the end of the time limit mentioned below or before if the competitor has finished.</i>
<i>Allocated time</i>	<i>0h50</i>
<i>Goals of this sub-part</i>	<i>Deploy a serverless Lambda function that processes real-time feedback data from a Kinesis stream. The function enriches the data using basic AI (NLP) and stores the output in DynamoDB for live monitoring. The goal of this section is to position you in the role of a Cloud Engineer specializing in cloud-native solutions. Even though some of the questions may seem to focus on specific technologies like NLP or AI, they are all deeply rooted in cloud technologies, particularly in building scalable, serverless systems. This exercise will deepen your understanding of how to leverage the cloud for real-time data processing and monitoring.</i>
<i>Context</i>	<p><i>This task is part of the real-time data ingestion pipeline. Feedback sent from attendee tablets is streamed through Amazon Kinesis. The Lambda function, triggered by this stream, processes the data using lightweight sentiment analysis before writing it to DynamoDB. This setup enables dynamic, serverless, and scalable event monitoring.</i></p> <p><i>You will find the zipped code for the Lambda function in the bucket named "code-for-part-2-2."</i></p>

	Q2.2.1	What roles does your Lambda function need? Justify your answer and keep in mind every action it takes.	
	P2.2.1	Create the IAM role for the Lambda function with the correct naming convention.	
	P2.2.2	Create the Lambda function using the zipped code and the created IAM role	
	Q2.2.2	How did you get Lambda to retrieve the data from the stream? Please explain.	

	Q2.2.3	<p>What is the purpose of the following code lines (under this table) ?</p> <p>Here are a few questions that might help you, but which are not marked:</p> <ul style="list-style-type: none"> <li>- In the "Code" tab, review the provided code. Why does the zipped folder contain multiple subdirectories such as regex and nltk?</li> <li>- Description of the NLTK library</li> </ul>	

```
# nltk.download('vader_lexicon', download_dir=os.getcwd())
nltk.data.path.append("/var/task/nltk_data")
from nltk.sentiment import SentimentIntensityAnalyzer
```

	P2.2.3	<p>In the function, replace <b>TO_REPLACE</b> with the correct values. Each value is used only once, and two are not used at all :</p> <ul style="list-style-type: none"> <li>⊄ NEUTRAL</li> <li>⊄ SentimentIntensityAnalyzer()</li> <li>⊄ allcap_differential(words)</li> <li>⊄ NEGATIVE</li> <li>⊄ sid.polarity_scores(comment)</li> <li>⊄ make_lex_dict()</li> <li>⊄ POSITIVE</li> </ul>	
	Q2.2.4	<p>What is the AWS equivalent of nltk for performing sentiment analysis, and how does it complement boto3?</p>	<p>A) AWS Lex – A chatbot service that provides sentiment analysis through natural language processing, integrated with boto3 for conversation management.</p> <p>B) AWS SageMaker – It is a machine learning service that requires manual model training for sentiment analysis, but can be used in combination with boto3.</p>

		<p>C) Amazon Comprehend – It is a fully managed service that performs sentiment analysis and integrates with boto3 for accessing AWS resources like DynamoDB.</p> <p>D) AWS Polly – It is a text-to-speech service and does not perform sentiment analysis, though it can be used with boto3.</p>
	P2.2.4	<p>In the lambda_handler function, based on the payload processing of fdb-food, replace the TO_DO section with the correct code to send fdb-conf data to DynamoDB.</p> <p><b>Informations about <i>fdb-conf</i>:</b></p> <p>The data lines follow this format:  <code>internal_id;user_id;timestamp;conf_id;conf_grade;comment</code></p> <p>To help you, the line above contains the attribute names in the correct format, but one attribute is missing. This missing attribute is added by the code, and its name is the same as the one used in <code>fdb-food</code>.</p> <p>Test your function to ensure no errors occur, that you have access to logs, and that the function runs correctly, even if no data has been sent yet.</p>

## SUB-PART 3: VISUALIZE THE DATA

Time of submission	Grade	Comment

<i>Do I keep this document</i>	<i>This document will be collected at the end of the time limit mentioned below or before if the competitor has finished.</i>
<i>Allocated time</i>	<i>0h50</i>
<i>Goals of this sub-part</i>	<p><i>Use Amazon QuickSight to visualize the real-time feedback data stored in DynamoDB. You will create interactive dashboards to visualize user reviews for conferences and food corners. The goal is to create clear and meaningful visualizations that help event organizers understand attendee sentiments and ratings.</i></p> <p><b>Submission Instructions</b></p> <p><i>The folder should be named in the following format:</i> <i>part2-comp-{id-comp}</i> (where {id-comp} is your competition ID of the day). <i>Place the folder in the Documents directory of your Windows user profile.</i></p> <p><i>The folder should include screenshots that visually verify the completion of each step. Ensure that your screenshots clearly display the calculated fields, aggregations, and visualizations in QuickSight as described in the tasks below.</i></p>
<i>Context</i>	<p><i>In this section, you will work with one of the data collected in DynamoDB. You will choose between those two distinct datasets:</i></p> <ul style="list-style-type: none"><li><i>– <b>fdb-conf:</b> Each attendee can submit one review per conference. Queries should retrieve all reviews for a given conference.</i></li><li><i>– <b>fdb-food:</b> Each attendee can submit one review per food corner. Queries will retrieve all reviews for a specific food corner.</i></li></ul> <p><i>The two subjects are quite similar in terms of the tasks involved but focus on different aspects of the event (conferences vs food corners). Each subject requires you to create meaningful visualizations that allow event organizers to make real-time decisions based on attendee feedback. The goal is to test your ability to use Amazon QuickSight to handle real-time data visualizations, regardless of whether you choose to work with fdb-conf (conference feedback) or fdb-food (food corner feedback).</i></p>

In this task, you will work with **hourly granular** data and create visualizations based on reviews for food corners or conferences.

**Create a folder named `part2-comp-{id-comp}` (where {id-comp} is your competition ID for the day) and place it in the Documents directory of your Windows user profile.**

Each question specifies what the screenshots should show, and the exact naming convention to follow. Make sure to copy-paste the screenshot names to ensure the correct format. If you need to capture multiple screenshots for the same question, append -1, -2, etc., at the end of the screenshot name.

Here are the two subjects. **Cross out the one you don't want.** The questions that follow are common to both subjects and can be adapted to your chosen table.

<b><i>Visualizing Conference Feedback</i></b>	<b><i>Visualizing Food Corner Feedback</i></b>
The data includes reviews from attendees for various conferences. Each attendee can submit only one review per conference.	The data includes reviews from attendees for various food corners. Each attendee can submit only one review per food corner.

### Introduction to QuickSight

QuickSight is AWS's cloud-based business intelligence (BI) service that lets you easily create visualizations and analyze data from various sources. In this task, you will connect your data stored in DynamoDB to QuickSight, perform some transformations, and create visualizations to analyze attendee feedback.

### Steps to create visualizations

	P2.3.1	<p>The first step is to connect your DynamoDB data source to QuickSight. This can be done directly via the QuickSight console. Make sure your DynamoDB table is accessible and connected to QuickSight.</p> <p><b>Screenshot Name:</b> P2-3-1-Connect.png</p> <p><b>What it should show:</b></p> <ul style="list-style-type: none"><li>– The DynamoDB connection to QuickSight.</li><li>– The correct DynamoDB table listed under data sources.</li></ul> <p>IMPORTANT: If you're stuck for too long, raise your hand for help. You'll only lose points on this question.</p>
	P2.3.2	<p>Once you've connected your data to QuickSight, you'll need to transform and aggregate it to derive useful metrics. In this case, you'll use calculated fields.</p> <p>Average Rating Calculation: You need to calculate the average rating for each conference or food corner by hour. This can be done by applying an aggregation function on the rating field.</p> <p><b>Screenshot Name:</b> P2-3-2-Avg-Rating.png</p> <p><b>What it should show:</b></p> <ul style="list-style-type: none"><li>– The calculated field for average rating by hour.</li><li>– The Average aggregation applied to the rating field.</li><li>– Grouping by timestamp for hourly data.</li></ul>
	P2.3.3	<p>You will also need to calculate the total number of reviews for each food corner or conference. To do this, use a count aggregation on the user_id field.</p>



		<p><b>Screenshot Name:</b> P2-3-3-Num-Reviews.png</p> <p><b>What it should show:</b></p> <ul style="list-style-type: none"> <li>– The calculated field for the total number of reviews.</li> <li>– The Count aggregation applied to the user_id field.</li> </ul>
--	--	---

#### Creating the visualizations

Once you have calculated the necessary fields, you will create the following visualizations in QuickSight.

	P2.3.4	<p>Bar Chart – Display the average rating for each food corner or conference.</p> <p><b>Screenshot Name:</b> P2-3-4-Bar-Chart.png</p> <p><b>What it should show:</b></p> <ul style="list-style-type: none"> <li>– A bar chart displaying the average rating for each food corner or conference.</li> <li>– Food corner or conference as the dimension (X-axis).</li> <li>– Average rating as the measure (Y-axis).</li> </ul>
	P2.3.5	<p>Line Graph – Show the sentiment trend over time for each food corner or conference.</p> <p><b>Screenshot Name:</b> P2-3-5-Line-Graph.png</p> <p><b>What it should show:</b></p> <ul style="list-style-type: none"> <li>– A line graph showing sentiment trends over time.</li> <li>– Timestamp as the dimension (X-axis).</li> <li>– Average sentiment score as the measure (Y-axis).</li> </ul>

## SUB-PART 4: FINAL ASSESSMENT QUESTIONNAIRE

Time of submission	Grade	Comment

<i>Do I keep this document</i>	<i>This document will be collected at the end of the time limit mentioned below or before if the competitor has finished.</i>
<i>Allocated time</i>	<i>0h15</i>
<i>Goals of this sub-part</i>	<i>The goal of this section is to evaluate your understanding of the cloud services and architecture components used throughout today's project. Through a series of short, scenario-based multiple-choice questions, you will demonstrate your ability to choose appropriate AWS services, configure them effectively, and understand how they interact in a real-time, serverless infrastructure.</i>
<i>Context</i>	<i>You will answer 12 multiple-choice questions in 15 minutes. These questions cover the key technologies and design choices applied during today's hands-on exercises, such as data streaming, serverless computing, event-driven architecture, and data visualization. You do not need internet access, as the focus is on testing your conceptual understanding and ability to reason about common cloud use cases and architectures.</i>

	Q2.4.1	<b>CRON Expression</b> Imagine you need to configure a CloudWatch Event Rule to execute a Lambda function every day at 2 AM from Monday to Friday. What would be the correct CRON expression for this task?	a) 0 2 * * 1-5 b) 2 0 * * 1-5 c) 0 2 * * 0-6 d) 0 2 * * 0-5
	Q2.4.2	<b>Choosing the Right Database DynamoDB or RDS?</b>  Use case: You are managing a real-time sales tracking application (eg. Amazon) where every sale needs to be recorded with a precise timestamp, product ID, and quantity. Users will mainly query sales summaries by product, rather than individual sales details.  Which database would you choose for this application? Choose the better justification.	a) DynamoDB, as it's designed for handling real-time data.  b) RDS, because it's relational and supports complex queries involving joins.  c) DynamoDB, because it allows for efficient queries on large datasets and handles partitioning automatically.  d) RDS, because sales data involves complex relationships.

	Q2.4.3	<p><b>Choosing Keys in DynamoDB</b></p> <p>Use Case:</p> <p>You are managing a task management application where each task is assigned to a user, and tasks are grouped by their project. The main queries will:</p> <ul style="list-style-type: none"> <li>– Retrieve all tasks for a specific user.</li> <li>– Retrieve all tasks within a project.</li> </ul> <p>What would be the best combination of Partition Key and Sort Key for this use case in DynamoDB?</p>	<p>a) Partition Key: user_id, Sort Key: task_id</p> <p>b) Partition Key: project_id, Sort Key: task_id</p> <p>c) Partition Key: user_id, Sort Key: project_id</p> <p>d) Partition Key: project_id, Sort Key: user_id</p>
	Q2.4.4	<p><b>Choosing DynamoDB vs RDS for a Solution</b></p> <p>You are tasked with building a user database for a mobile application. Users will log in at any time, and there are complex transactional use cases, such as tracking purchases, loyalty points, etc.</p> <p>Which database would you recommend and why?</p>	<p>a) DynamoDB, as it handles high-throughput operations well.</p> <p>b) RDS, as it can handle complex transactions and relational data.</p> <p>c) DynamoDB, for scalability and performance in managing real-time data.</p> <p>d) RDS, because it supports high-volume reads and writes.</p>
	Q2.4.5	<p><b>When to Use S3 for Data Storage</b></p> <p>Use Case:</p> <p>You are building a data pipeline for processing conference attendee data. The system needs to collect and store large files (CSV, JSON) before further processing them into a database. The files will be collected from different sources, including mobile apps and web applications, and will be processed asynchronously.</p> <p>Why would Amazon S3 be a good</p>	<p>a) S3 provides a relational database model with efficient querying.</p> <p>b) S3 allows for inexpensive, scalable storage of large files and integrates seamlessly with AWS services for processing.</p> <p>c) S3 is designed to store structured data, making it easier to query and process directly without needing additional storage solutions.</p> <p>d) S3 automatically processes data in real-time</p>

		choice for storing these files before processing?	and ensures high availability of files for processing.
	Q2.4.6	<p><b>Understanding SNS</b></p> <p>Use Case:</p> <p>You are working on a financial application that handles transactions. Whenever a user makes a large transaction (e.g., over \$10,000), the system needs to perform multiple tasks:</p> <ul style="list-style-type: none"> <li>– Log the transaction for auditing purposes.</li> <li>– Notify the user via email that the transaction has been processed.</li> <li>– Trigger a fraud detection process to evaluate whether the transaction is suspicious.</li> <li>– Notify the customer support team about high-value transactions for further review.</li> </ul> <p>Which SNS setup would you use to distribute the transaction alert to all relevant services?</p>	<p>a) Create a single SNS topic for "high-value transactions" and subscribe each service (logging, email, fraud detection, customer support) to this topic.</p> <p>b) Create a separate SNS topic for each service to avoid message duplication.</p> <p>c) Use an SNS topic for logging only and have the fraud detection and customer support services poll for updates from an SQS queue.</p> <p>d) Use multiple SNS topics for each type of transaction (e.g., one for high-value transactions, one for low-value transactions), and subscribe each service to both topics.</p>
	Q2.4.7	<p><b>Understanding Kinesis Streams</b></p> <p>What is the primary purpose of a Kinesis stream?</p>	<p>A) To perform complex queries on static datasets</p> <p>B) To store data indefinitely for large-scale storage solutions</p> <p>C) To enable real-time data streaming and processing with high throughput</p> <p>D) To deliver notifications to multiple endpoints in a pub/sub pattern</p>
	Q2.4.8	<p><b>SNS vs Kinesis Stream</b></p> <p>Use Case:</p> <p>You are developing a system that processes user actions on a high-traffic website. Every time a user clicks a button, submits a form, or</p>	<p>A) SNS is ideal because it allows you to send notifications to multiple subscribers (such as email, SMS, or another service) each time a user performs an action.</p> <p>B) Kinesis Streams is better because it allows you to stream user events in real-time, process the</p>

		<p>triggers a specific action, an event is generated. You need to track these events in real-time to calculate user engagement metrics, display live updates, and feed the data into an analytics platform for further processing.</p> <p>Which service would you choose and why?</p>	<p>data continuously, and feed the data into analytics platforms for aggregation and reporting.</p> <p>C) SNS should be used because it provides low latency and can handle a high volume of user actions without requiring special setup.</p> <p>D) Kinesis Streams is better for batch processing user events and should not be used for real-time event tracking.</p>
	Q2.4.9	<p><b>Lambda function roles</b></p> <p>When deploying a Lambda function that processes real-time feedback from a Kinesis stream and writes the output to S3 instead of DynamoDB, which IAM policy should be assigned to the Lambda function?</p>	<p>A) Read access to Kinesis, Write access to S3, and Log access to CloudWatch</p> <p>B) Full access to S3 and SNS</p> <p>C) Write access to Kinesis, Read access to DynamoDB</p> <p>D) Full access to Lambda, Kinesis, and DynamoDB</p>
	Q2.4.10	<p><b>Lambda data processing</b></p> <p>In a Lambda function processing Kinesis stream data, how does the function retrieve the data from the stream?</p>	<p>A) By invoking a Kinesis API call within the Lambda code</p> <p>B) By using a Kinesis trigger that automatically passes the stream records to Lambda</p> <p>C) By polling the stream every 30 seconds</p> <p>D) By manually reading from the S3 bucket that stores stream data</p>
	Q2.4.11	<p><b>Using Amazon Comprehend for Sentiment Analysis</b></p> <p>Imagine that instead of using the NLTK library for sentiment analysis, you decide to use Amazon Comprehend in your Lambda function to analyze attendees' comments in real-time.</p> <p>Which of the following code snippets correctly shows how to call Amazon Comprehend to analyze the sentiment of a text?</p>	<p>A)</p> <pre>import comprehend result = comprehend.get_sentiment(comment, 'en') sentiment = result.output</pre> <p>B)</p> <pre>import boto3 comprehend = boto3.resource('comprehend') sentiment = comprehend.detect(comment)</pre>

			<p>C)</p> <pre>import boto3  comprehend = boto3.client('comprehend')  response = comprehend.detect_sentiment(Text=comment, LanguageCode='en')  sentiment = response['Sentiment']</pre> <p>D)</p> <pre>from aws import comprehend  sentiment = comprehend.analyze(text=comment, lang='en')</pre>
	Q2.4.12	<p><b>Visualizing Ride Feedback in QuickSight</b></p> <p>Use Case:</p> <p>You're managing an amusement park that collects visitor feedback for each ride via digital kiosks. Each feedback includes the ride_id, timestamp, and a rating from 1 to 5. This data is stored in DynamoDB and streamed into QuickSight for visualization.</p> <p>You want to help park managers identify the most appreciated rides by time of day to optimize maintenance schedules and crowd control.</p> <p>Which type of aggregation should be used on the rating field to visualize this?</p>	<p>a) Count aggregation</p> <p>b) Average aggregation</p> <p>c) Sum aggregation</p> <p>d) Minimum aggregation</p>

# DEPLOYING A SCALABLE AND SECURE API INFRASTRUCTURE ON AWS WITH ECS, LOAD BALANCING, AND CDN

DUREE TOTALE DE L'ÉPREUVE	4 heures
---------------------------	----------

DIFFUSION DU SUJET	C3 <i>Découvert le jour de la compétition</i>
--------------------	--

**Grading :** The grading for this part is 68 points.

You have the right to the Internet but no AI, whether in a code editor or online or in any other form. Should you be caught using this technology, the sub-section in question will be retrieved and the comment “AI cheats” with a grade of 0 will be awarded. If you're caught a second time, you'll be eliminated from the day's event or even from the competition.

If you're really stuck, or need a break, just raise your hand.

To complete this project, you have access to an AWS account. To demonstrate your ability to properly use this cloud provider, we require that you follow all best practices defined in the Well-Architected Framework.

## CONTEXT

As part of securing data and ensuring traceability of information, your company needs to transform sensitive strings into cryptographically secure values using hash functions. This allows for storing the information without exposing raw data, while ensuring its uniqueness.

## GOAL

Based on the information provided, your task is to design and deploy the necessary infrastructure to support the company's API. Your work will be assessed on how well it follows the project instructions and applies AWS best practices.

# DIAGRAM

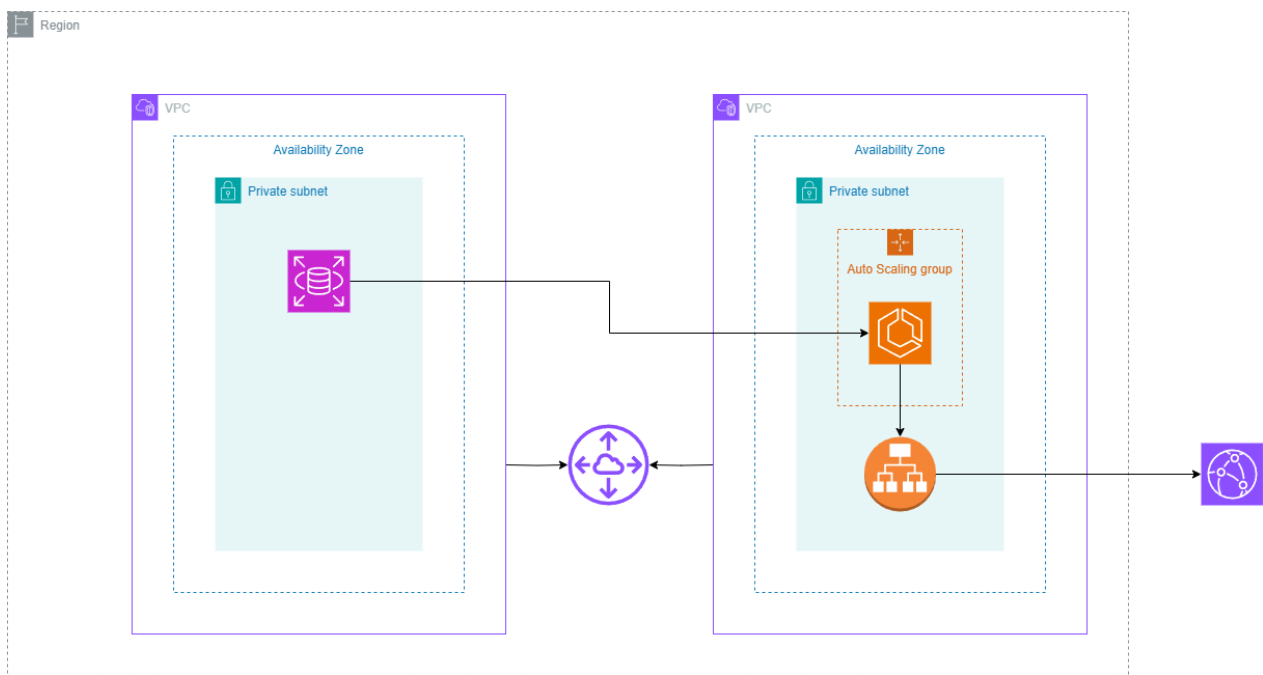


Figure 1 Non-detailed diagram

# TECHNICAL DETAILS

## Container

To run your company's API, a binary has been compiled on a Debian Bookworm image. Deploy this binary using ECS (Elastic Container Service) from an image that you will build for this purpose.

Description	Valeur
Cluster name and/or tag name	app-cluster
Task name and/or tag name	app-task
Service name and/or tag name	app-service
Launch type	Fargate
Max capacity	3
Min capacity	1
Auto scaling policy	CPU

## Registry

Your container image must be stored in a private registry on the AWS account provided by your company.

Description	Valeur
Repositories name and/or tag name	app _repository



## Database

The API relies on a MariaDB database to log the most recent requests made. You may choose any service you prefer to deploy this database. The database schema will be initialized by the binary, so there is no need to perform reverse engineering to define it.

Description	Valeur
Database name and/or tag name	app_data
Engine	MariaDB
Version	11.4.5
Instance class	db.t3.micro
Allocated storage	10 Gb
Username	Admin

## Load balancer

The binary provides an API using the HTTP protocol. Deploy a load balancer to distribute traffic across your containers. Choose and deploy the service you believe best meets this requirement.

Description	Valeur
Load balancer name and/or tag name	app-lb

## Network

For this infrastructure, you need to deploy two VPCs:

- VPC A for the database,
- VPC B for your containers.

Since other services located in different VPCs within the company may need to access the database, choose the method of VPC interconnection wisely.

Description	Valeur
VPC name and/or tag name	app_vpc_a
VPC CIDR	10.0.0.0/16
Subnet1 name and/or tag name	subnet_a1
Subnet1 CIDR	10.0.1.0/24
Subnet2 name and/or tag name	subnet_a2
Subnet2 CIDR	10.0.2.0/24
Subnet3 name and/or tag name	subnet_a3
Subnet3 CIDR	10.0.3.0/24
Route table name and/or tag name	app_route_a

Table VPC A

Description	Valeur
VPC name and/or tag name	app_vpc_b

VPC CIDR	10.1.0.0/16
Subnet1 name and/or tag name	subnet_b1
Subnet1 CIDR	10.1.1.0/24
Subnet2 name and/or tag name	subnet_b2
Subnet2 CIDR	10.1.2.0/24
Subnet3 name and/or tag name	subnet_b3
Subnet3 CIDR	10.1.3.0/24
Route table name and/or tag name	app_route_b

Table 2 VPC B

## CDN

To distribute your API across the USA and Europe, use a CDN (Content Delivery Network). Deploy the service you believe best meets this requirement.

# LISTE DES ANNEXES

*Description :*

*Document contenant les schémas d'architecture liés aux parties 1 et 2 du sujet C2. Ces schémas permettent aux compétiteurs de visualiser la structure cible de l'infrastructure à mettre en place*

## **ANNEXE 1 : DIAGRAMS – ARCHITECTURE REFERENCE C2-PART1**

**Fichiers inclus :**

48CNAT\_02\_Cloud\_Annexe1\_DIAG1.jpg – Final architecture of part 1

48CNAT\_02\_Cloud\_Annexe1\_DIAG2.jpg – Final architecture of sub-part 2

## **ANNEXE 2 : DIAGRAMS – ARCHITECTURE REFERENCE C2-PART2**

**Fichiers inclus :**

48CNAT\_02\_Cloud\_Annexe1\_DIAG3.jpg – Final architecture of part 2

48CNAT\_02\_Cloud\_Annexe1\_DIAG4.jpg – Combined architecture of part 1 and 2