

Sujet d'épreuves de la 48^e Compétition Nationale des Métiers

MÉTIER N°53

CLOUD COMPUTING

Jour 2

Soumis par :

Jade Marquette Expert WorldSkills France

CLOUD DATA PROCESSING AND ANALYTICS WITH SERVERLESS AND STREAMING TECHNOLOGIES

DUREE TOTALE DE L'ÉPREUVE

4 heures et 50 minutes d'épreuve

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 20,5 points</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</i></p>

	<p><i>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 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
	/5,5	

<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 23 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**.

0,5	Q1.1.1	How many SNS topics have been created?	
0,5	Q1.1.2	What type of topics are they ("Details" section)?	
0,5	Q1.1.3	What other type.s of SNS topics could exist that are not present here?	
0,5	Q1.1.4	Do you think a different type of SNS topic should have been used for the deployed SNS topics? Justify your answer.	

0,5	Q1.1.5	Which service is responsible for processing the messages arriving in the topics?	
0,5	Q1.1.6	One of these resources already exists. What is its name?	

Consumer service analysis

0,5	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?	
0,5	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:
0,5	Q1.1.9	What is the programming language used?	
0,5	Q1.1.10	What is the purpose of the <code>lambda_handler</code> function?	Main purpose:
0,5	Q1.1.11	Where can you find the logs of this function? Which AWS service is specifically designed for this purpose?	
0,5	Q1.1.12	What will be the name of the “directory” where the logs will be stored in the service mentioned in this section ?	
0,5	Q1.1.13	Does the function require environment variables? If yes, which ones?	
0,5	Q1.1.14	Where can you add environment variables?	Tab name above the code: Subsection name:
0,5	Q1.1.15	What IAM role is assigned to the service?	

0,5	Q1.1.16	What permissions does this role provide?	
-----	---------	--	--

SNS and S3 Integration

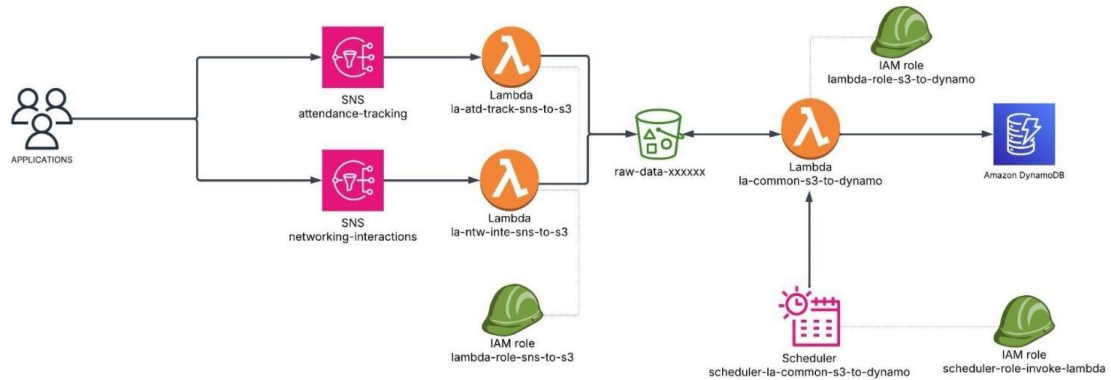
0,5	Q1.1.17	Based on your answers, what is the name of the S3 bucket where the data is written?	
0,5	Q1.1.18	What happens when the SNS topic "networking interactions" receives data with the currently deployed infrastructure?	

DynamoDB integration

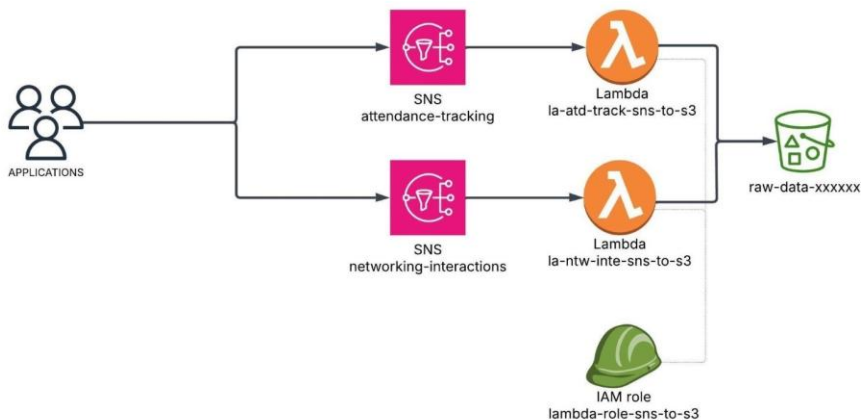
The chosen database for the project is **DynamoDB**.

0,5	Q1.1.19	Does the database already exist?	
0,5	Q1.1.20	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?	
0,5	Q1.1.21	How would your solution be triggered? List the AWS service or services involved.	

Given at the same time as C2 Part 1 sub-section 2 - this document can be kept once given



DIAG 1: final architecture of part 1



DIAG 2: final architecture of sub-part 2

*The bucket name is not necessarily the same as the one shown on the diagram, as the x's are replaced by numbers.

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

Time of submission	Grade	Comment
	/5,25	

<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>	<p><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></p>

Questions that do not require data reception

1	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.
0,5	Q1.2.1	What's wrong with the lambda function in its settings? 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.

1,75	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. The zip code of the first lambda function is available on the following bucket : "zip-code-bucket-xxx/code-01". Don't forget to connect the SNS topic to your new lambda function.
------	--------	--

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.

0,5	Q1.2.2	Is there a pattern in the file naming? What information does it provide (short explanation)	
0,25	Q1.2.3	What is the ID for data related to attendance tracking?	
0,25	Q1.2.4	What is the ID for data related to networking interaction?	
0,75	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?	data / file format : Network interaction column names: Attendance tracking column names:

Debugging an issue with a conference checkpoint

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

0,25	Q1.2.6	A tester informs you that some data arrives without the id of the room while checking the attendance tracking. How can you be informed of this problem so that it isn't noticed 20 minutes later? The best thing would be to be notified as soon as possible.	
------	--------	---	--

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

Time of submission	Grade	Comment
	/1	

<i>Do I keep this document</i>	<i>This document will be collected before the end of the allocated time, after 22 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>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	atd-track;user_id;room_id;timestamp
Conference Feedback	fdb-conf;user_id;timestamp;conf_id;conf_grade;comment

Food Feedback	fdb-food;user_id;timestamp;food_corner_id;dish_type;dish_grade;waiting_time_grade;comment
Network Interaction	ntw-inte;user_id;contact_id;timestamp

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

1	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
	/1	

<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>

1	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
	17,75	

<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>

0,25	Q1.4.1	What roles (policy) does your Lambda function need? Justify	
------	--------	---	--

		your answer and keep in mind every action it takes.	
0,75	P1.4.1	Create the IAM role and policy for the Lambda function with the correct naming convention.	
1,25	P1.4.2	Create the Lambda function using the zipped code (bucket zip-code-bucket-xxx/code-02) with the created IAM role. Look at the code for clues.	

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

0,5	Q1.4.2	What is the purpose of the regex in the lambda_handler function? Would the code work without this line?	
0,5	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.	<ul style="list-style-type: none"> • ABC-123-XYZ-2024-04-04_12-30_2024-04-04_14-30.txt • ABC-123-XYZ-2024-04-04_12-30_to_2024-04-04_14-30.txt • file-5678-2023-12-31_23-59_to_2024-01-01_00-30.txt • file_5678-2023-12-31_23-59_to_2024-01-01_00-30.txt
0,5	Q1.4.4	Which of the following best describes the main operations performed by the lambda_handler function?	<p>A. It lists files from an S3 bucket, filters them by a specific filename pattern using regex to extract an identifier, reads their content, saves data to different DynamoDB tables based on the identifier, and deletes the files after processing.</p> <p>B. It listens for SNS messages, parses each message, and saves the data directly into a single DynamoDB table without validation or cleanup.</p> <p>C. It scans S3 objects, uploads new files to the bucket if the filename matches a pattern, then triggers another Lambda function to process those files asynchronously.</p>

			D. It reads all files in an S3 bucket and immediately deletes any files that don't have a valid JSON structure, without further processing.
0,5	P1.4.3	Complete the delete_file_from_s3 function. Hint: only one line needs to be added.	

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.

1,25	P1.4.4	Create the IAM role and the policy for the scheduler.	
0,25	Q1.4.6	What is the correct CRON expression for scheduling every 15 minutes from Monday to Friday between 08:00 and 20:00?	
2	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. Verify that the lambda is authorized to be invoked.	

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>2h05 (2h20 with the break)</i>
<i>Grading</i>	<i>The grading for this first part is out of 15 points.</i>
<i>Goals of this sub-part</i>	<p><i>Collect and process feedback data directly from attendees' tablets in real time.</i></p> <p><i>Write feedback data directly to the DynamoDB tables</i></p> <p><i>Use serverless components to handle the ingestion, processing, and reporting pipelines.</i></p> <p><i>Develop the Python logic for processing feedback and preparing the data for analysis.</i></p> <p><i>Offer a data visualization solution</i></p>
<i>Context</i>	<p><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></p> <p><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></p> <p><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></p> <p><i>You'll be introduced to new concepts and tools, such as capturing data from multiple sources simultaneously, processing it with low latency. 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></p>

Breakdown of the session

- 2.1 Create the Kinesis stream (0h30)
 - Submission of responses at the end of the allocated time.
- 2.2 Deploy the lambda function (1h15)
 - Data will start being sent 30 minutes after the beginning to allow storage to begin.
 - Submission of responses at the end of the allocated time.

BREAK (15min)

- 2.4 Final Assessment Questionnaire (20 minutes)
 - Submission of responses at the end of the allocated time.

A 20- to 30-minute review of the day will then be conducted.

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
	/4 (2,5 + 1,5)	

<i>Do I keep this document</i>	<i>This document will be collected before the end of the sub-part allocated time, 12min over 20min. So, after 20 minutes or earlier if the competitor has finished before we will collect this document. A second document will then be given in exchange. This second document can be kept.</i>
<i>Allocated time</i>	<i>0h20</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>

0,5	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>
0,5	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</p>

			<p>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>
0,5	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>
0,5	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>
0,5	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>

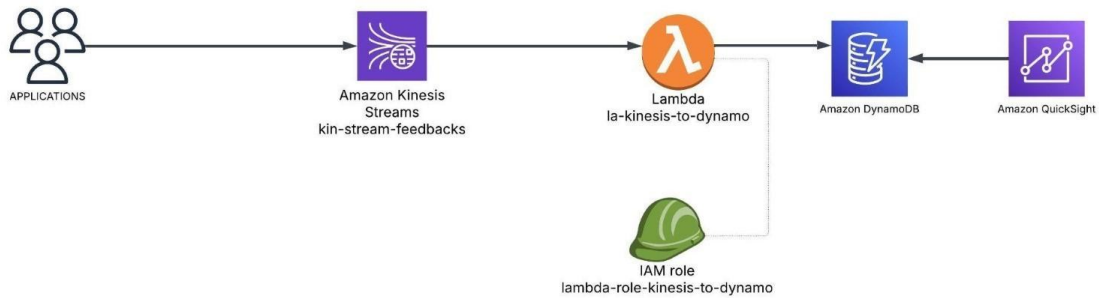
SUB-PART 1 CONTINUED : CREATE THE KINESIS STREAM

Time of submission	Grade	Comment
	/4 (2,5 + 1,5)	

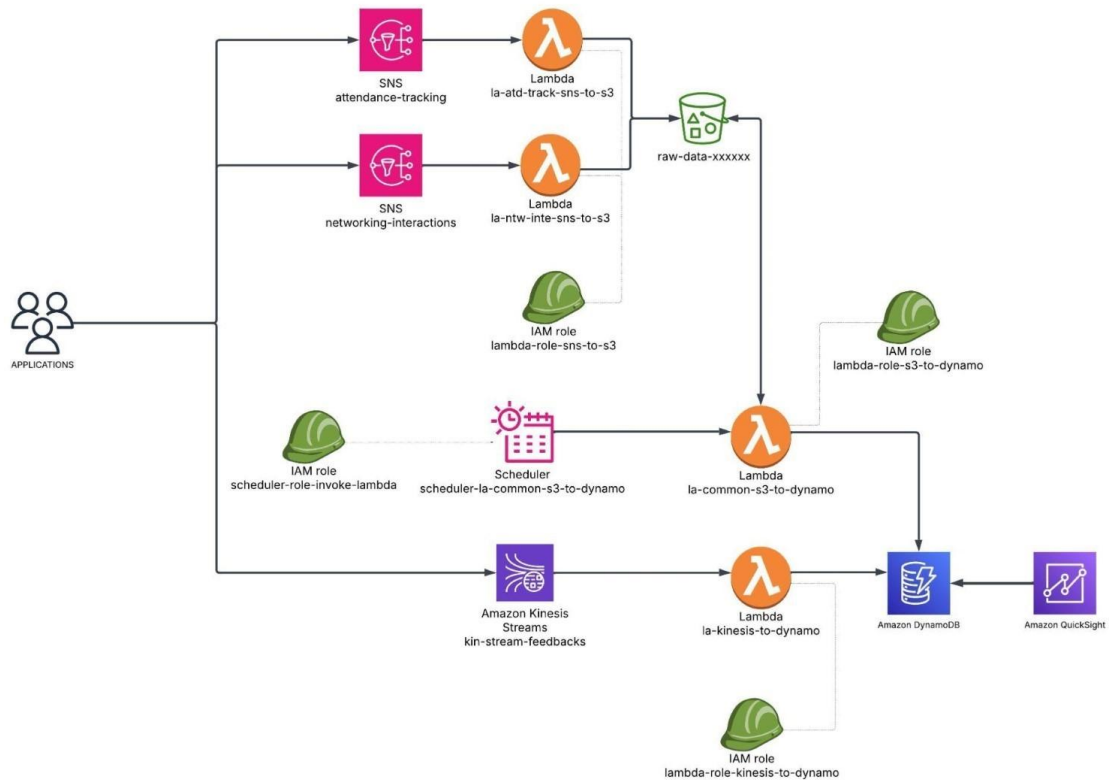
<i>Do I keep this document</i>	<i>This document can be kept throughout the competition.</i>
<i>Allocated time</i>	<i>minimum 10min or a little more if the competitor has handed in his question sheet early</i>
<i>Goals of this sub-part</i>	<i>In this part, you will create a Kinesis stream</i>

1,5	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.
-----	--------	--

Given at the same time as C2 Part 2 sub-part 2 - this document can be kept once given



DIAG 3 : final architecture of part 2



DIAG 4 : final architecture of part 1 and 2

SUB-PART 2: DEPLOY THE LAMBDA FUNCTION

Time of submission	Grade	Comment
	/7,5	

<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>1h15</i>
<i>Goals of this sub-part</i>	<p><i>The goal of this section is to put you in the role of a Cloud Engineer specializing in cloud-native solutions. By completing this exercise, you will strengthen your understanding of how to leverage the cloud for real-time data processing while applying fundamental coding skills.</i></p> <p><i>Note: The feedback data will be made available 25 minutes after the start of this section.</i></p>
<i>Context</i>	<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>

0,75	Q2.2.1	What roles does your Lambda function need? Justify your answer and keep in mind every action it takes.	
1,6	P2.2.1	Create the IAM role and policy for the Lambda function with the correct naming convention.	
1	P2.2.2	Create the Lambda function using the zipped code (bucket zip-code-bucket-xxx/code-03) and the created IAM role	
0,5	Q2.2.2	How did you get Lambda to retrieve the data from the stream? Please explain.	

0,5	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"> - In the "Code" tab, review the provided code. Why does the zipped folder contain multiple subdirectories such as regex and nltk? - Description of the NLTK library 	

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

0,9	P2.2.3	<p>In the function, replace TO_REPLACE with the correct values. Each value is used only once, and two are not used at all :</p> <ul style="list-style-type: none"> • NEUTRAL • SentimentIntensityAnalyzer() • allcap_differential(words) • NEGATIVE • sid.polarity_scores(comment) • make_lex_dict() • POSITIVE 	
0,5	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>
1,75	P2.2.4	<p>In the lambda_handler function, based on the payload processing of fdb-conf, replace the TO_DO section with the correct code to send fdb-food data to DynamoDB.</p> <p>Informations about <i>fdb-conf</i>:</p> <p>The data lines follow this format:</p> <pre>internal_id;user_id;timestamp;conf_id;conf_grade;comment</pre> <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 <i>fdb-food</i>.</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: FINAL ASSESSMENT QUESTIONNAIRE

Time of submission	Grade	Comment
	/3,5	

<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>0h20</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>

0,25	Q2.4.1	CRON Expression 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
0,25	Q2.4.2	Choosing the Right Database DynamoDB or RDS? Use case: You are managing a real-time sales tracking application (eg. Amazon) where every sale needs to be recorded with a	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

		<p>precise timestamp, product ID, and quantity. Users will mainly query sales summaries by product, rather than individual sales details.</p> <p>Which database would you choose for this application? Choose the better justification.</p>	<p>queries on large datasets and handles partitioning automatically.</p> <p>d) RDS, because sales data involves complex relationships.</p>
0,25	Q2.4.3	<p>Choosing Keys in DynamoDB</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"> – Retrieve all tasks for a specific user. – Retrieve all tasks within a project. <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>
0,25	Q2.4.4	<p>Choosing DynamoDB vs RDS for a Solution</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 complex joins, multi-row ACID (wrong, it doesn't)</p> <p>b) RDS, as it can handle complex transactions and relational data.</p> <p>c) RDS, because it supports high-volume reads and writes.</p>

0,25	Q2.4.5	<p>When to Use S3 for Data Storage</p> <p>Use Case: 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. Why would Amazon S3 be a good choice for storing these files before processing?</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 and ensures high availability of files for processing.</p>
0,25	Q2.4.6	<p>Understanding SNS</p> <p>Use Case: 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"> – Log the transaction for auditing purposes. – Notify the user via email that the transaction has been processed. – Trigger a fraud detection process to evaluate whether the transaction is suspicious. – Notify the customer support team about high-value transactions for further review. <p>Which SNS setup would you use to distribute the transaction</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>

		alert to all relevant services?	
0,25	Q2.4.7	Understanding Kinesis Streams What is the primary purpose of a Kinesis stream?	A) To perform complex queries on static datasets B) To store data indefinitely for large-scale storage solutions C) To enable real-time data streaming and processing with high throughput D) To deliver notifications to multiple endpoints in a pub/sub pattern
0,25	Q2.4.8	SNS vs Kinesis Stream Use Case: 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 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. Which service would you choose and why?	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. B) Kinesis Streams is better because it allows you to stream user events in real-time, process the data continuously, and feed the data into analytics platforms for aggregation and reporting. C) SNS should be used because it provides low latency and can handle a high volume of user actions without requiring special setup. (SNS is more suited for notifications to multiple subscribers, not for processing and analyzing a continuous stream of events.) D) Kinesis Streams is better for batch processing user events and should not be used for real-time event tracking.
0,25	Q2.4.9	Lambda function roles When deploying a Lambda function that processes real-time feedback from a Kinesis stream and writes the output to S3 instead of DynamoDB,	A) Read access to Kinesis, Write access to S3, and Log access to CloudWatch B) Full access to S3 and SNS C) Write access to Kinesis, Read access to

		which IAM policy should be assigned to the Lambda function?	<p>DynamoDB</p> <p>D) Full access to Lambda, Kinesis, and DynamoDB</p>
0,5	Q2.4.10	<p>Lambda data processing</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>
0,5	Q2.4.11	<p>Using Amazon Comprehend for Sentiment Analysis</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>
0,25	Q2.4.12	The organization wants to analyze visitor feedback but isn't sure which AWS service can help. The data is stored in DynamoDB and needs to be	<p>a) Amazon Athena</p> <p>b) Amazon QuickSight</p> <p>c) Amazon CloudWatch</p> <p>d) AWS Glue</p>

		visualized in interactive dashboards to track trends by ride and by time of day. Which service will you use ?	
--	--	---	--