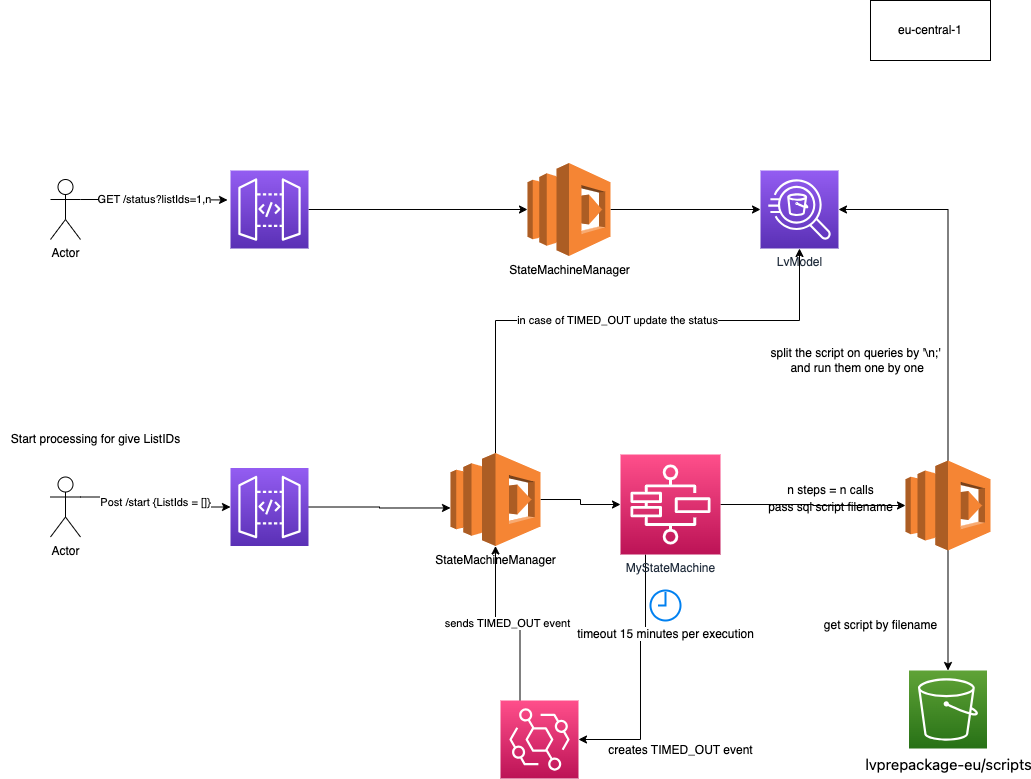
# Solution design: LV Prepackage (Athena Data Processing)

## Problem statement:

Need to execute SQL queries one by one until the whole processing is finished. If the entire processing takes more than 15 minutes, need to abort it and update the status accordingly. An API should allow the client to get the status for the given ListID(s) and start processing.

### High level design



## Api Requests

* Get status for listIds:
  + **PROD**
    - https://tc5qrc9qx7.execute-api.eu-central-1.amazonaws.com/default/status?listIds=468814,467473,1
    - Response {"467473": "1", "468814": "1"}
  + **STAGE**
    - <https://bgbicwt19h.execute-api.us-east-1.amazonaws.com/default/status?listIds=16417,16421,1>
  + DEV
    - https://yzq9s6wsrh.execute-api.us-east-1.amazonaws.com/default/status?listIds=468814,467473,1
* Start execution for listIDs:
  + **PROD**
    - <https://tc5qrc9qx7.execute-api.eu-central-1.amazonaws.com/default/start>
      * Request body : {"listIds": [1,2,3]}
    - Response {"16417": "arn:aws:states:us-east-1:237702658854:execution:MyStateMachine:Execution\_16417\_2023-07-07T10\_49"}
  + **STAGE**
    - <https://bgbicwt19h.execute-api.us-east-1.amazonaws.com/default/start>
      * Request body : {"listIds": [1,2,3]}
    - Response: {"16417": "arn:aws:states:us-east-1:237702658854:execution:StagePrepackageModelByListId:Execution\_16417\_2023-07-07T10\_49"}
  + **DEV**
    - https://yzq9s6wsrh.execute-api.us-east-1.amazonaws.com/default/start
      * Request body : {"listIds": [1,2,3]}
    - Response: {"16417": "arn:aws:states:us-east-1:237702658854:execution:DevPrepackageModelByListId:Execution\_16417\_2023-07-07T10\_49"}

## Implementation details

StateMachine on each step calls runSQLquery lambda and passes the payload consisting of the following params:

* key - sql script path in s3 bucket including filename
* bucket - s3 bucket
* MyTaskToken - for callback

**Each script can contain multiple sql queries, but they should be separated by a new line symbol plus a semicolon:**

*Select 1 from table*

*;*

If state machine PrepackageModelAthenaFromHistory is running, the pipeline will not be executed

## ARN

### PROD

1. Main lambda that serves as a backend for API and StepFunctions event handling  
   *arn:aws:lambda:eu-central-1:237702658854:function:stateMachineManager*
2. Lambda to execute sql scripts *arn:aws:lambda:eu-central-1:237702658854:function:runSQLquery*
3. Bucket with scripts and athena metadata *arn:aws:s3:::lvprepackage-eu*
4. API Gateway dashboard <https://eu-central-1.console.aws.amazon.com/apigateway/main/api-detail?api=tc5qrc9qx7&region=eu-central-1>
5. State machine arn:aws:states:eu-central-1:237702658854:stateMachine:MyStateMachine

### STAGE

1. arn:aws:lambda:us-east-1:237702658854:function:stageStateMachineManager
2. arn:aws:lambda:us-east-1:237702658854:function:stageRunSQLquery
3. Arn:aws:s3:::lvprepackage
4. <https://us-east-1.console.aws.amazon.com/apigateway/main/api-detail?api=bgbicwt19h&region=us-east-1>
5. arn:aws:states:us-east-1:237702658854:stateMachine:StagePrepackageModelByListId

### DEV

1. arn:aws:lambda:us-east-1:237702658854:function:devStateMachineManager
2. arn:aws:lambda:us-east-1:237702658854:function:devRunSQLquery
3. arn:aws:s3:::lvprepackage/dev/
4. <https://us-east-1.console.aws.amazon.com/apigateway/main/api-detail?api=yzq9s6wsrh&region=us-east-1>
5. arn:aws:states:us-east-1:237702658854:stateMachine:DevPrepackageModelByListId

How to connect to Athena from EC2

1. Attach IAM role with required permissions to EC2 (Should be done by DevOps / Denis)

2. Install ODBC driver, <https://docs.aws.amazon.com/athena/latest/ug/connect-with-odbc-driver-and-documentation-download-links.html>   
See attached screenshots:

