* AWS CLI ADD step 명령어

> aws emr add-steps --cluster-id {my-cluster-id}

--steps Type=CUSTOM\_JAR, Name="My SPARK JOB NAME”,

ActionOnFailure=CANCEL\_AND\_WAIT,

Jar="command-runner.jar",

Args=[spark-submit, --deploy-mode,cluster, --deploy-mode,client,--master,yarn,

--py-files, s3://skb-oasis-pgm-workload/lib/oasis\_pyspark\_module.py\,

s3://skb-oasis-pgm-workload/env/env\_common.py\,s3://skb-oasis-pgm-workload/env/env\_dw.py\,

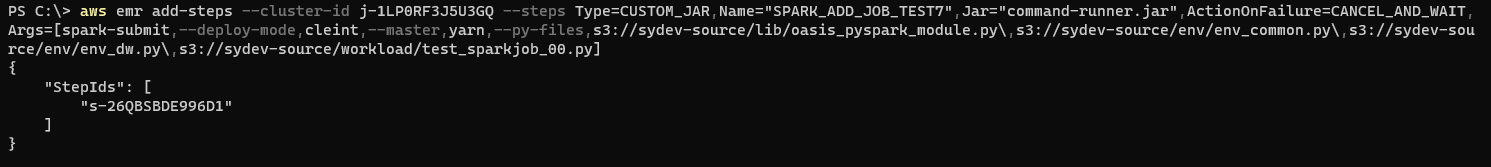
step 에 추가할 파일 , --D1,YYYYMMDD ex:20220318,--D2,YYYYMMDD ex:20220319]

--endpoint-url https://vpce-08c301a70496086d0-mqcikbd2.elasticmapreduce.ap-northeast-2.vpce.amazonaws.com

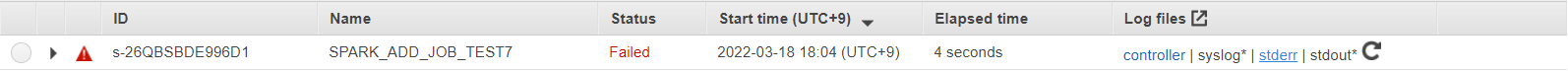
--profile 695473834042\_SKB-DevOps-ForOasis

* { } : 개발자 수정부분
* 가독성을 위해 줄바꿈, 띄어쓰기를 했지만 실제로 등록할 때는 전부 붙어야 합니다.
* Error message

Aws cli로 add step을 했으나,



EMR step 에서 fail로 뜸



로그 :

Error: Missing application resource.

Usage: spark-submit [options] <app jar | python file | R file> [app arguments]

Usage: spark-submit --kill [submission ID] --master [spark://...]

Usage: spark-submit --status [submission ID] --master [spark://...]

Usage: spark-submit run-example [options] example-class [example args]

Options:

--master MASTER\_URL spark://host:port, mesos://host:port, yarn,

k8s://https://host:port, or local (Default: local[\*]).

--deploy-mode DEPLOY\_MODE Whether to launch the driver program locally ("client") or

on one of the worker machines inside the cluster ("cluster")

(Default: client).

--class CLASS\_NAME Your application's main class (for Java / Scala apps).

--name NAME A name of your application.

--jars JARS Comma-separated list of jars to include on the driver

and executor classpaths.

--packages Comma-separated list of maven coordinates of jars to include

on the driver and executor classpaths. Will search the local

maven repo, then maven central and any additional remote

repositories given by --repositories. The format for the

coordinates should be groupId:artifactId:version.

--exclude-packages Comma-separated list of groupId:artifactId, to exclude while

resolving the dependencies provided in --packages to avoid

dependency conflicts.

--repositories Comma-separated list of additional remote repositories to

search for the maven coordinates given with --packages.

--py-files PY\_FILES Comma-separated list of .zip, .egg, or .py files to place

on the PYTHONPATH for Python apps.

--files FILES Comma-separated list of files to be placed in the working

directory of each executor. File paths of these files

in executors can be accessed via SparkFiles.get(fileName).

--conf, -c PROP=VALUE Arbitrary Spark configuration property.

--properties-file FILE Path to a file from which to load extra properties. If not

specified, this will look for conf/spark-defaults.conf.

--driver-memory MEM Memory for driver (e.g. 1000M, 2G) (Default: 1024M).

--driver-java-options Extra Java options to pass to the driver.

--driver-library-path Extra library path entries to pass to the driver.

--driver-class-path Extra class path entries to pass to the driver. Note that

jars added with --jars are automatically included in the

classpath.

--executor-memory MEM Memory per executor (e.g. 1000M, 2G) (Default: 1G).

--proxy-user NAME User to impersonate when submitting the application.

This argument does not work with --principal / --keytab.

--help, -h Show this help message and exit.

--verbose, -v Print additional debug output.

--version, Print the version of current Spark.

Cluster deploy mode only:

--driver-cores NUM Number of cores used by the driver, only in cluster mode

(Default: 1).

Spark standalone or Mesos with cluster deploy mode only:

--supervise If given, restarts the driver on failure.

Spark standalone, Mesos or K8s with cluster deploy mode only:

--kill SUBMISSION\_ID If given, kills the driver specified.

--status SUBMISSION\_ID If given, requests the status of the driver specified.

Spark standalone, Mesos and Kubernetes only:

--total-executor-cores NUM Total cores for all executors.

Spark standalone, YARN and Kubernetes only:

--executor-cores NUM Number of cores used by each executor. (Default: 1 in

YARN and K8S modes, or all available cores on the worker

in standalone mode).

Spark on YARN and Kubernetes only:

--num-executors NUM Number of executors to launch (Default: 2).

If dynamic allocation is enabled, the initial number of

executors will be at least NUM.

--principal PRINCIPAL Principal to be used to login to KDC.

--keytab KEYTAB The full path to the file that contains the keytab for the

principal specified above.

Spark on YARN only:

--queue QUEUE\_NAME The YARN queue to submit to (Default: "default").

--archives ARCHIVES Comma separated list of archives to be extracted into the

working directory of each executor.

Command exiting with ret '1'