**Overview of Where Logs Can Be Captured**

| **Step** | **Component** | **Where Logs Go** | **How to Access** |
| --- | --- | --- | --- |
| 1️⃣ | **Lambda** | CloudWatch Logs | /aws/lambda/<lambda\_name> |
| 2️⃣ | **Airflow (MWAA)** | CloudWatch Logs under MWAA | /aws/airflow/<your\_mwaa\_env> |
| 3️⃣ | **Spark on EKS** | CloudWatch, S3, or kubectl logs | Depends on how Spark logging is configured |
| 4️⃣ | **S3 Output (Optional)** | S3 bucket (if configured) | <your-s3-bucket>/spark-logs/... |

When your Spark job runs on EKS, you want all Spark logs (events, metrics, stages, tasks) to be stored in an S3 bucket, so they’re:

* Persisted beyond pod lifecycle
* Viewable later via the Spark History UI
* Useful for debugging and audit

**How Spark Logging Works**

Spark has a built-in event logging system.  
When you set the following configs, Spark writes event logs (JSON files) to your chosen directory — in this case, an S3 bucket.

--conf spark.eventLog.enabled=true

--conf spark.eventLog.dir=s3a://<your-log-bucket>/spark-events/

These logs record:

* Application ID
* Stage start / completion
* Task metrics (input/output, shuffle, CPU time, etc.)
* Executor events
* Job success/failure info

**Full Example — spark-submit Command (EKS)**

Here’s how you’d modify your Spark command in Airflow or shell:

spark-submit \

--master k8s://https://<EKS\_API\_ENDPOINT> \

--deploy-mode cluster \

--name spark-s3-logger-job \

--conf spark.kubernetes.namespace=default \

--conf spark.kubernetes.container.image=<your-spark-image> \

--conf spark.kubernetes.authenticate.driver.serviceAccountName=spark-sa \

--conf spark.eventLog.enabled=true \

--conf spark.eventLog.dir=s3a://my-spark-logs-bucket/spark-events/ \

--conf spark.hadoop.fs.s3a.aws.credentials.provider=com.amazonaws.auth.WebIdentityTokenCredentialsProvider \

--conf spark.hadoop.fs.s3a.impl=org.apache.hadoop.fs.s3a.S3AFileSystem \

local:///opt/spark/jars/SparkS3CSVFilter-assembly-0.1.jar

**Key Notes**

* spark.eventLog.enabled=true → turns on event logging.
* spark.eventLog.dir=s3a://... → where Spark writes JSON event files.
* spark.hadoop.fs.s3a.aws.credentials.provider → required if you’re using IRSA for S3 access from EKS.
* These logs are automatically uploaded as the job runs.