

When was the last time you tried/learned something new? What did you do and how?

Since I chose big data engineering as my career path, I never stopped learning about the latest technologies related to big data, trying to apply that on my job tasks so I can get actual and on-hands experience and to grow within the corporate or startup I'm working for.

Six months ago, I had the chance to learn about Hadoop by buying different video courses from Udemy and useful ebooks from Amazon. I could learn more by applying what I learned about Spark and Hive on my current tasks and introduce the data team to such tech and even mentor them how to use and get experienced using them.

Hence, the current ETL workflow I'm working on is being migrated from using web services using Node JS to Hadoop environment, using AWS EMR, Athena and later Glue or Airflow for ETL orchestration.

What is a Lambda architecture? Which problems does it solve?

AWS Lambda is a serverless service which allows you to run your code with zero-administration of the servers your code needs to run on.

AWS Lambda provides various solution use cases which helps data engineers build their data pipelines faster, better and burden-free:

- Stream Processing: Using Kinesis streams, Lambda helps reading Kinesis shards and writes them to a data storage such as S3 and DynamoDB.
- Real-time File Processing: Whenever a new file is being uploaded to one of our S3 buckets, Lambda can be triggered to process the file using the code uploaded by an engineer. Processing can be compressing the file, or creating a thumbnail of the photo uploaded, or document validation.
- ETL: Data cleansing, validation, transformation and filtering can be held by Lambda which can later be loaded into S3, DynamoDB, RDS or Redshift.

What to consider when using Lambda:

- Lambda execution time is 300 seconds (5 minutes).
- Memory default size is 512MB, can be expanded to 3,000MB.

So, Lambda is finite which can be a constraint if the code is not optimized to finish within the timeout and memory restrictions. Lambda also support some popular languages which may be a drawback.