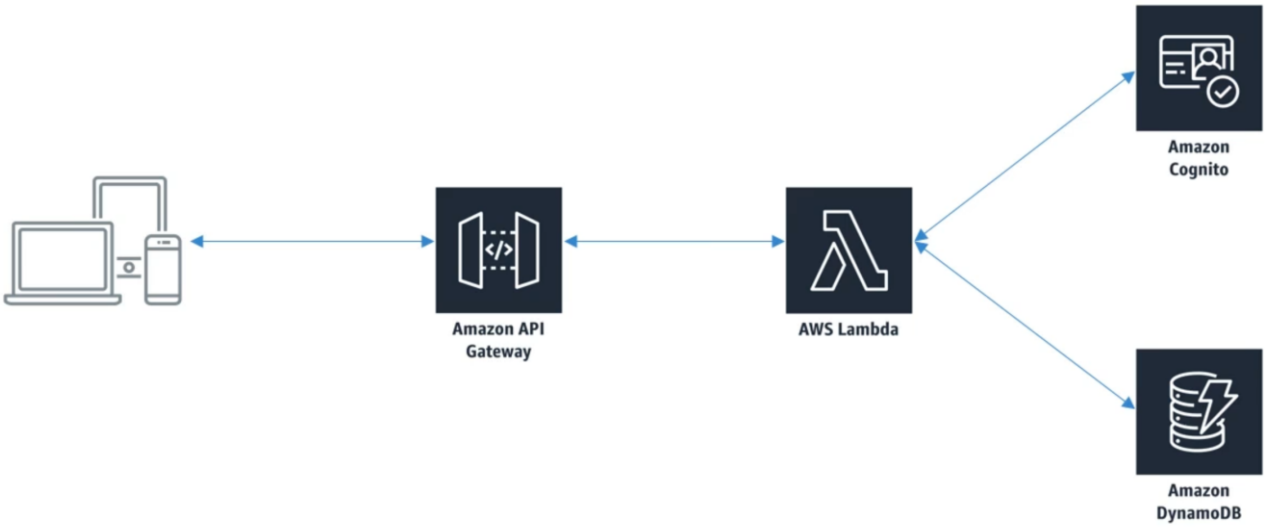
**AWS Lambda –** A serverless data processing tool. It helps us run snippets of code in the cloud. Will scale out the hardware as needed. Used to glue different services together in Big Data.

Has high availability (no scheduled down time), unlimited scalability , high performance.

Eg -



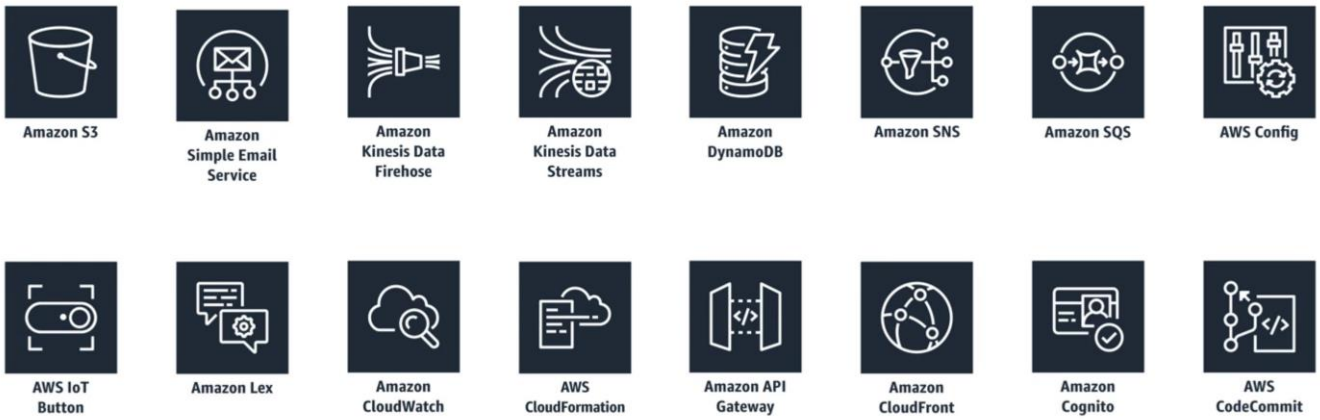

The reasons for which this is better alternative than running server are –

- Servers require constant management.
- Servers can be cheap but scaling gets quite expensive
- You don’t pay for the processing time you don’t use
- Easier to split up development between the front end and back end

Real world uses of Lambda –

- Real-time file processing
- Real-time stream processing
- ETL
- Cron replacement
- Process AWS events

# Lambda triggers




Lambda and Pipelines – can be useful to build pipelines. It can trigger at certain times but if you don't know when the data might come, you can use a lambda trigger to run the pipeline when the data is received.

Since Lambda is stateless, it can't keep track of the work that has been done already. We can use DynamoDB to to keep track of what's been loaded.

If the batch size of streaming data is too big, it might cause Lambda to crash.

Anti-Patterns :

- If the application takes more than 900 seconds then an EC2 instance is more preferrable to use
- Not suitable for stateful applications