Computer Science & Engineering, University of Nevada Reno

Using AWS Lambda to Solve Big Data Problems

Authors
Brian Smith
Max Wiegant
Vladislav Savranschi

Instructor
Dr. Feng Yan

April 14, 2017

Table of Contents

Background and Related Work

AWS Lambda is an event-driven, serverless computing platform provided by Amazon as a part of the Amazon Web Services. You write a snippet of code that you want to run, then that code is executed every time a particular event is triggered. For instance, let's say you add an image to your S3 bucket, you can make a Lambda function detect the event and then run some code to convert that image into another format on the fly. Lambda only runs the specific job, then kills itself, which means if you upload 100 photos at once, 100 Lambda jobs will be generated and they will all execute in parallel then kill themselves when they have completed their task.

Objective

The purpose of this project is to show several use cases and advantages of using AWS Lambda to solve big data problems. To show the capabilities of using Lambda, we will put together several examples that demonstrate how Lambda can tackle big data computing, data retrieval, and data analysis problems. Video conversion and image conversion is a great way to show how Lambda can solve big data computing problems. To show this, we could put several thousand images into an S3 bucket and then use Lambda to spin up several hundred instances to convert or modify them in parallel. To demonstrate big data retrieval and analysis, we could create a small application that grabs Airbnb data and cross references it with Zillow data to come up with the best listings. Instead of just getting information about a single area, Lambda can target many places at once and get a big data result all at once.

Note: the examples provided here may change slightly as we move forward in our implementation.

Challenges and motivation

Challenges

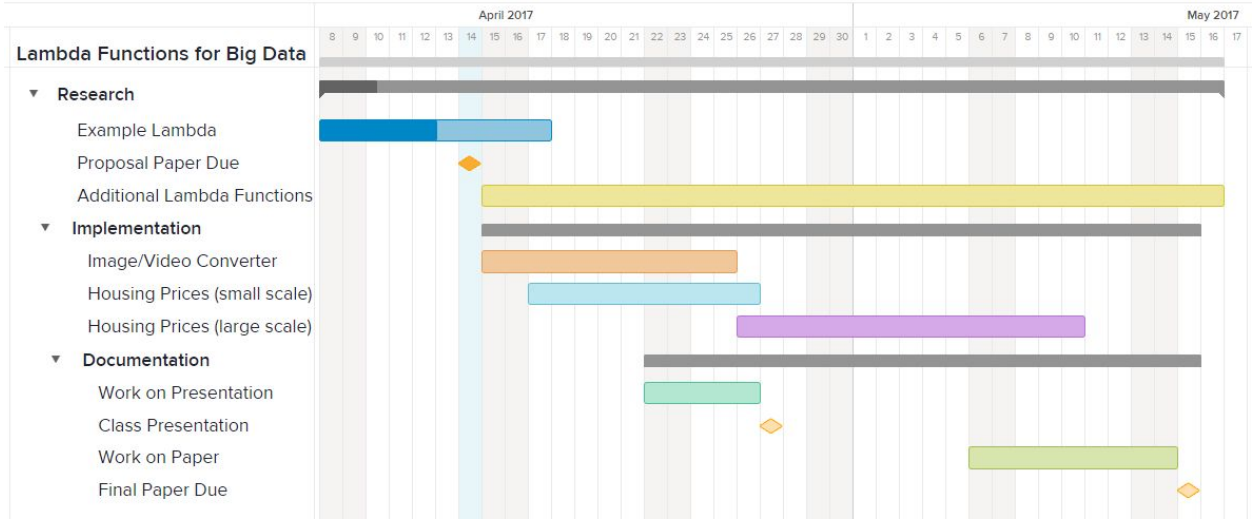
- Learning how to use AWS Lambda
- Figuring out how to create test examples
- Testing and analysing various examples

Motivation

- To learn and create big data solutions to complex problems
- Maximising speed and efficiency with large data sets
- Getting hands on experience with AWS Lambda

Timeline and Milestones

Project Gantt Chart



Explanation

The first goal beyond the proposal is to complete a working example function, which will help with learning how to use AWS Lambda for more complicated functions. Next, set up a simple image or video converter by executing commands in an open source Javascript library. For “Housing Prices” implementations, we will utilize Zillow and/or Airbnb API’s to pull relevant data on a small scale mesh of overlapping radii, run it through Lambda, and create various statistics. After a successful small scale implementation, the scale will be vastly increased to demonstrate the effectiveness of Lambda in Big Data usage. We will keep researching for more examples of utilizing Lambda for Big Data, whilst working on our presentation and final paper - due April 26th and May 15, respectively.

Tools

AWS Lambda

AWS Lambda is an event-driven, serverless computing platform provided by Amazon as a part of the Amazon Web Services. It is a compute service that runs code in response to events and automatically manages the compute resources required by that code. It was introduced in 2014.

The purpose of Lambda, as compared to AWS EC2, is to simplify building smaller, on-demand applications that are responsive to events and new information.

Node.js

Node.js is an open-source, cross-platform JavaScript runtime environment for executing JavaScript code server-side. Node.js enables JavaScript to be used for server-side scripting.

Responsibilities

Lambda Big Data: Computing - Brian Smith

Description: Computes large tasks, converting something from one form to another.

Lambda Big Data: Retrieval - Max Wiegant

Description: Retrieves API information from many different websites.

Lambda Big Data: Analysis - Vladislav Savranschi

Description: Analyses gathered big data sets and interprets it into useable statistical data (price per time, price per square footage of listing)