

**GEORGIA INSTITUTE OF TECHNOLOGY**  
**SCHOOL of ELECTRICAL and COMPUTER ENGINEERING**

By Michelle Feng and Elisa Miller

**ECE 4150 Spring 2024**

**Final Project: Big Data Analytics with Apache Spark**

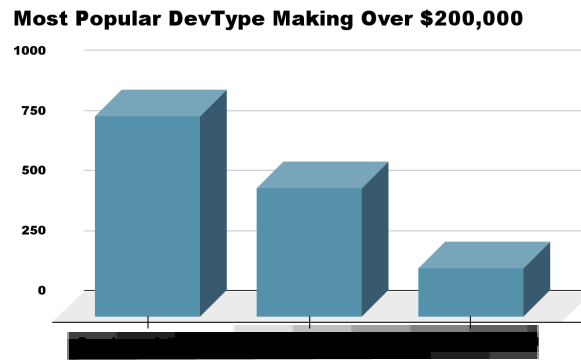
1. Create and configure EMR cluster for Spark
2. Create and configure S3 Bucket file system
3. Write data processing code in python
4. Create and run Spark code in the Hadoop terminal
5. View results in S3 bucket
6. Create a visual representation of the data to display your results

## Data Source

Download your data source from [here](#)

## Challenges

1. Process the data to only include those who live in the USA and have been learning to code for 15+ years
2. Process the data to only include those who are a developer by profession and make \$150,000+ USD
3. Process the data to only include those who make \$200,000+ USD
4. Filter that data to determine the top three DevTypes that make \$200,000+
5. Create a visual representation of the top 3 DevTypes that make \$200,000+ to understand what DevType most commonly makes the most money!



**Figure 1.** Results may vary but should look like this