

# John Doe

john.doe@duke.edu | +1 xxx-xxx-xxxx | <https://github.com/JDoe>

## Education

Duke University, Durham, NC May 2019  
Master of Science in Computer Science  
GPA: 3.7

Great University, Granville, OH May 2017  
Bachelor of Science in Computer Science, Double Major in Mathematics, Minor in Economics  
Major GPA: 3.7/Overall: 3.6

## Programming Skills

**Programming Languages:** Python, C/C++, Java, R, SQL, Scala, Spark

**Skills:** AWS, Hadoop, MongoDB, Django, Arduino, Stata, Mathematica, HTML, LaTeX

## Projects

Klaviyo Weather Powered Email, Durham, NC, Full Stack Project Jan 2017

- Designed a full stack web application in **Django**, taking the sign-up forms from users to send personalized emails based on the current weather at the subscribers' location
- Wrote front-end in **HTML** with **CSS** styling and the back-end stores the subscribers' information in **MySQL**

Perceptron Clustering on Wikipedia Data, Granville, OH, Senior Project Mar - May 2017

- Designed a **MPI** program utilizing 20 **Linux** machines to download 5 Terabytes data in an efficient manner
- Wrote two **MapReduce** jobs in Java to clean the data in **HDFS**
- Wrote **bash** scripts to handle massive data transportation from local machines to **HDFS**.
- Incorporated two parallel computing libraries, **MPI** and **Pthreads** to achieve the parallel querying on Wikipedia query API website, and achieved a dramatic speedup
- Designed a data analysis program with **SQL** database in serialized **Python** version (same result was also achieved in parallel **MapReduce** version after assigning the database into a distributed file system)
- Demonstrated the procedure and the mechanism behind such multi-parallel hybrid system in paper
- Presented and addressed the result of the analysis, the limitations of existing hardware, and the potential future problems

## Related Experience

**Software Engineer & Research Scholar** May - July 2016

*Data Streaming Algorithms for the Chi-Square Test, Granville, OH*

- Designed 4 streaming algorithms for 3 variants of the Chi-square Test with minimum assumption of the stream
- Implemented the algorithms with experiment evaluations and basic streaming fashion statistics, open-sourced on GitHub
- Validated the performance of results through extensive testing on both real and synthetic data sets on a large-scale (stream size  $n \approx 10^7$ )
- Submitted a ten-page paper to IEEE 2017 conference and presented research findings at the Anderson Program Science Symposium

## Honors & Awards

- 1st Place, Regional Programming Contest, Ohio Wesleyan University Nov 2017
- 1st Place, Ohio Four College Math Contest, Wittenberg University Mar 2017
- Upsilon Pi Epsilon, Computer Science Honor Society Apr 2016
- Pi Mu Epsilon, Mathematics Honor Society Apr 2016
- Mortar Board National Senior Honor Society Apr 2016