

Robert Thome

1779 Blue Course Drive, State College, PA 16801
robert.thome@protonmail.com
(215) 518 8022

Objective Software engineer with expertise in data engineering and analytics, seeking work with Blue Orange Digital to utilize experience in designing and developing scalable, maintainable and fault-tolerant data pipelines and models for data intensive systems.

Education **The Pennsylvania State University**, State College, PA
Master of Professional Studies, Data Analytics, 2020

The Pennsylvania State University, State College, PA
Bachelor of Science, Electrical Engineering, 2012 – 2016
3.87 / 4.00 GPA

Professional Experience **Applied Research Laboratory, the Pennsylvania State University**, State College, PA
R&D Engineer, 2016 – 2019

Project One: *application to identify trends and anomalies in and across cyber and cryptocurrency data sets*

- Designed and implemented ETL pipelines to process historical and real-time data on terabyte scale
- Responsible for the configuration, management and tuning of on-premise computing cluster
- Managed dataflow and workflow processes across twenty-plus node network
- Responsible for data science intern tasking and mentoring
- Utilized Java, Scala, Python, HBase, Phoenix, Hadoop (HDFS, MapReduce) Spark, Kafka, NiFi

Project Two: *application to train NLP models to perform name entity recognition*

- Implemented ETL pipeline to process message traffic for natural language processing tasks
- Developed workflow to perform name entity recognition using trained models and regex rules
- Responsible for intern tasking and coordination for labeling training data for model development
- Utilized Java, Postgres

Project Three: *web application to provide users tools to investigate and visualize complex data sets*

- Responsible for backend development of Java servlets, SQL queries, and Tomcat server configuration
- Implemented utilities to automate user queries, report generation, and alerting
- Implemented web security filters and configurations to secure application
- Wrote client interfaces to third-party REST and RPC API services
- Utilized Java, Postgres, Hadoop (HDFS)

| | | |
|--------------------|------------------------|---------------------------------|
| Skills | Programming Languages: | C/C++, Python, Java, Scala |
| | NoSQL Databases: | HBase, Phoenix |
| | SQL Databases: | Postgres |
| | Cloud Computing: | AWS (EC2, Lambda, S3, IAM, EMR) |
| | Distributed Computing: | Hadoop, Spark |
| | Streaming Platforms: | Kafka |
| | Dataflow Management: | NiFi |
| | Machine Learning: | TensorFlow, Scikit-Learn |
| Build Systems: | Ant, Maven | |
| Operating Systems: | Linux/Unix, Windows | |