

# Li Pei

<http://www.linkedin.com/in/lip17>

Email : [lipei.work@gmail.com](mailto:lipei.work@gmail.com)

Mobile : +1-412-419-4227

## EDUCATION

---

- **Carnegie Mellon University** Pittsburgh, PA  
*Master of Electrical And Computer Engineering, GPA: 3.92/4.00*  
May 2016
- **Harbin Institute of Technology** Harbin, China  
*Bachelor of Science, Communication Engineering, GPA: 87.78/100*  
June 2014

## SKILLS

---

- **Programming Language:** Mainly use of Scala and Java, hands on coding in Bash, project experience in Go and C.
- **Microservice Building:** Working experience in building service with Finagle, Zookeeper, Docker, Mesos, Nginx
- **Big Data ETL:** Experienced in implementing ETL pipeline with Scalding, Spark, Oozie, Sqoop, Hive/Impala, HDFS. Real life work on building data streaming with Kafka.
- **API Implementation:** Design and implement API using Scala in Play Framework and Typed.js in Node.js. Relational DB design in MySQL, Vertica and PostgreSQL
- **Real-time Bidding:** Experienced in DSP build, maintenance and trouble-shooting; Familiar with Open RTB protocol and integration workflow; knowledged about ID-Sync, audience targeting

## EXPERIENCE

---

- **Tapad Inc.** New York, NY  
*Software Engineer*  
July 2016 - Present
  - **Tensorflow:** TensorFlow is an open source software library for numerical computation using data flow graphs; primarily used for training deep learning models.
  - **Apache Beam:** Apache Beam is a unified model for defining both batch and streaming data-parallel processing pipelines, as well as a set of language-specific SDKs for constructing pipelines and runners.
- **EEme, LLC** Pittsburgh, PA  
*Software Engineer, Part-time Intern*  
Mar. 2016
  - **Notifications:** Service for sending email, push and in-app notifications. Involved in features such as delivery time optimization, tracking, queuing and A/B testing. Built an internal app to run batch campaigns for marketing etc.
  - **Nostos:** Bulk data processing and injection service from Hadoop to Cassandra and provides a thin REST layer on top for serving offline computed data online.
  - **Workflows:** Dataduct an open source workflow framework to create and manage data pipelines leveraging reusables patterns to expedite developer productivity.
  - **Data Collection:** Designed the internal survey and crowd sourcing platform which allowed for creating various tasks for crowd sourcing or embedding surveys across the Coursera platform.
  - **Dev Environment:** Analytics environment based on docker and AWS, standardized the python and R dependencies. Wrote the core libraries that are shared by all data scientists.
  - **Data Warehousing:** Setup, schema design and management of Amazon Redshift. Built an internal app for access to the data using a web interface. Dataduct integration for daily ETL injection into Redshift.
  - **Recommendations:** Core service for all recommendation systems at Coursera, currently used on the homepage and throughout the content discovery process. Worked on both offline training and online serving.
  - **Content Discovery:** Improved content discovery by building a new onboarding experience on coursera. Using this to personalize the search and browse experience. Also worked on ranking and indexing improvements.
  - **Course Dashboards:** Instructor dashboards and learner surveying tools, which helped instructors run their class better by providing data on Assignments and Learner Activity.
- **Lucena Research** Atlanta, GA  
*Data Scientist*  
Summer 2012 and 2013
  - **Portfolio Management:** Created models for portfolio hedging, portfolio optimization and price forecasting. Also creating a strategy backtesting engine used for simulating and backtesting strategies.
  - **QuantDesk:** Python backend for a web application used by hedge fund managers for portfolio management.
- **Georgia Institute of Technology** Atlanta, GA  
*Research and Teaching Assistant*  
Jan 2012 - Dec 2013

- **Research Assistant - Machine Learning:** Research on machine learning for portfolio hedging and replication algorithms. Modeling low-risk & continuous-return strategies. Developed the python library QSTK.
- **Teaching Assistant - Computational Investing:** The online course on Coursera, had more than 100,000 students enrolled. It was featured on the 11 Alive News and the Atlanta Journal Constitution. Involved in creating assignment, exams and conducting recitation sessions. Also taught the on-campus version of the course.

## PROJECTS

---

- **QuantSoftware Toolkit:** Open source python library for financial data analysis and machine learning for finance.
- **Github Visualization:** Data Visualization of Git Log data using D3 to analyze project trends over time.
- **Recommendation System:** Music and Movie recommender systems using collaborative filtering on public datasets.
- **Mac Setup:** Book that gives step by step instructions on setting up developer environment on Mac OS.