

Daniel Pascal Lamblin

Brooklyn NY U.S.A. & Remote

m:dlamblin@gmail.com

t:+1 646 644 3587

gh:dlamblin.github.io

in:linkedin.com/in/daniellamblin

Skills *JVM*: Java, Scala, Guava, Guice, Dagger, SpringBoot, Lucene, GWT & Kotlin; *Unix*: Python, Go, C, Perl, Bash & Zsh; *C#*: ASP.Net; *Data*: Spark, Airflow, MySQL, BigQuery, Hive, Presto, Megastore, Beam/Cloud Dataflow; *Web*: NodeJS, Vue.js, HTML, CSS & JavaScript; *AWS*: ALB, ASG, EBS, EC2, EFS, EKS, EMR, Lambda, RDS, CloudTrail, CloudWatch, CloudFormation, S3 ...; Agile & TDD

Education Worcester Polytechnic Institute; Worcester, MA – BS Computer Science Sep 1995–May 1999
Columbia University; New York, NY – Non-Degree Semester Sep 2007–Dec 2007
Andrew Ng's Stanford ML-Class.org; Online – Became Coursera Oct 2011–Dec 2011

Experience **Etsy**, *Senior Software Engineer* Experimentation Platform Brooklyn Jul 2022–Present

- (Selected) ▶ Extended self-service no-code custom experiment metrics for teams to track funnel metrics in experiment variants
- ▶ Added some api endpoints in php to support custom metrics and events management from commandline and python notebooks
- ▶ Consolidated jobs with similar sources to get 45% run time improvements & increased argument granularity to allow back-fills without customizing jobs
- ▶ Added data quality validations to new and old experimentation pipeline tables with Great Expectations
- ▶ Developed Self-repair pipeline steps for event mis-namings & extend monitoring with on-call rotations

Coupang, *Senior Backend Software Engineer* Clickstream Web Logging Seattle Nov 2020–Apr 2022

- ▶ Developed and delivered Customer Experience Analytics Platform tool for self-service funnel, journey and trend analysis of web log data using Vue.js 2, Spark, and Scala
- ▶ Mentored intern project to port and extend UI on Vue.js 3 & hired them
- ▶ Assisted Retail Delivery team's mission-critical multi-AZ expansion projects

Senior Data Engineer Data Platform Tools and Infra Seattle Jul 2019–Nov 2020

- ▶ Migrated ~4000 Airflow ETL jobs from Airflow 1.8 to 1.10+, with custom tooling
- ▶ Optimized teams' resource usage of EMR and Spark jobs
- ▶ Split and upgraded Airflow into deployments by priority

Big Data Engineer Big Data Platform Seoul Aug 2017–Jul 2019

- ▶ Onboard teams onto Airflow & and maintained EMR clusters for their jobs
- ▶ Developed Data Platform Portal tool with Cluster management and data discovery features
- ▶ Scale out Airflow to ~6000 dags with multiple deployments
- ▶ On call rotation for Airflow, Presto, Hive, Hue, Zeppelin, Spark, HDFS, Zookeeper etc.

Data Engineer Data Platform Seattle Jul 2016–Aug 2017

- ▶ Migrated on-prem Netezza and HDFS Hive to EMR Hive on S3
- ▶ Update ETL into Airflow from Oozie and Talend
- ▶ Monitored data readiness with on-call rotations

Insight Data Science, *Data Engineering Fellow* New York City Sep 2015–Jan 2016

- ▶ Realtime and batch processing on NYC MTA's GTFS stream with Kafka, Spark, HDFS, HBase, S3
- ▶ Generated user data in large scale for testing aim of notifying users of train delays
- ▶ Cluster on AWS EC2; project information and presentation at dlamblin.github.io/mta-delay-monitoring

Google, *Software Engineer* New York City Jul 2010–Nov 2014

- ▶ Ported legacy Studio product, a rich advertisement authoring and QA web-app, to Google's Web Toolkit front-end with a stubby rpc backend and megastore datastore
- ▶ Developed a dashboard to track component usage data of ads by comparing html5 vs. flash authoring, common formats, layouts, and generated impressions. Utilized cross-team apis and internal versions of GFS, Cloud Dataflow and Drill
- ▶ Implemented critical preview features for monitoring ad unit interactions and compliance
- ▶ Migrated user records and assets to support new multi-account users and unified asset library view
- ▶ Reduced reprocessing and conserved storage of assets by fingerprinting uploads, both on individual files and within archives

Travelocity, *Senior Web Application Developer* New York City Jul 2006–Jan 2009

- ▶ Established features for IgoUgo.com using the ASP.NET 2.0 framework, C#, PrototypeJS
- ▶ Introduced a Lucene based index of content to offload db search as a wsdl service in Java and Spring with auto-completed suggestions for key geo-entities
- ▶ Boosted traffic ten fold through optimization of page structure, URLs and image file names

EMC Corp, *Associate Software Engineer* Milford, MA Jun 2000–Oct 2001

- ▶ Extended EMC Data Manager Volume & Tape Library Manager's media duplication processes targeting Petabyte capable systems like Sony PetaSite with multiple robots and drives
- ▶ Improved and maintained EDM as a multi-process C based system with Sun RPC, threading, and IPC signals & pipes; Resolved deadlock by refactoring mutex hierarchy