

Daniel Pascal Lamblin		m: dlamblin@gmail.com
Brooklyn NY U.S.A. & Remote		t: +1 646 644 3587
		gh: dlamblin.github.io
		in: linkedin.com/in/daniellamblin
Skills	<i>JVM:</i> Java, Guava, Guice, Dagger, SpringBoot, Lucene, GWT, Scala & Kotlin; <i>Unix:</i> Python, Go, C, Perl & Bash; <i>C#:</i> ASP.Net; <i>Data:</i> Airflow, MySQL, Hive, Presto, Spark, Megastore, Beam/Cloud Dataflow; <i>Web:</i> NodeJS, Vue.js, HTML, CSS & JavaScript; <i>AWS:</i> ALB, ASG, EBS, EC2, EFS, EKS, EMR, Lambda, RDS, CloudTrail, CloudWatch, CloudFormation, S3 ...; Agile & TDD	
Education	Andrew Ng's Stanford ML-Class.org; Online – Became Coursera	2011–2011
	Columbia University; New York, NY – Non-Degree Semester	2007–2007
	Worcester Polytechnic Institute; Worcester, MA – BS Computer Science	1995–1999
Experience (Selected)	Coufang , <i>Senior Backend Software Engineer</i> Clickstream Web Logging Seattle 2020–Present	
	▶ Developed Customer Experience Analytics Platform tool for self-service funnel, journey and trend analysis of web logging data ▶ MVP for CXAP tool: Stood up the web front on Vue.js 2 for customers of Business Analytics team ▶ Designed the scheduling of CXAP jobs ▶ Delivered Journey and Trend analysis for backend of CXAP with Spark and Scala ▶ Mentored intern project to port and extend UI on Vue.js 3 & converted to hire ▶ Assisted Retail Delivery team's mission-critical multi-AZ expansion projects	
	<i>Senior Data Engineer</i> Data Platform Tools and Infra Seattle 2019–2020	
	▶ Migrated enterprise data warehouse's ~4000 Airflow ETL jobs from Airflow 1.8 to 1.10+, with custom semi-automated tooling. ▶ Right-size teams' infrastructure resource usage by monitoring for efficient use of EMR and Spark jobs ▶ Split Airflow into multiple deployments and onboard users to latest versions ▶ On call rotation as in Seoul.	
	<i>Big Data Engineer</i> Big Data Platform Seoul 2017–2019	
	▶ Onboard teams (edw, growth, pricing, catalog, search, retail etc.) into airflow; configure and maintain EMR clusters for teams' jobs ▶ Develop Data Platform Portal tool with Cluster start-stop API, management, and Data Discovery ▶ Scale out Airflow to ~6000 dags with multiple deployments ▶ On call rotation for Airflow, Presto, Hive, Hue, Zeppelin, Spark, HDFS, Zookeeper etc.	
	<i>Data Engineer</i> Data Platform Seattle 2016–2017	
	▶ Cloud migration. Lift and Shift on-prem IBM Netezza and HDFS Hive to EMR Hive on S3 ▶ Update ETL into Airflow from Oozie and Talend ▶ Monitor data readiness with oncall rotations	
	Insight Data Science , <i>Data Engineering Fellow</i> New York City 2015–2016	
	▶ 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 , <i>Software Engineer</i> New York City 2010–2014	
	▶ 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 , <i>Senior Web Application Developer</i> New York City 2006–2009	
	▶ 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 , <i>Associate Software Engineer</i> Milford, MA 2000–2001	
	▶ 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	