Holden Karau   http://www.holdenkarau.com

San Francisco, CA, USA

E-mail: holden@pigscanfly.ca, Cell: (425) 233-8271

Highlights of Achievements

- Apache Spark Committer, PMC member, and ASF member
- Received IBM OTAA award and Google Open Source Peer Bonus for work on Apache Spark
- Frequent Conference Speaker
- Co-Author of Spark books: Learning Spark - `https://amzn.to/3bTg434` and High Performance Spark - `https://amzn.to/2LMqneN`
- Co-Author of Kubeflow for Machine Learning: From Lab to Production - `https://amzn.to/3sIQtjl`
- Co-Author of Python Books - Scaling Python With Ray `https://amzn.to/42q89kv` and Scaling Python With Dask `https://amzn.to/3qBN9da`
- For more see `https://github.com/holdenk`

COMPUTER   SKILLS

Languages Scala, Java, Python

Not so recent languages Scheme, Perl, C, C++, SQL

Development Tools Emacs, SVN, GIT

EXPERIENCE

Netflix   June 2021 - Present

OSS Software Engineer

- Lead Spark-Upgrade Project to automate Spark Migrations
- Improved Spark dynamic scaling and Kubernetes integration.
- Created internal and external big data training material.

Apple   August 2019 - June 2021

Software Engineer, TL Apache Spark

- Improved Spark dynamic scaling and Kubernetes integration.
- Shared best practices and mentored new contributors.
- Encouraged OSS participation and best practices.

Google (Cloud)   October 2017 - August 2019

Open Source Big Data Developer Advocate

- Joined Apache Spark PMC
- Became a member of the Apache Software foundation
- Worked on improvements for Spark focused in Core, ML, and Python
- Provided guidance for OSS based big data products, including Dataproc and Dataflow.
- Helped grow Beam and Spark contributors and community.
- Gave talks and training sessions for Spark, Beam, and Kafka.

IBM - Spark Technology Center (Watson Data Platform)   October 2015 - October 2017

Principal Software Engineer

- Became Apache Spark Committer
- Worked on improvements for Spark focused in Core, ML, and Python
- Helped other developers get up to speed and contribute to Spark (both in code & community)
- Gave talks and training sessions for Spark

Alpine Data Labs   May 2015 - October 2015

Senior Software Development Engineer

- Worked on improvements to Apache Spark and Alpine's Spark integration

Databricks   February 2014-May 2015
Senior Software Development Engineer
- Worked on backend improvements for Databricks Cloud
- Worked on improvements for Spark
- Created and improved integrations with external libraries and projects
- Taught Spark Trainings

Google   August 2012-February 2014
Software Development Engineer
- Worked on improvements to the Google+ storage backend
- Worked on different methods to improve recommendations on Google+
- Taught new hire engineering class

Foursquare   March 2011-August 2012
Software Development Engineer
- Worked on improving venue & user search in Scala using Elasticsearch, Solr, and Lucene

Amazon   May 2008 - August 2008, September 2009-2011
Software Development Engineer
Software Development Intern
- Created product miss-classification detection system using AI techniques
- Helped create a system to recommend categories for search keywords

Microsoft   January 2008 - April 2008
Program Manager Intern
- Spec-d and costed various performance improvements

Google   Summer 2007
Summer of Code
- Gained commit access to Subversion repository and contributed bindings for Subversion in Scheme

Pigs Can Fly Computing and Totally Legit Co.   Summer 2002 - Present
Software Developer and Everything Else
- Employed, supervised, and lead teams of programmers in software development
- Worked collaboratively with clients to specify their objectives and interpret these goals to the team
- Proposed, developed, and implemented source code search engine, All The Code
- Scaled Ruby on Rails application in collaboration with AideRSS
- Created LLM for generating health insurance appeals

Xandros - Linux Distro   September 2006 - December 2006
Software Developer
- Improved wireless signal strength reporting in numerous drivers
- Forward ported various wireless drivers

Net Integration - Linux based intelligent servers   Jan 2006 - April 2006
Evil Death Ray
- Created a transparent dynamic mirror selection module
- Wrote a library for automated cross-platform integration testing

Department of Housing Technology Waterloo - Residence computer network   Summer 2005, Fall 2005, May 2006 - September 2007
Senior Residence Computer Consultant & Technical Support Specialist
- Trained and supported team in maintaining and utilizing networking tools
- Created virus removal tools to handle viruses not removed by Norton Antivirus

EDUCATION

Bachelor of Math, Computer Science, Business Option (Co-Op) 2004-2009
University of Waterloo, Waterloo, Ontario, Canada