# Didier Deshommes

*1413 Oriole Ave.*
*McAllen, TX 78504*
*(919) 616-0344*
*dfdeshom@gmail.com*

## Executive Summary

**Summary** Python data engineer who has designed and implemented large-scale production systems from conception to deployment. Experienced with writing, deploying and maintaining data processing pipelines for analytics and search workloads, high-throughput RESTful APIs, web crawlers and internal backend tools.

**Production Experience** Python, AWS/EMR, Spark, Chef/Ansible, Postgres, Elasticsearch, Redshift, Kafka, MongoDB, Docker, DroneCI, Redis

**Languages** Python, Golang

## Experience

January 2017 - October 2019

**Data Engineer**, *Restless Bandit*, San Francisco, CA.

- Responsible for maintaining and improving our data pipeline using Spark and ElasticSearch. Increased the speed of out data-processing by doing a combination of: optimizing critical part of our code (using Cython) and designing better ways to run our daily Spark jobs.
- Reduced cost of running daily EMR jobs by over 50% in terms of normalized instance hours by analyzing job resource usage . Reduced the total time time it took jobs to run from 4 hours to less than 2 hours. Made data pipeline more resilient by installing monitoring and tracing checks throughout our code for faster debugging.
- Created internal RESTful web APIs for use by our other data pipelines using Python, Flask and Celery. Created various ansible/Docker playbooks for provisioning and deploying machines running these APIs.
- Lead migration of over 30 internal python projects from Python2 to Python3. Added integration and unit tests to all python projects, along with CI integration (using Drone CI) to ensure a safer transition.

May 2016 - January 2017

**Software Engineer**, *Shareaholic*, Boston, MA .

- Lead design and implementation of a system that built user profiles (stored in S3) from various attributes such as geographic location, content of pages visited, referrer data, etc. These profiles were then used for highly-specific targeting to various data partners. Wrote the tooling to launch and run the data pipeline in our AWS cluster, using Python and Spark . The system has delivered thousands of additional revenue to data partners in its first 3 months existence by combing through over 100 million data points every day.
- Lead engineering effort to revamp our 100-machine infrastructure managed by Chef. Re-wrote many of our custom recipes and roles, added more recipes to increase automation. Wrote a tool to make cloning or creating new machines from AWS simpler.

Spring 2011 - February 2016 **Backend Lead**, *Parse.ly*, New York, NY (Fully Remote Team).

- Lead implementation of major backend systems for an engineering team that grew from 3 to 20 over the course of 4 years. During this period, the company also grew to hundreds of customers, millions in revenue and massive data volumes. An overview of Parse.ly's core product is available on its website.
- After seed financing round in 2011, company hired other engineers to focus on areas like dashboard design and time series data. Took full ownership over backend systems and APIs while company pivoted into the content analytics market, where it ended up finding success.
- Built our API, which allowed customers to access analytic and recommendations data. It was adopted by 40% of customers and by 2015 was serving 2 billion calls per month. It serves 500 requests per second and peaks up to 2,000 requests per second. It powers popular websites like The New Yorker and Arstechnica. Also wrote its public documentation: https://www.parsely.com/docs/api/overview/endpoint.html . An overview of how the recommendation engine works can be found here: http://blog.parsely.com/post/3406/reco-engine/
- Created the crawling system that processes over 100 million URLs per month. Crawlers collect informative metadata from publishers' pages for display on the Parse.ly's Dash analytics dashboard.
- Implemented several features in our real-time and batch analytics pipeline that processes hundreds of millions of events per day. The pipeline powers the Parse.ly's Dash analytics product.
- Designed and implemented flexible auto-complete search for our dashboard analytics users. Users could complete either the beginning, middle or end of several words. Users were able to search millions of titles, authors and sections in a unified way in less than 200ms.
- Designed and developed named entity recognition system using Solr, the Wikipedia index, Wikipedia traffic data, and NLTK. With this feature, users were able to see what topics were trending on the web based on the traffic they got from Wikipedia traffic.
- Automated configuration and deployment for machines that were used to do API, crawling or search work using Chef and Vagrant.
- During period of high growth, performed several technology evaluations as company refreshed its technology stack for scale. Company raised series A financing in 2013 and grew to a total headcount of over 40, while landing new customers that included the highest-traffic websites on the Internet, such as Reuters, The Huffington Post, Business Insider, and WIRED. Evaluated distributed systems like Elasticsearch and Cassandra, which were eventually adopted to replace MongoDB and Solr.

Summer 2009 - Spring 2011 **Software Engineer & Employee #1**, *Parse.ly*, New York, NY (Fully Remote Team).

- First engineering hire at an early-stage startup that was part of Dreamit Ventures 2009, a YCombinator-like startup accelerator program in Philadelphia, PA.
- Wrote and launched the backend to the Parse.ly Reader, a news article reader (similar concept to Flipboard, but before Flipboard's launch) that displayed articles from the web based on a user interests. Worked directly with CTO. A review of the Parse.ly Reader is here: http://blog.louisgray.com/2009/08/parsely-spices-up-news-based-on-your.html
- Rewrote backend to support a content recommendation API for publishers, as company moved into the enterprise API space serving content publishers. These systems helped the company land its first customers. Worked as sole full-time engineer through Parse.ly's bootstrapping period.
- Company raised a seed financing round in 2011 and changed its focus, while re-using technology developed during this founding period.

Winter 2007 - Spring 2008 **Python Consultant**, *Wordstream*, Boston, MA.

- Designed and implemented a keyword-matching algorithm that accepts documents as its input and produces a list of suggested hyperlinks to be added by a user. The suggested keywords are based on the user's profile data.

2007 - 2008 **Software Engineer**, *Intelligent Information Systems*, Durham, NC.

- Designed and implemented a website so clients could access and browse property inspection information
- Implemented and improved a procedure to port and test programs written in VB6 to VB.NET. Converted over 100 programs written in VB6 code to VB.NET using a combination of the MS conversion tool and Python.

## Education

Fall 2006 - Winter 2008 **MS Applied Mathematics, North Carolina State University**, *Raleigh, NC.*

Fall 2001 - Winter 2006 **BS Computer Science, North Carolina State University**, *Raleigh, NC.*

Fall 2001 - Winter 2006 **BS Applied Mathematics, North Carolina State University**, *Raleigh, NC.*

## Extra

- Presenter,PyData: "Wikipedia Indexing and Analysis" - see https://vimeo.com/53091620
- Contributor to open source projects - see https://github.com/dfdeshom
- Fluent in French
- B1-B2 level in Spanish