

# Charles Menguy

Data science engineer with 6+ years of professional experience in the online advertising industry. Fluent in Java and Python, intermediate in C++ and Scala. Comfortable with Unix/shell scripting. Well-versed in big data technologies and applying data science techniques at large scale.

## Work Experience

Since 2012

**Senior Data Engineer** for [Adobe](#) (New York City, United States).

- Complete backend development of a customer segmentation platform from scratch in *Java* using *Hadoop*, *HBase* and *Cassandra*, managing more than 10 Billion profiles. This was done by leveraging *Amazon Web Services* in multiple regions, in particular *EC2*, *EMR* and *SQS*.
- Development of a *Python*-based *data science platform* to provide analytics on multi-dimensional data coming from different data sources with *Hive* and *Redshift* processing dozens of Terabytes daily. 150+ Terabytes stored.
- Participation in the development of an algorithmic segmentation framework based on *Hive* using *TF-IDF*.
- High-frequency distributed multi-threaded cookie exchange service built in *Java* with a custom streaming protocol based on *SQS*, making tens of thousands HTTP requests per second to dozens of ad networks/DSPs/DMPs.
- Research and development of a cross-device stitching engine processing hundreds of Terabytes weekly via *EMR* and *Spark* using graph-based cluster algorithms, coordination via *AWS Data Pipeline*. Main development in *Scala*, analysis and evaluation via *Python notebooks*.

2011-2012

**Senior Software Engineer** for [Proclivity Media](#) (New York City, United States).

- Development of a scoring engine using *Bayesian inference* to compute business value of website visitors.
- Using *Hadoop* and *Pig* to develop reporting infrastructure, and operational improvements.
- Implementation of a job controller to standardize the company's *Hadoop* pipeline and scheduling.
- High-frequency monitoring system to collect *Hadoop* metrics via *ZeroMQ* and analyzed via *Esper* to detect outliers and trends and visualized in *Graphite*.

2010-2011

**Software Engineer** for [24/7 Real Media](#) (New York City, United States).

- Implementation of scaling improvements on a custom *C++*-based distributed system.
- Research and development of customer segmentation engine in *C++*.
- Modernization of backend using *MongoDB* after extensive benchmarking of several *NoSQL* solutions.

2009 (5 months)

**Internship** at [GFI Informatique](#) (Sophia-Antipolis, France).

Research on *image processing* techniques to recognize QR codes on mobile and implementation of a mobile virtual shop using *J2EE* technologies.

## Open-Source Contributions

Python

Author of [ndopt](#), a library for optimizing n-dimensional spaces using particle swarms. Author of [griddle](#), a library for optimizing grid-like patterns using simulated annealing.

Java

Author of the [cloudwatch-metrics](#) library to collect Hadoop metrics into Amazon CloudWatch.

Misc

My [StackOverflow reputation](#) is close to 20k, among the top 2% contributors, and top 5 in the Hadoop community.

Open Data

I like taking open data apart to find something useful, and [blog](#) about it.

Author of [crime-analytics](#), an analysis of real crime data from San Francisco and Seattle to find temporal and geospatial patterns.

## Education

- 2007-2010 *MS in Computer Science in [EPITA](#) (Paris, France).*  
Top French school of Computer Science and Applied Mathematics with highly selective recruitment. Graduated with highest honors. Major in cognitive science and machine learning.
- 2005-2007 *Preparatory classes in [Lycée Chateaubriand](#) (Rennes, France).*  
Mathematics and Physics courses before a national competitive examination to Grandes Écoles. Equivalent to a US Bachelor's degree.
- 2005 *High school diploma (SAT equivalent) with highest honors.*

## Activities and Interests

- Hobbies I like traveling around the world and discovering new cultures and ways of thinking. I enjoy learning more about various topics, using platforms such as [Coursera](#), [Udacity](#) or [edX](#).
- Activities I have been teaching assistant at [Coursera](#) for several computer science and AI-related courses.  
Also a technical news editor at [InfoQ](#) focusing on the big data and data science topics.

## Recommendations

Available upon request.

---

[menguy.charles@gmail.com](mailto:menguy.charles@gmail.com) • +1 (917) 294-3487 • 28 years old • permanent resident  
350 E 62nd Street, New York - NY 10065, United States

---