

**Andrew Ehrlich**  
3486 17th St  
San Francisco, CA 94115  
andrew@aehrllich.com

## SUMMARY

I build data pipelines and analytics platforms for businesses. I can provide all components of data collection and ingestion pipelines, including software development, schema design, ETL, data quality best practices, infrastructure planning, server deployment and monitoring, disaster prevention and recovery, real time processing, and Amazon Web services.

## EXPERIENCE

**Centro, Senior Software Engineer** **May 2016 – Present**

**Centro, Data Engineer** **September 2015 – May 2016**

- Apache Spark to HDFS ingestion pipeline.

**Unified, Platform Engineer** **February 2014 – September 2015**

- Replaced RabbitMQ with a distributed data collector, Kafka cluster, Spark jobs, and Redshift tables.
- Developed distributed Java framework to pull JSON records from external APIs. Used Kafka for both job scheduling and storing incoming data.
- Wrote ETLs in Cascading + Hadoop 2.4 and Scala + Spark 1.3 to clean, validate, flatten, dedupe, and load into Redshift.
- Deployed, monitored, and maintained Kafka 0.8.1 brokers and consumer groups in production and staging.
- Optimized Redshift tables for consistency and fast queries.
- Deployed servers in Amazon EC2 and EMR.
- Created multiple deployers, implemented a source control branching model and ran releases every 2 weeks.

**Kixeye, Lead Analytics Engineer** **January 2013 – October 2013**

- Transitioned data storage from MongoDB to Hadoop.
- Developed Python ETLs to extract gameplay events from Hive tables into MySQL.
- Developed JSON schemas for games to send data to analytics APIs.
- Weekly 1:1s with developers.

**Kixeye, Junior Analytics Engineer** **September 2011 – January 2013**

- Developed MySQL data warehouse and Python ETLs for player profiles and game sessions.
- Developed Python ETLs to integrate Facebook Ads API reports into data warehouse.

**Interaction Design Lab, Cornell University, Developer** **June 2010 – August 2011**

- Projects to encourage healthy decisions by using special mobile apps to increase social awareness.

## EDUCATION

- Cornell University, Dyson School of Applied Economics and Management, Environmental and Resource Economics, May 2011.

## SKILLS

Languages: Scala, Java, Python.

Databases/Frameworks: Spark, MapReduce, Hive, Pig, MySQL.

Amazon Web Services: EC2, Elastic MapReduce, VPC, S3, Redshift.