

# DANIEL DIXON

## Senior Data Engineer

256.509.9178



daniel@dixon.xyz



San Diego, CA



[www.daniel.dixon.xyz](http://www.daniel.dixon.xyz)



### Professional Summary

Lead Big Data Engineer with 10+ years of distributed computing experience. Most recently in the financial services industry where I'm responsible for designing, building, testing petabyte-sized data pipelines with Spark and Python. My role also includes supporting rapid model deployment into complex low-latency real-time environments. I have experience with cloud platforms including AWS and GCP and I'm looking for opportunities to further my proficiency in that area.

### Experience

#### Wells Fargo, San Diego, CA

2015 - Current

##### Senior Data Engineer

- Lead Spark engineer responsible for designing, building, and optimizing Terabyte-sized data pipelines for both model training and scoring in the fraud space. Work closely with Data Scientist in small, agile development pods to build deployable data pipelines and models.
- Heavily involved in model deployment, guiding discussions and requirements around data flow and service architecture for low-latency real-time fraud models
- Guide discussions with business partners to understand needs, gather domain knowledge, and ensure objectives are met. Take ownership of key objectives and artifacts to drive project to completion.
- Principal tester for new ML tools and platforms to help modernize the company-wide analytics environment. Multiple successful adoptions have led to improvements in all phases of the model development life cycle.
- Directed the creation of internal standards and best practices around data engineering company-wide. These changes were critical in helping to reduce model development time from a year to weeks.
- Regularly create and conduct internal training sessions for hundreds of analysts within the company. Founded and currently moderate a large internal Spark community.

#### Aventrix, San Francisco, CA

2017 - 2018

##### Machine Learning Engineer Consultant

- Designed and constructed an online activity recommender system hosted on AWS. The system involved EMR to tag and vectorize historical entities, SQS to manage asynchronous processing, Lambda and EMR to perform a similarity search (content-based filtering) in near real time, RDS to store recommendations, and CloudFormation to provision the infrastructure.
- The provided recommendations measurably increased the click-through rate when compared to a baseline of random recommendations (determined through A/B testing).

#### Claraview, San Francisco, CA

2011 - 2015

##### BI Consultant

##### Engagements:

##### Adobe, San Jose, CA

- Lead Tableau advisor and subject matter expert for Adobe's Data Science and Analytics team.
- Headed team to develop several key dashboards used internally across multiple product teams. The dashboards were challenging as performance was key, and most of the reports included non-aggregatable metrics across numerous dimensions.
- Analyzed slow performing Tableau server and dashboards and made suggestions for improvement. In one case, the

initial dashboard load speed was improved from ~15 min to ~5 sec.

- Responsible for crafting stylistic guidelines to be used across Tableau reports and dashboards.
- Developed several key Tableau workbooks against very large datasets stored in a Hadoop environment.
- Wrote and modified SQL queries from a variety of platforms including SAP Hana, SAP Sybase, and Hive.
- Responsible for organizing and scheduling Tableau Server data extracts, ingesting data from a wide variety of sources including SAP Hana, SAP Sybase, Postgres, Hive, Excel, and CSV.

#### Dell, Austin, TX

- Technical lead on project to examine customer satisfaction with online media dispatch system.
- Project involved using Aster Database to combine online interactions (Omniture) with dispatch events to evaluate dispatch failures and customer frustration with the online system.
- Extensively used Aster pathing (nPath), attribution, text manipulation, and Regex functions to merge disparate datasets and chart the customer journey.
- Identified most common customer journeys that did not lead to a success event (media dispatch). Several customer frustration events were hypothesized and examined.
- Developed several summary and pathing reports using both Tableau and Aster GraphGen.

#### HealthSparq, Portland, OR

- Project to analyze Hadoop cluster performance using Hive, Impala, and MapReduce against different storage and compression formats.
- Developed MapReduce job to expand the given dataset, as sample member data provided by client was too small to test performance. This dataset was loaded into hBase tables and integrated with Hive/Impala.
- Recorded query performance and disk storage while using the storage formats Avro and Parquet and the compression formats Gzip and Snappy. This performance was then compared against a baseline of unformatted and uncompressed data.
- Explored Hadoop and R integration solutions and provide recommendations.

#### T-Mobile, Bellevue, WA

- Designed and led training sessions with the client in Aster SQL-MR functions. Most of the training focused on Aster Lens, GraphGen, Pathing and Attribution functions, and text analysis.
- Crafted training material, documentation, and several presentations detailing the abilities of the Aster Database.
- Communicated the value of the Aster Discovery Platform with examples and use cases built with the client's own data.

#### Cars.com, San Francisco, CA

- Assisted in developing ~100 Tableau reports focusing on web analytics data. The web analytics metrics were broken down by many dimensions including date, region, category, and section.
- Responsible for setting up a Tableau server to host the reports and data sources.
- Upgraded the Tableau workbooks, reports, and server to version 8, implementing new report features such as scoped filters, data source filters, improved heat maps, and multi-select comboboxes.
- Organized data sources to allow large amounts of data to quickly be aggregated and displayed.

#### Office Depot, San Francisco, CA

- Worked with a team to develop a proof-of-value solution using AsterData nCluster and Tableau.
- Using nCluster, developed several nPath and Attribution queries to analyze web traffic and search ad clicks over hundreds of millions of customer sessions. Customer pathing and attribution data allowed a detailed understanding of marketing campaign effectiveness.
- Constructed ~20 Tableau reports to visualize the nPath and Attribution functions and allow enhanced data consumption. Advanced pathing visualizations were created to display the flow of user interactions.

#### Bloomin Brands Inc, Tampa, FL

- Worked in a large team environment to develop numerous SnapLogic ETL pipelines focused on transferring restaurant invoicing and sales data. Acted as the Snaplogic subject matter expert, developer, primary QA contact, and server administrator.
- Acted as the sole consultant representing both Claraview and Snaplogic over a period of eight months.
- Worked with analysts and developers from SAP to design an interface between Snaplogic and SAP's invoicing and

billing system.

- Created a Snaplogic best practices guide and helped implement those practices with the development team.

AsterData, San Francisco, CA

- Worked with a team to develop multiple Tableau reports that were demoed at the 2011 Teradata Partners conference. The reports were used to demonstrate several new features of AsterData's nCluster.
- Assisted in constructing a data integration process to merge email campaign data from AsterData, web click data from LeadFormix, and contact data from Salesforce.com. The data was merged in a staging database and was used to populate the Tableau reports.

US Oncology, The Woodlands, TX

- Sole Jaspersoft developer
- Conducted end-user interviews to identify end user requirements
- Designed and developed several Jasper reports, including custom modifications to report behavior (jsp, JavaScript) and display (css).
- Assisted in the architecture and development of a MYSQL data warehouse back-end.
- Provided day-to-day maintenance and administration of Jaspersoft Servers.

---

## Education & Training

**Georgia Institute of Technology**, Atlanta, GA

Bachelor of Science, Electrical Engineering

Minor in Computer Science

Certification / Training:    MicroStrategy 9 CRD, Cloudera Certified Developer for Apache Hadoop (CCDH)

---

## Technical Skills

- **Programming:** Python, SQL, Scala, Java, C++, Golang, HTML, Javascript/NodeJS, Shell
- **Methodologies:** Agile, Data Modeling, Object Oriented Design
- **Database:** Postgres, MongoDB, Teradata, Oracle, Aster Database, MySQL, Sqlite
- **Hadoop:** Spark, Hive, MapReduce, HDFS, Impala, Sqoop
- **Business Intelligence:** Tableau, MicroStrategy, SnapLogic
- **Platforms:** AWS, Google Cloud Platform, Kubernetes, IBM Spectrum