


DANIEL DIXON

Senior Data Engineer

256.509.9178 

daniel@dixon.xyz 

San Diego, CA 

github.com/dnldxn 

Professional Summary

Senior Big Data Engineer with 10+ years of distributed computing experience. Most recently in the financial services industry where I was responsible for designing, building, and testing petabyte-sized data pipelines with Spark and Python. My role also includes supporting rapid model deployment into challenging low-latency real-time environments.

Experience

Wells Fargo, San Diego, CA

2015 - Current

Senior Data Engineer

- Responsible for designing, building, and optimizing Terabyte-sized Spark 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, helping to define the data flow and model service architecture
- Communicate with business partners to understand needs, gather and document domain knowledge, and ensure objectives are met
- Principal tester for new ML tools and platforms to help modernize the company-wide analytics environment
- Led the creation of internal standards and best practices around data engineering and software development
- Regularly created and conducted internal training sessions for hundreds of analysts
- Speaker at conferences including DataEdge, Teradata Partners, H2O MeetUp

Claraview, San Francisco, CA

2011 - 2015

BI Consultant

Recent 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 several 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
- 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

- Aster 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.

T-Mobile, Bellevue, WA

- 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

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
- Worked with analysts and developers from SAP to design an interface between Snaplogic and SAP's invoicing and billing system

AsterData, San Francisco, CA

- Worked with a team to develop 7 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

Education & Training

Georgia Institute of Technology, Atlanta, Georgia
Bachelor of Science, Electrical Engineering,
Minor in Computer Science

August 2010

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

Technical Skills

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