BIG DATA ENGINEER

Big Data Engineer

+1-931-273-8786 name@email.com San Francisco

EXPERIENCE

Big Data Engineer
San Francisco

Pinterest
2019 - Present

- Led the requirements gathering process with clients, resulting in accurate estimations for developing complex queries using Hive in a logistics application, achieving an average accuracy rate of 95% in meeting project timelines.
- Implemented a highly efficient data pipeline utilizing Flume, Pig, and Sqoop, resulting in a 30% reduction in data ingestion time for cargo data and customer histories into HDFS for analysis.
- Successfully imported data from MySQL DB to HDFS and vice versa using Sqoop, optimizing data transfer processes and reducing data transfer time by 40%, resulting in enhanced overall system performance.
- Developed advanced workflows in Oozie and efficiently scheduled jobs in Mainframes, resulting in a 20% reduction in overall job execution time.
- Installed and configured Hive, Pig, Sqoop, and Oozie on Hadoop clusters, optimizing cluster performance by conducting benchmarking tests. Achieved a 15% improvement in query execution time, enhancing overall system efficiency for internal use.

Data Engineer
San Francisco

Stripe
2016 - 2019

- Ingested streaming and transactional data across 8 diverse primary data sources using Spark, Redshift, S3, and Python
- Created Python library to parse and reformat data from external vendors, reducing error rate in the data pipeline by 15%
- Automated ETL processes across billions of rows of data, which saved 53 hours of manual hours per month
- Built tools to provide real-time data around international currency exchange, reducing latency by 18%

Data Engineering Intern
San Francisco

ABC Company
2015 - 2016

- Performed Data Migration from Amazon S3 to Amazon Redshift.
- Managed delivery of $1.8 million development work annually
- Developed a 6th App Store ranked health app
- Reduced overall team spending by $600,000 a year with dynamic resource allocation and flagging.

EDUCATION

Bachelor's in Computer Science
University of California, Irvine
2012 - 2015

BBA Risk Management and Insurance
University of Georgia
2003 - 2006

STRENGTHS

Data modeling

Created 15+ data models for clients, resulting in an increase in efficiency by 30%.

Problem-solving

Developed innovative solutions to complex data problems, reducing errors by 25%.

ACHIEVEMENTS

Former PMP

• Program managed global multi-organizational projects over $200M and with $1B revenue.

CRM - Infusionsoft

• Created a campaign with multiple apps that would sort leads via zip code and send a real-time email & SMS to the nearest sales rep.

PROJECTS

Summative Peer Review of Teaching

12/2018 - 01/2019
UNSW, Sydney

https://peerreview.unsw.edu.au

An online tool to facilitate end-to-end process of Peer Review of Teaching for UNSW staff members.

- Built with Django and React the application provides end-to-end support for booking, scheduling and conducting Peer Reviews of Teaching for UNSW academic staff
- Leverages Microsoft Graph API to extract staff calendar data to schedule reviews

Convolutional Neural Network Based Emotions Recognition System

05/2018 - 06/2018
UNSW, Sydney

https://github.com/vnihit/CNN-Emotions-Recognition

A CNN based emotions recognition system, built with Scikit-Learn, Tensorflow and OpenCV -

- High degree of accuracy (86%) and confidence for real-time emotions recognition
- Potential applications in - Behavioural analysis, real time analysis of customer satisfaction in retail settings, automatic photography

LANGUAGES

English (Native)

French (Proficient)