



John Smith

Data Engineer

Los Angeles, United States email@email.com

Details

1515 Pacific Ave
Los Angeles, CA 90291
United States

3868683442

Place of birth

San Antonio

Driving license

Full

Skills

SQL



Java



Apache Spark



Hadoop



Python



Coding



Languages

English



German



Hobbies

Cycling, Songwriting,
Running

Profile

Dedicated Data Engineer with 5+ years' experience dealing with large datasets. Eager to build robust databases that lay the groundwork for game-changing insights at LionHeart Algorithm LLC. Implemented natural language processing tools to ensure machine-readable databases were ready for the team of data scientists.

Employment History

Jan 2018 — Dec 2020

Nong Phai

Data Engineer at FNB

Responsible for scaling machine learning models and making these models fit within banking environments with the sci-kit-learn, Tensorflow, and Keras, also adapting these models/architectures created by data scientists to fit in with the UX and the UI of the brand.

- Translated business propositions into quantitative queries and organized the necessary data.
- Developed scalable databases capable of ETL processes using SQL and Spark.
- Estimated the workflow and increase the efficiency of data pipelines that process over 60 TB of data daily.
- Utilized MongoDB to create NoSQL databases that collect data from a variety of sources.
- Involved in creating tables, join conditions, partitioning tables, correlated subqueries, views, sequences, nested queries, and synonyms for business application development.

Jan 2015 — Dec 2017

Washington

Data Engineer at ABSA

Responsible for developing database triggers, packages, functions, and stored procedures using PL/SQL and maintain the scripts for various data feeds across multiple regional and international offices of the company

- Co-develop a SQL server database system to maximize performance benefits for clientele.
- Assisted senior-level Data Scientists in the design of ETL processes, including SSIS packages.
- Developed coherent Logical Data Models that helped guide important client business decisions.
- Collaborate and coordinate with development teams to deploy data quality solutions and create and maintain standard operating procedure documentation.

Education

Jan 2021 — Jan 2021

Austin

Texas University

BS, Computer Science

- Dual concentration Machine Learning,
- Received a Business Foundations Certificate
- Part of the winning team for the 2016 Longhorn.

Courses

Jan 2021 — Jan 2021

CCA Cloudera Certified Associate at Cloudera, Online.

Achievements

- Developed a data pipeline with Delta Lake that led to process optimization and a corresponding revenue increase of 21%.
- Successfully figured out ETL issue while following PL/SQL best practices that resulted in an insight that increased the client's customer base by 37%.