

Wm. Keith van der Meulen

wkeithvan.com | +1 (206) 300-4878 | keith@wkeithvan.com

Data engineer with an extensive background working with petabyte-scale datasets and analytics-heavy workloads. Experienced at designing and optimizing for highly efficient pipelines with standardized, reusable, and understandable code structure. Skilled at designing high quality data models for use in analytics, data science, and machine learning.

Education

University of Calgary

Sep 2012 – Apr 2018
Computer Science, B.Sc.
Electrical Engineering, B.Sc.
Computer Engineering, minor

Skills

Programming Languages

Python, SQL, JavaScript, Java,
HTML, CSS, C, C++, C#

Cloud Services

Databricks, Microsoft Azure,
Google Cloud Platform, Amazon
Web Services

Databases

Snowflake, Delta Lake, SQL
Server, Google Big Query, Hive,
Presto, Cassandra, MySQL,
Cosmos, Mongo DB

Frameworks

Apache Spark, Streamlit, Flask,
GraphQL, React, REST APIs,
DataSwarm

Libraries

PySpark, NumPy, SciPy, Pydantic,
Keras, TensorFlow, jQuery,
CloudEvents

Software Products

Fivetran, Power BI, Jira, PyCharm,
VSCode, Git, Mercurial, SSMS, MS
Office

Technical Interests

Computer vision, robotics, data
science, economics, mathematics,
synthetic biology, rocketry,
computational linguistics, online
communities

Work Experience

Data Engineer

Stack Overflow

Jan 2022 – Present

- Lead data ingestion strategy and prioritization of first party data sources for use in generative AI, search algorithms, and click-based advertising
- Author company-wide event message standard for use in data streaming, microservices, and AI products
- Automate usage of Azure Data Lake metadata to create and maintain Snowflake warehouse schemas, data, and roles for company-wide data access
- Develop hierarchical role-based access patterns for flexible data governance
- Design cloud-based data architecture to deprecate on-premise data processing servers, event collection systems, and A/B user testing tools
- Lead sales discussions for onboarding Fivetran as a SaaS partner and develop internal Git-based governance software for rollout across multiple departments

Data Engineering

Accenture AI

Aug 2019 – Jan 2022

Data Engineer

Meta (Facebook)

May 2021 – Jan 2022

- Build financial forecasting pipelines in SQL for data science & analytics teams
- Explore datasets for discovering new predictive insights across terabyte-scale datasets and productionize results
- Refactor financial tables to improve data quality for data science teams

Lead Analytics Engineer

Accenture Research

Aug 2020 – Apr 2021

- Manage two separate projects simultaneously with data engineers, analysts, and QA testers spread across multiple continents
- Coordinate business managers' objectives into technical requirements
- Optimize Spark jobs to reduce run time by up to 95% on customer datasets
- Unify multiple codebases into one cohesive, standardized internal library
- Refactor Python codebase to zero linting and static code analysis errors

Data Engineer

Jones Lang LaSalle, Inc. (JLL)

Oct 2019 – Jul 2020

- Produce automated data ingestion tools on Azure Functions for 3rd party data
- Develop enterprise GraphQL API that generates itself from an external data schema configuration file using Python meta-programming techniques
- Write job orchestration scripts for running daily Databricks jobs

Data Engineer

Talroo, Inc.

Feb 2019 – Apr 2019

- Build production ETL notebooks for generating datasets from raw data streams
- Explore terabyte scale datasets with Spark for geographic trends for targeted advertising opportunities across multiple market segments
- Construct Docker-based React/Flask/SQL web app for creating labelled datasets

ML Engineer

Hotchkiss Brain Institute

May 2018 – Aug 2018

- Train and test machine learning (ML) networks to analyze MRIs for brain lesions
- Create Python-based ML models with TensorFlow and Keras frameworks
- Analyze large medical datasets to find statistical correlations and prep for ML