

# DUSTIN SMITH

Hanoi, Vietnam

+84 (0)90 469 9938

[✉ Dustin.William.Smith@gmail.com](mailto:Dustin.William.Smith@gmail.com)

[in linkedin.com/in/dustin-s-photo](https://www.linkedin.com/in/dustin-s-photo)

[github.com/dwsmith1983](https://github.com/dwsmith1983)

[dustinsmith.info](https://dustinsmith.info)

## Experience

### Director of Site Reliability Engineering

January 2025 – Present

*Techcombank – Data & Analytics Division*

*Hanoi, Vietnam*

- Built and led SRE team from zero headcount in 9 months, scaling organizational monitoring from 600 to 3,721 data pipelines with comprehensive observability frameworks
- Directed team to achieve 97.8% SLA compliance (exceeding 97.5% target) by implementing reliability frameworks and error budget tracking across 401+ critical jobs spanning 13+ teams
- Led cost optimization initiatives delivering \$1,034,919 in savings: \$954,525 recurring (Feb-Dec) and \$80,394 from advanced S3 optimization strategies
- Orchestrated 22% AWS production cost reduction (from \$346k to \$270k monthly) through strategic S3 lifecycle policies, storage class transitions, and multi-region environment cleanup
- Championed organization-wide adoption of Python/Scala coding standards with standardized templates (Hatch, Ruff, Black, pre-commit hooks, scalafmt, jacoco) and 75% minimum code coverage requirements
- Guided team delivery of 6 production repositories and 2 code templates including Delta Lake optimization, CoB automation, shared SRE utilities, IaC monitoring, and SLO-based reliability tracking
- Introduced STAMP safety analysis methodology for data pipeline safety
- Established RACI framework defining clear ownership structures across pipeline, team, and functional group levels for SLIs, SLOs, and error budget management

### Solutions Architect

June 2023 – January 2025

*Databricks*

*London, UK*

- Developed analytics strategies with executives across Central and Middle East markets, specializing in data engineering and telecom
- Mentored SAs organization-wide on data engineering best practices and conducted field engineering pilot for technical hiring
- Led data migrations from on-premise and legacy tech stacks to Databricks, optimizing cluster configurations and code performance
- Contributed to GitHub action integration for internal tooling with code refactoring initiatives

### MLOps Manager

July 2022 – April 2023

*Delivery Hero*

*Berlin, Germany*

- Developed and implemented MLOps roadmap with OKRs, growing and mentoring the MLOps team
- Built MLOps framework on GCP supporting multiple ML frameworks and deployment pipelines

### Associate Director – Solutions Engineer Lead

March 2019 – June 2022

*True Digital Group*

*Bangkok, Thailand*

- **Platform Operations (Aug 2021-Jun 2022):** Migrated credit models from on-premise to GCP/Databricks, implementing MLOps infrastructure with framework-agnostic approach (TensorFlow, MLMD, TFX, OSS Spark, H2O, MLflow)
- **Analytics Platform (Jan 2020-Aug 2021):** Optimized HDFS data compaction achieving 2.331 PB savings over 2 years (40M THB cost savings); increased Kafka geo-streaming capacity 1200% from 50 to 630 locations
- **Analytics Product (Mar 2019-Jan 2020):** Developed automated customer segmentation engine using DPI data; created BTS user tracking methodology without triangulation

### Lead Data Scientist

May 2018 – February 2019

*AppSmiths Technology*

*Bangkok, Thailand*

### Data Analyst – Contractor for Caterpillar Inc

September 2017 – April 2018

*Tata Technology*

*Bangkok, Thailand*

### Data Scientist III

January 2016 – September 2017

*Caterpillar Inc*

*Champaign, Illinois*

Technical Skills

**Languages:** Python, Scala, SQL, Bash, LaTeX, Markdown  
**Data Engineering:** PySpark, Spark, Databricks, Delta Lake, Airflow, Kafka, HDFS  
**MLOps & DevOps:** MLflow, TensorFlow, Docker, Kubernetes, Git, GitHub Actions, Jenkins  
**Cloud Platforms:** AWS (S3, RDS, EKS, CloudWatch), GCP, Azure  
**Monitoring & Observability:** Prometheus, Grafana, CloudWatch, Error Budget Tracking  
**Operating Systems:** Linux, macOS, Windows

Education

<b>University of Vermont</b> <i>Master of Science in Mechanical Engineering</i>	<b>Burlington, Vermont</b> <i>January 2015</i>
<b>University of Tampa</b> <i>Bachelor of Science in Mathematics; Bachelor of Science in Finance</i>	<b>Tampa, Florida</b> <i>May 2011</i>

Publications & Open Source

<b>Data Optimization for Compacted Partitions</b>   <i>True Analytics Tech Blog, Towards Data Science</i> <ul style="list-style-type: none"><li>Semi-linear and z-ordering optimization for compacting partitions achieving 2.311 PB storage savings</li></ul>	<b>2021</b>
<b>Configuration Files in Python using dataclasses</b>   <i>True Analytics Tech Blog</i> <ul style="list-style-type: none"><li>Demonstrated dataclasses and dataconf for config parsing similar to Scala’s pureconfig</li></ul>	<b>2021</b>
<b>dataconf Contributor</b>   <i>GitHub</i> <ul style="list-style-type: none"><li>Added Scala case class parsing behavior for Python dataclasses; created <a href="#">demo repository</a></li></ul>	<b>2020–2021</b>
<b>spark-pipeline-framework</b>   <i>GitHub</i> <ul style="list-style-type: none"><li>Configuration-driven Spark pipeline framework with HOCON configs; published to Maven Central</li></ul>	<b>2025–Present</b>
<b>spark-bestfit</b>   <i>GitHub</i> <ul style="list-style-type: none"><li>Library for fitting 100+ scipy distributions using Spark parallelization; full documentation on ReadTheDocs</li></ul>	<b>2025–Present</b>

CliftonStrengths

**Top Strengths:** Learner, Competition, Achiever, Deliberative, Consistency, Analytic, Intellection, Adaptability, Input, Command