

James Soehlke

Dunedin, FL

-Email me on Indeed: <http://www.indeed.com/r/James-Soehlke/37b3b169a7717e58>

Professional Summary

With over a quarter-century in the data realm, I bring a blend of seasoned expertise in data science, data architecture, data/software engineering, and business intelligence. My proficiency covers an array of programming languages, from Python and Java to SQL and R, complemented by hands-on experience with development tools like Jira, Bitbucket, GitHub, and GitLab. My DBMS experience includes PostgreSQL, MySQL, Oracle, Neo4J,

Snowflake and Teradata. Additionally, I'm well-versed in visualization tools such as Power BI and Tableau and have experience with data solutions implemented in Azure and Amazon Web Services. My data management acumen encompasses practices from architectural design to governance, and I've employed tools like ERWin, Oracle SQL Developer Data Modeler, MySQL Workbench, and Visio for modeling both relational and dimensional data structures. I am actively seeking senior or lead technical positions where I can harness my expertise to enable top-tier management's ability to make informed, data-driven decisions.

Work Experience

Azure Cloud Data Platform Engineer - USCENTCOM MacDill AFB (Hybrid)

TEKsystems

October 2023 to Present

- Participated in the implementation of the Data Lakehouse of the Data Analytic Environment (DAE) in Azure Synapse Analytics, aligning with the framework outlined in the influential whitepaper co-authored at MITRE in 2021.
- Played a pivotal role in defining user personas for Azure Synapse Analytics, facilitating tailored solutions and optimizing platform usability.
- Conducted thorough testing of Azure Synapse Analytics capabilities, meticulously evaluating performance and functionality against defined personas.
- Innovated the demonstration medallion structure within the lakehouse environment, enhancing data organization and accessibility.
- Implemented a serverless database solution to effectively manage external tables residing within the data lakehouse, streamlining data exploration and analysis.
- Engage with data analytic stakeholders to deploy and transition DAE Data Lakehouse solutions, providing comprehensive support and training throughout the process.

Databricks Engineer (Contract to 1099 Position) - (Fully Remote)

540, Inc.

September 2023 to October 2023

- Spearheaded the development of data pipelines facilitating the transition from a major suite OBIEE analytics to a more robust Databricks and Qlik Sense solution hosted in AWS (ADVANA) for a Department of Defense financial unit.
- Utilized PySpark with Python and SQL to construct efficient data pipelines, enabling the seamless migration of data from Oracle to Databricks.

Senior Data Engineer

Watsco, Inc.-Miami, FL

November 2021 to August 2023

- **Subsidiary Data Pipeline Development:** Implemented first SAP R\3 ERP pipeline, ingesting data from Watsco's subsidiary in Mexico City into the

Watsco Enterprise Data Warehouse (EDW) which is a star schema fact and dimension implementation in Snowflake on AWS. Integrated TE subsidiary's Mincron ERP supply chain data into the Watsco EDW. Programmed in pipelines in Java and dimensional and relational modeled using

Oracle SQL Developer Data Modeler.

- **Watsco Environmental, Social & Governance (ESG):** Designed and deployed an application to quantify both cost and carbon emission savings attained from sales of high-efficiency HVAC equipment. Utilized R, Python, Jupyter Notebook, Snowflake, and Sigma for this development.

- **Proof of Concepts (POC):** Spearheaded a POC using Nifi, successfully extracting Watsco's enterprise data from its private cloud to the Snowflake

EDW on AWS. Conducted analytic POCs using Sigma, Snowflake's Python Snowpark, and Jupyter Notebook.

- **Other Responsibilities:** Develop Python automation scripts, manage Terraform HCL files for provisioning infrastructure, and support Watsco's proprietary ETL tool in AWS.

Data Engineer/Scientist

MITRE-Tampa, FL

December 1999 to November 2021

(2018 - 2021)

- **RECAST Planner:** Implemented a workstation variant of USCYBERCOM's RECAST network intrusion detection game to expedite planner algorithm development, utilizing Python, Jupyter Notebook, and Kafka.

- **Software Evaluation Test Harness for Common Operating Picture (COP):** Designed a test harness to uniformly capture and reproduce sensor data for COP test assessments. Developed data pipelines for diverse sensor data, transforming metadata into XML messages that align with COP standards.

- **Enterprise Data Analytic and Application Development Analysis:** Served as the technical leader for a team that architected the data analytics environment for CENTCOM.

- **CENTCOM Commander's Daily Brief:** Piloted a team that developed a prototype Power BI dashboard, automating the manual CENTCOM

Commander's daily briefing process. Utilized Python for data extraction and integration into a MySQL database, with project management using

JIRA and SAFe agile methodologies. Used MySQL Workbench Data Modeler to model data.

- **Counter-Narcotics Impact & Trends POC:** Initiated a Power BI dashboard, later migrated to Tableau to incorporate ArcGIS location functionalities. Established NiFi data pipelines for data extraction and integration from multiple law enforcement and governmental agencies.

- **Person of Interest Database:** Demonstrated automated integration and reconciliation across varied data flows from numerous law enforcement agencies. Deployed NiFi data pipelines to support a Laravel web application backed by a MySQL database. Used MySQL Workbench Data Modeler to model data.

- **Records Management Document Auto-tagging POC:** Spearheaded a team developing auto-tagging capabilities for USCENTCOM. The automated process utilized entity and temporal data extraction for legal compliance and enhanced organizational knowledge management.

- **AFSOC personnel management dashboards:** Orchestrated a set of personnel management dashboards in Tableau. Implemented the career assessment tool for AFSOC and all Python data pipelines supporting dashboards while ensuring PII data security. Created a PostgreSQL database to manage personnel and supporting data for dashboards and extracted data via python from source systems.

- **FAA ARTCC Route & Center Controller Assignment:** Conceived two optimization algorithms for the FAA. Utilized Neo4J for a graph database implementation to apply community detection to determine air route placement to an ARTCC and a Python-based algorithm for optimal air route center controller positions within an ARTCC.

- **Tower Visibility API migration to AWS:** Successfully transitioned an Apache TOMCAT Java API to an AWS API Gateway and Java Lambda function to bolster the FAA Tower Obstruction Visibility application.

- **MITRE Social Integrity Research Project:** Initiated a project targeting disinformation sources on social media. Introduced a Python lambda function for sentiment analysis using the NLTK as an AWS

Comprehend alternative. Delved into efficient techniques for social media thread extraction using AWS, leveraging services such as EC2, S3, Lambda, Aurora, Athena, and Redshift.

- Cursor on Target (CoT) Validator: Developed a Java tool for AFSOC to validate CoT messages against established standards.

- Evaluate Gold Standard Corpus: Collaborated with MITRE linguistic experts to review a Gold Standard Corpus crafted by the John Hopkins

Applied Physics Lab for SOCOM. This collaboration resulted in a MITRE-authored report detailing the corpus's merits and areas for improvement.

Lead Data Analyst/Architect

Position-Scott AFB, IL

January 2013 to December 2018

- Supply Chain Analytics MITRE Research Project: Controlled a three-year research initiative to employ innovative data science techniques on diverse transportation data streams, aiming to detect fraudulent activities. Designed and oversaw the R-based implementation of an anomaly detection algorithm that identified historically reported fraud schemes. Conducted data discovery using Teradata & Oracle, and developed data pipelines with Teradata BTEQs, Python, Pentaho Kettle Community, and Talend Open Studio. Used MySQL Workbench Data Modeler to model data.

- Single Source Distribution Data Analytic (S2DA) Environment: Devised potential courses of action to rectify inconsistencies in metrics and analytics. The aim was to establish a unified analytic logistics environment catering to the US Distribution & Logistic Agency, USTRANSCOM, and the Army Materiel Command.

- Transportation Control Movement Document (TCMD) Process Analysis: Led a team to probe enterprise wide TCMD metrics, which indicated prevalent issues with advance shipment notices to ports. Pinpointed primary issues in data flow, enhancing the quality of information and instigating system modifications, including user input adjustments and automated data sharing processes. Detected and addressed inaccuracies in the TCMD

Shipper Performance Metrics that erroneously depicted subpar shipper performance.

Enterprise Data Architect

Position-Scott AFB, IL

January 1999 to December 2013

- US Transportation Command (USTRANSCOM) Enterprise Data Management Strategy: Headed the overhaul of data management policies and practices to align with agile development.

- Agile Transportation for 21 st Century (AT21) Agile Team Technical Lead: Agile Technical Lead on several AT21 agile application deliveries, enabling enhanced analysis of transportation methods for global delivery of personnel and supplies.

- Air Mobility Command (AMC) Enterprise Data Management Strategy: Pivoted enterprise data management strategies to ensure alignment with primary system implementations. This move catalyzed a jump from under 10% adherence to enterprise standards in 1999 to an impressive 70% compliance by 2006. The high degree of compliance across mission execution systems eased the implementation of enterprise corporate services.

- AMC Data Architecture Implementation: Directed a robust technical team of 12 data modelers and engineers, focusing on documenting and standardizing data within the client's core mission execution systems. All models were created and maintained in ERWin.

- Conference Participation: Presented and published the AMC Data Management Strategy at 2001 Armed Forces Communications and Electronics

Association (AFCEA) conference in San Diego and presented to the USAF Scientific Advisory Board (SAB) at Eglin Air Force Base. Attended Data

Management Association (DAMA) conference from 2003 through 2007 to keep abreast of topics in relational and dimensional modeling.

Education

Master of Science in Computer Science

Missouri University of Science & Technology - Rolla, MO

Master of Business Administration in Business Administration

Fontbonne University - St. Louis, MO

Bachelor of Science in Electrical Engineering

Southern Illinois University - Edwardsville, IL

Skills

- Responsibilities: Develop Python automation scripts, manage Terraform HCL files for provisioning infrastructure, and support Watsco's proprietary ETL tool in AWS.
MITRE DEC 1999
- SQL
- Jira
- Git
- Data Warehouse
- Agile
- Azure
- Java
- APIs
- Tableau
- Python
- AWS
- Spark
- Data modeling
- Azure Data Lake
- Databricks
- Snowflake
- Power BI
- Kafka
- Analytics
- Communication skills
- Teradata
- Business intelligence
- Cloud computing
- Data visualization
- Apache
- SDLC
- Scripting
- Supervising experience
- Workday

Links

<https://www.linkedin.com/in/jimsoehlke>

Certifications and Licenses

CompTIA Security+

ITIL Certification

APICS Certification