Shiv Pratap Singh Tomar

Phone: +1-917-379-3242   LinkedIn: www.linkedin.com/in/shivtomar
Email: shivtomarwork@gmail.com   Website: https://shiv-tomar.com/

EDUCATION

– **Stevens Institute of Technology, Hoboken, New Jersey**   **Aug 2018 - Dec 2019**
   Master of Science in Information Systems, GPA: 3.92/4.00
   - Graduate Certificate in Business Intelligence and Analytics
– **NMIMS University, Mumbai, India**   **Aug 2014 - May 2018**
   Bachelor of Technology in Computer Science

SKILLS

- **Programming**: Python, SQL (Expert-Level), Bash, R. HTML, CSS
- **BI & Analytics Tools**: Tableau, Power BI, Looker, Google Analytics, Advanced Excel, Google Sheets
- **Databases**: Snowflake, MySQL, PostgreSQL, MongoDB, Google BigQuery
- **Cloud & Big Data**: AWS (S3, EMR, Glue, DynamoDB, Lambda), Apache Spark, Databricks, Hadoop
- **Platforms**: Windows, Linux, MacOS
- **DevOps & Workflow Tools**: Apache Airflow, GitHub, JIRA, Jenkins, Datadog, Confluence, PagerDuty

EXPERIENCE

Pycube Inc., Sterling, VA   **Nov 2024 - Present**
Lead Engineer – Prototyping & Full-Stack Applications
- Leading a team of 3 to develop rapid full-stack prototypes using **CursorAI, Swift (iOS)**, and web technologies for client demos and pilots.
- Built production-ready solutions for **healthcare and logistics**, integrating frontend, backend, and **AI-based components**.
- Integrated **API services, backend automation, and data workflows**, enabling real-time application logic and client-ready performance.

Capital One Bank, Mclean, Virginia   **Nov 2021 – Nov 2024**
Data Engineer (Python, AWS, Airflow, Jenkins, SQL, Spark, Snowflake, Docker)
- Orchestrated end-to-end data pipelines using **Apache Airflow DAGs**, enabling data scientists to successfully run and manage ML feature workflows.
- Processed and loaded datasets from **Snowflake, S3, and DynamoDB**, with outputs written as **Parquet files to OneLake** for machine learning model consumption.
- Utilized **Postman and cURL-based API calls** to trigger Airflow workflows, dynamically create EMR clusters, and manage feature runs across **AWS East and West regions**.
- Maintained modular **GitHub repositories** for Spark jobs, orchestration logic, and EMR configurations using **Spark Submit**.
- Deployed and managed pipelines with **Jenkins**, transitioning from manual processes to a standardized managed deployment model via internal **bogie files**.
- Built TTL-based cleanup and **pagination logic in DynamoDB** to manage feature table lifecycle and query efficiency.
- Led efforts to **scale EMR clusters**, optimize memory and compute utilization, and reduce overall pipeline costs.
- Documented architecture, workflows, and best practices in **Confluence**, and participated in engineering discussions to improve framework speed and efficiency.
- Provided **24/7** production support, led cross-team triage, and enhanced observability by migrating from **Datadog** to **New Relic** and implementing **Splunk** loggers.

Pycube Inc., Sterling, Virginia   **Jun 2020 - Nov 2021**
Business Data Analyst (Python, AWS, Tableau, SQL, Excel, Management)
- Analyzed healthcare supply chain data using **Python**, **SQL**, and **Excel**, driving a 20% boost in operational efficiency and identifying $1M+ in cost-saving opportunities.
- Built interactive **Tableau dashboards** to track KPIs and streamline inventory audits, achieving 100% item count accuracy.

PROJECTS

- **Local Dockerized Airflow for QA Acceleration**: Set up a local Dockerized Airflow environment to enable local testing, reducing QA time by **60%** and removing reliance on external teams.
- **EMR Log Retrieval Automation**: Built an Airflow-integrated **S3 Sensor** and automated EMR log collection, cutting debugging time by **70%**.
- **New Relic Monitoring & Alerting**: Created a custom monitoring class in Python for **New Relic**, integrating alerts with **Slack** and **PagerDuty** to streamline on-call workflows and improve observability.