

# Vivek Anupindi Data Engineer

[vivekanupindi@gmail.com](mailto:vivekanupindi@gmail.com) | 314-327-1714 | MO, USA | [LinkedIn](#)

## Summary

Detailed-oriented Data Engineer and Analyst with 4 years of experience across cloud data platforms, including IBM and Deloitte. Proficient in building scalable ETL pipelines using Python, Apache Airflow, and Snowflake, with expertise in data quality (Great Expectations), SQL optimization, and Power BI reporting. At IBM, modernized DataStage on Cloud Pak for Data with containerization and CI/CD, reducing deployment time by 60% and improving hybrid cloud scalability. Recently contributed to Deloitte's Cloud DataOps accelerator and completed AWS Cloud Practitioner and dbt certifications. Strong track record in delivering high-performance, cost-effective data solutions across enterprise-scale, hybrid-cloud environments.

## Technical Skills

- Data Analysis & Processing:** SQL (JOIN, Subqueries), Python (OOPs, Pandas, NumPy, Datetime, Regex), Exploratory Data Analysis (EDA)
- ETL & Data Engineering:** Apache Airflow, Apache Kafka, IBM DataStage (Containerized ETL), Azure Data Factory, Azure Blob Storage, Hadoop, Cassandra DB, MongoDB, SQL & NoSQL, Data Lake Ingestion, Remote Engine Patterns, Data Virtualization
- Big Data Ecosystem:** HDFS, Kafka Pipelines, Azure Big Data Stack, Data Governance (Watson Knowledge Catalog, Lineage, GDPR), Data Modeling, Hybrid Cloud Integration
- Data Validation & Quality:** Automated Quality Checks, Schema Validation, Watson Knowledge Catalog (Lineage & Audit), Metadata Management
- Testing Expertise:** CI/CD Pipeline Testing, A/B Testing, Production & Regression Testing, Load & Integration Testing
- Cloud Platforms:** Microsoft Azure, IBM Cloud Pak for Data, OpenShift, Kubernetes, Hybrid Cloud Environments
- Visualization & Reporting:** Power BI (DAX, drill-through, slicers), Tableau (LOD, dynamic filters, parameters), Python Reports (Matplotlib, Seaborn)
- Operating System:** Linux/Unix (bash scripting, basic command line tasks), Windows

## Work Experience

### Data Engineer Intern, Deloitte

01/2025 – Present | Remote, USA

- Built Airflow and Python-based ETL pipelines to ingest 2M+ daily API records into Snowflake with 98% uptime, eliminating manual intervention using REST APIs and parameterized DAGs.
- Optimized SQL and dbt transformations in Snowflake; reduced Power BI dashboard query time by 35% across 5+ views using refactored joins, filters, and materialized models.
- Automated 15+ Great Expectations checks integrated with Snowflake and S3, boosting SLA compliance to 99.4% and reducing data defect tickets by 60%.
- Designed YAML-based, config-driven pipeline templates using Jinja2 for Deloitte's Cloud DataOps Accelerator, reducing code duplication by 40% and onboarding 3+ new projects efficiently.
- Applied AWS Cloud Practitioner and dbt Fundamentals skills to develop modular dbt models; improved Snowflake lineage tracking and enabled better team collaboration through Git and dbt Cloud.

### Data Engineer, IBM

01/2020 – 05/2023 | Bangalore, India

- Modernized DataStage using Docker, OpenShift, and Cloud Pak for Data; reduced deployment time by 60% and improved provisioning speed 3x across hybrid cloud environments.
- Integrated IBM Data Virtualization for multi-source queries; reduced ETL latency by 35% and data movement by 45% across PostgreSQL, Oracle, and NoSQL sources.
- Designed remote engine and Kubernetes-based workload balancing; increased job throughput by 40% and reduced compute costs by 25% across cloud-native and hybrid ETL jobs.
- Enabled end-to-end data lineage using Watson Knowledge Catalog; automated Great Expectations checks and improved audit readiness by 70% across 85% of governed pipelines.
- Built CI/CD pipelines using Jenkins, Git, OpenShift Pipelines, and ArgoCD; accelerated release frequency by 3x and eliminated 90% of manual deployment tasks.
- Resolved integration blockers 30% faster by collaborating with architects and DevOps; improved delivery speed by 40% across 5 enterprise client deployments.

## Projects

- Integration of ipass with AD**
  - Automated group membership provisioning for new joiners using iPass workflows integrated with Active Directory.
  - Streamlined IAM operations by enabling role-based access provisioning through workflow-driven automation.
- Smart Health Monitoring System using Python**
  - Built a Python app to track vitals (heart rate, BP, oxygen, temperature, sugar) and flag anomalies.
  - Generated basic health reports for continuous monitoring and early alerts.

## Education

### Master of Science in Information Systems

08/2023 – 05/2025 | Missouri, USA

Saint Louis University

### Bachelor of Technology in Computer Science

2016-2020 | Hyderabad, India

B. V. Raju Institute of Technology

## Certifications

---

- **IBM Cloud Core certification**  
Gained foundational expertise in cloud computing, including IaaS, PaaS, and DevOps principles using IBM Cloud services.
- **IBM Python Certification**  
Demonstrated proficiency in Python programming for data manipulation, control flow, and basic application development.
- **Microsoft Azure fundamentals**  
Developed understanding of Azure services, cloud architecture, pricing, and security principles aligned with enterprise cloud adoption.
- **DataCamp Power BI Analyst**  
Completed 10-course track focused on data visualization, DAX functions, and building interactive dashboards with Power BI.
- **Data bricks fundamental's accreditation**  
Certification covers ETL pipelines in Databricks, Azure integration, and data processing using Apache Spark within the Databricks environment.