

Bishwajit Prasad Singh

Data Engineer | Data Analytics | Cloud Platforms (Azure, AWS) | Python | SQL

📞 +91 8882218924 ✉ bishwajit.1804@gmail.com in <https://linkedin.com/in/bishwajitsingh>
GitHub: <https://github.com/bishwajitSingh123> | Portfolio: bishwajitsingh123.github.io

Professional Summary

Aspiring Data Engineer with a strong foundation in Python, SQL, and cloud platforms (Azure, AWS), backed by hands-on experience designing scalable ETL workflows, real-time data pipelines, and analytics-ready data lakehouse architectures. Proficient in Spark, Pandas, and PySpark for data transformation, with practical knowledge of data modeling, normalization, and schema design. Demonstrated ability to ingest and process structured and semi-structured data from APIs, flat files, and relational databases. Passionate about solving complex data challenges, optimizing data systems for performance and reliability, and contributing to impactful analytics solutions that empower decision-making at scale.

Technical Skills

- Programming & Scripting:** Python (Pandas, NumPy, OOP, DSA), PySpark (RDDs & DataFrames), Shell Scripting
- Databases & SQL:** Joins, Window Functions, Stored Procedures, Data Modeling (ER Diagrams, Normalization), MySQL, PostgreSQL
- Data Engineering:** ETL/ELT (Batch & Streaming), Data Warehousing Concepts, Delta Lake (Bronze-Silver-Gold), REST API Ingestion, JSON/Parquet, Airflow (Basics), Docker (Basics)
- Big Data & Processing:** Apache Spark (Structured Streaming, DataFrames, RDDs)
- Cloud Platforms:** Azure (ADF, Databricks, Event Hub, Stream Analytics, Data Lake Gen2, Key Vault), AWS (S3, EMR, Glue, Redshift, RDS), Exposure to GCP
- Visualization & Reporting:** Power BI, Matplotlib, Seaborn
- AI/ML & NLP Tools:** Scikit-learn, SpaCy, NLTK, TF-IDF, Cosine Similarity, Model Deployment (Flask, Streamlit)
- Collaboration & Version Control:** Git, GitHub, Jira, Confluence, Postman, VS Code, Jupyter Notebook
- Mathematics & Statistics:** Probability, Linear Algebra, Statistical Analysis, Machine Learning Basics

Work Experience

Apr 2024 - Oct 2024 | Cygnus Advertising, Kolkata

PHP Developer (Client Projects)

- Designed and maintained data-driven web applications using PHP, MySQL, and JavaScript.
- Optimized SQL queries & CRUD operations for improved performance.
- Delivered projects for IIT Kharagpur, IIM Amritsar, and CNCI Hospital, contributing to data collection, transformation, and reporting workflows.
- Engaged in requirement gathering, demos, and delivery alignment.

Education & Certifications

- B.Tech in Computer Science Engineering** – NSHM Knowledge Campus, Durgapur, WB, 2019-2023 | CGPA: 8.4

CERTIFICATIONS:

- Microsoft Certified: Azure Fundamentals
- Microsoft Certified: Azure Data Fundamentals
- Microsoft Certified: Azure Data Engineer Associate (*In Progress*)

Projects

AI Resume Matcher Suite | [GitHub](#):

Technologies: Python, Streamlit, TF-IDF, SpaCy, Azure Databricks, Cosine Similarity

- Applied NLP techniques to analyze unstructured resume data, Developed 2 intelligent NLP-based tools to match resumes with job descriptions using TF-IDF and cosine similarity.
 - Extracted top resume skills with over 90% accuracy, added filters (match %, location), and export options.
- Reduced manual screening workload by 70% through automated matching engine.

Olympics Data Lakehouse | [GitHub](#):

Technologies: Azure Data Factory, Azure Databricks, PySpark, Delta Lake, CI/CD

- Designed scalable Lakehouse on Azure (ADF, Databricks, PySpark, Delta Lake).
- Integrated APIs and flat files into Delta Lake, ensuring consistency and reliability.
- Automated ETL with parameterized pipelines, Deployed SaaS-ready Gold layer via CI/CD & optimized Delta for seamless incremental updates.

NYC Yellow Taxi Streaming Pipeline | [GitHub](#):

Technologies: Azure Event Hub, Stream Analytics, Azure Databricks, Delta Lake

- Built real-time streaming Lakehouse with Event Hub, Stream Analytics & Databricks.
- Designed Spark Structured Streaming for ingestion → Bronze → Silver → Gold, Ensured data quality, low latency & analytics-ready dashboards.

Personal Details

- [Languages: "English, Hindi and Maithali", Address: "Madhubani, Bihar – 847108"]**