

# RAKESH KUMAR SINGH

DNACLOUD -SENIOR ASSOCIATE CONSULTANT – INFOSYS LIMITED

[rakeshkumar.singh05@ad.infosys.com](mailto:rakeshkumar.singh05@ad.infosys.com)

## WORK EXPERIENCE

---

- Built and maintained data pipelines using TPA, SoR, USPS, etc., to ADLS Data Lake, handling over 1 TB of data daily.
- Enhanced efficiency by debugging production notebooks and streamlining workflow processes, reducing runtime by 35%.
- Authored comprehensive documentation for data pipelines and ETL processes, improving on-boarding efficiency by 20%.
- Optimized SQL queries for performance, resulting in a 50% reduction in data retrieval times.
- Engineered Intelligent Pipelines for parallel data loading with periodic scheduling.
- Re-constructed 250+ data pipelines with comprehensive functionality, including triggers and integration runtimes.
- Merged multiple Azure Data Pipelines, improving efficiency by 40%.
- Debugged many Databricks notebooks, resolving issues and fixing failures.
- Managed multiple 50 CRs for live reports, ensuring modifications were accurate, achieving a 99% accuracy rate.
- Collaborated with various development team for follow-up tasks, ensuring error-free reports and reducing post-deployment issues by 40%.
- Collaborated with Burberry business users to secure official approval, ensuring seamless production deployment.
- Conducted rigorous Unit and Regression Testing to ensure robustness and reliability.
- Made an EDA Dashboard with interactive graphs and ML algorithms using Python (Plotly, Streamlit, Seaborn) and was capable of generating detailed report within 5 seconds.
- Integrated few algorithms such as regression, decision trees, and random forest models for enhanced analytics with an accuracy of 92%.
- Implemented data cleaning and validation techniques to ensure accuracy and reliability.
- The dashboard was officially launched internally for sharing with colleagues for internal purposes.
- Designed, developed, maintained, and deployed more than 25 websites and applications, managing both front-end and back-end tasks.

## TECHNICAL SKILLS

---

- **Softwares:** Anaconda, Power BI, MS Excel, Visual Studio
- **Cloud Platforms:** Azure Data Factory, Azure Synapse, Azure SQL
- **Management Tools:** Confluence, Jira
- **Programming Languages:** SQL, Python, Pyspark, Spark, Mongo DB, Scala, Hadoop, HTML, CSS, PHP, Bootstrap
- **Data Visualization:** Plotly, Streamlit, Seaborn
- **Big Data Technologies:** Databricks, Hadoop, Spark, ADLS

## CERTIFICATES

---

- **Introduction To Big Data** - Learned about the terminologies and fundamental concepts underlying big data challenges, applications, and systems. ([See Certificate](#))
- **Associate Data Analytics** - Learned how to extract a useful result out of raw data with the use of Python Libraries, Excel, Tableau and R Programming. ([See Certificate](#))
- **Azure Fundamentals (AZ-900)** - Microsoft Certification on Azure Fundamentals certification which confirms my grasp of essential Azure concepts and services. ([See Certificate](#))
- **IBM Big Data** - Learned about Big Data tools including Hadoop, Hive, HBase, and Pig to manage and analyze large datasets, enhancing my capabilities in the world of data processing and analytics. ([See Certificate](#))
- **Data Engineer Capgemini Certified** - A professional certified in data engineering, skilled in designing, building, and maintaining data infrastructure and pipelines. ([See Certificate](#))