Prashant Dixit

Senior Data Engineer with expertise in ETL & Data Analytics.

Gurugram, India | Born 23/07/1992 | +91 9555541393
prshntdxt7@gmail.com | **LinkedIn** | **Github**
I'm an Indian citizen and my notice period is 60 days.

TECHNICAL SKILLS

Data Engineering (8 years of project experience) | Azure (4) | Azure Data Factory (2) | Azure Synapse (2) | Azure Kubernetes Service (2) | Azure Key Vault (4) | ADLS gen2 (4) | Azure AI Document Intelligence (1) | GCP (3) | BigQuery (3) | Dataproc (3) | Dataflow (2) | Snowflake (4) | Databricks (3) | Data Lake (3) | Delta Lake (3) | Apache Spark (3) | Apache Hadoop (2) | Hive (2) | SQL (5) | PostgreSQL (2) | Python (7) | Pandas (5) | Airflow (7) | Power BI (1) | Docker (2) | Data Warehousing (5) | Data Modelling (3) | ETL pipelines (6) | MongoDB (1) | STTM (3) | GitLab (7) | CI/CD (5) | Bitbucket (2) | Rally (2) | JIRA (3) | Agile/Scrum (5)

WORK EXPERIENCE

09/2020 – today

Senior Data Engineer
Deloitte Consulting India Private Limited, Gurugram, India
Project: Clinical Trials Data Platform (**CTDP**) - Centralized reporting platform for US-American pharma company **CVS Health** to manage 9K+ stores across the US.
- Developing the complete end-to-end data architecture using Azure, Snowflake, and Airflow providing a single touchpoint for business users to generate store performance reports.
- Built the ETL data pipelines Built a Python Audit Control Balance ETL framework to load data from multiple sources into a centralized data warehouse.
- Implemented Airflow on Azure Kubernetes Service, which increased completion time by 30%.
- Improved the decision-making of the management board process by building the dashboard for real-time data, which increased revenue by 35%. Developing the visualization of the curated data using Power BI.
- Performed Snowflake data model optimizations that improved the query performance by 40%. Implemented Azure Key vaults for better security and control.
Technologies used: Azure Data Lake | Snowflake | Apache Airflow | Python | SQL | Azure Key Vault | Azure Container Registry | Docker | Power BI.

Project: Platform Migration - Enterprise cloud data warehouse for US-based cybersecurity multinational **Palo Alto Networks**, with over 85k+ customers and reports published to 20+ vendors across the US.
- Led the architecture and design of the data platform using Azure, Azure Data Factory, Azure Synapse, SQL, and Python.
- Based the architecture on Lake House using Azure Data Lake and Azure Synapse, following best practices, designed and built the architecture of new big data pipelines using Spark, SQL, and Databricks.
- Implemented medallion architecture on Delta Lake which helped to maintain data quality and consistency..
- Developed 15 efficient and reliable data pipelines, using Azure Data Factory and Azure Synapse, for different stakeholders. Utilized Python

and PySpark to perform data cleansing, validation covering 90+% of the data, and enrichment tasks on the Data Lake.
- Received the applause award for my contribution to this project.

Technologies used: Databricks | PySpark | Lakehouse | Azure Data Factory | Azure Synapse | SQL | Azure Data Lake.

03/2018 – 08/2020	**Big Data Developer**

Infosys Technologies Private Limited, Pune, India
Project: Optimum Insights - Reporting and Analytics platform for **Altice USA,** serving nearly 5M customers across 21 states.
- Worked in an agile team of 3 (10 developers in the project) to modernize the analytics and data platform by migrating from a legacy Hadoop cluster running Spark to Google Cloud.
- Developed fully managed and highly scalable Dataproc pipelines to reduce infrastructure cost around $4M annually. The transition also boosted data processing speeds by 50%.
- Achieved high levels of automation in Composer Workflows, reducing manual intervention by 70%. Enhanced data reliability and governance, ensuring secure and compliant data operations.

Technologies used: Google Cloud Platform GCP | BigQuery | Dataproc | Composer | Cloud Storage | Cloud Functions | Cloud Logging | Hadoop | Hive | PySpark | Python | PostgreSQL.

03/2016 – 02/2018 **Programmer Analyst**
Cognizant Technology Solutions India Private Limited, Pune, India
Project: Relay Cards - KPI analysis for a leading private banking corporation **Bank Julius Baer & Co. Ltd.**, based in Zürich, Switzerland.
- Worked in an agile team of 3 (6 developers were on the project) to develop the data platform using Apache Hadoop, Spark and Airflow that helped to create BI dashboards for better decision-making.
- Re-engineered 20+ Hive scripts and UDFs to PySpark pipelines, reduced data processing speeds. Implemented DistCp transfers reducing data transfer time by 50%.

Technologies used: Apache Hadoop | Apache Spark | Apache Hive | Apache Oozie | Python | SQL | Linux | Shell Scripting.

EDUCATION

07/2011 – 07/2015   Bachelors in Electronics and Communication Engineering
Galgotias University, Greater Noida, India

CERTIFICATIONS

Azure Solutions Architect Expert | Designing Microsoft Azure Infrastructure Solutions (AZ-305) | Azure Administrator Associate (AZ-104) | Azure Data Engineer Associate (DP-203) | SnowPro Advanced Data Engineer (DEA-C01) | SnowPro Core (COF-C02) | Microsoft Technology Associate (MTA): Python (98-381)

LANGUAGE SKILLS

English (C2) | Hindi (Mother tongue)

HOBBIES

Poker | Swimming