

DEEPAK PANDEY

Gurugram, India

📞 (+91) 82188-65360 ✉ d.pandey11585@gmail.com [in linkedin.com/in/d-pandey](https://www.linkedin.com/in/d-pandey) github.com/freakypandit

EXPERIENCE

Sr. Associate - Data Engineer

February 2023 - Present

American Express

Gurugram, India

- Spearheaded the migration of Teradata-based reports to big data ETL using HiveQL and Pyspark (spark SQL), enabling 12 engineers to migrate 100% of the inventory before the target time.
- Architected a JIRA time tracking dashboard using Python, JIRA Rest API, SQL, and Tableau to enhance visibility and oversight of crucial agile metrics.
- Enhanced Hive job efficiency by optimizing Hadoop cluster metrics, and reducing iNode usage in the Linux namespace container bringing down disk I/O by 40%.
- Leading a team of 4 data engineers, and collaborating with leaders to define the roadmap, performance, and priorities for the team.

Associate - Data Engineer

August 2021 - January 2023

American Express

Gurugram, India

- Directed the design of the next-generation analytical product framework, resulting in a 40% improvement in data quality and the solutions capable of handling >99% of engagements.
- Improved the execution time of existing HiveQL ETL pipelines by 35%, analyzing over 20k+ lines of code (LOC), and optimizing the queries, reducing 4k+ LOC in the process.
- Achieved a 95% reduction in time and human effort by automating over 20 Excel leadership reports through Hadoop, HiveQL, and SSIS ETL pipelines integrated with SSAS OLAP cubes for pivot-based reporting.
- Created MIS reports on the \$200B+ USA personal loans market using Experian, AWS EMR, and SSIS, resulting in timely and accurate delivery of key insights on 17+ companies to strategy teams.

Analyst - Intern

January 2021 - June 2021

American Express

Gurugram, India

- Engineered a micro-segment insights engine using Python and PySpark to identify the top 1% high-risk groups from over 1 million segment permutations.
- Eliminated over 100 hours of monthly manual effort by engineering data pipelines with Bash, Python, Hive, and an internal orchestration tool for Amex risk team leadership.

EDUCATION

Bachelor of Engineering in Information Technology

July 2017 - June 2021

Panjab University

Chandigarh, India

Coursework: Computer Programming, Data Structure and Algorithms, Database Management Systems (DBMS), Operating Systems

TECHNICAL SKILLS

Languages/Tools/Systems: Python, SQL, Bash, ETL (Hive, SSIS), Git

Technologies/Frameworks: Hadoop, Apache PySpark, Linux, Apache Airflow, GitHub, GCP(BigQuery, Compute Engine), Docker

Soft Skills: Decision Making, Stakeholder Management, Ownership, Mentorship, Technical Communication

PROJECTS

Artha.io | React JS, Hardhat, JSON RPC

May 2022

- Conceptualized and developed a web application allowing digital brands to mint unique NFTs and create DApps with a one-click app deployment feature. Link: Artha.io

Forgery Detection in Digital Images | Python, Streamlit, SQL, Pytorch

September 2020

- Developed an image classification algorithm using deep transfer learning and image error level analysis, achieving over 95% accuracy on real-world datasets.

CERTIFICATIONS

Harvard Leadership Edge

December 2022

Issued By Harvard Business Publishing Corporate Learning

Statistical Analysis and Machine Learning

March 2020

Issued by WorldQuant University