

HIMANSHU PANT *Software Engineer | Backend & Data Systems*

✉ hpant.data@gmail.com 📞 6239994043 📍 USA 🔗 [Portfolio](#) [LinkedIn](#) [Github](#)

EXPERIENCE

Backend & AI Developer (Intern), MyEdMaster 01/2026 – Present | Virginia, US

- Building a multi-agent **Agentic RAG** system that delivers deeply personalized legal guidance for users who cannot afford traditional counsel
- Designed query refinement pipeline using vector databases and specialized agents that collaborate independently to retrieve, reason, and respond with high accuracy
- Creating backend infrastructure powering an interactive interface that helps users understand their legal options and make informed decisions

Consultant - II, Data Engineer, EXL Services Pvt. Ltd. 07/2023 – 03/2024 | Remote

- Worked for a US based Insurance Client. Eliminated a critical **90-min bottleneck** in insurance data processing by redesigning PySpark ETL pipelines on AWS EMR — cutting runtime by **66%** through **query optimization** and partitioning strategies.
- Established a governed data quality layer across 15+ insurance domain sources using Great Expectations and Python unittest, flagging anomalies before data reached production.
- Accelerated deployment cycles by **40%** by engineering CI/CD pipelines in Jenkins, automating Spark job delivery to AWS Glue and EMR

Data Engineer, Super Six Sports Ltd 08/2022 – 07/2023 | Gurgaon, India

- Built end-to-end ML data infrastructure from scratch, encompassing data modeling, feature engineering, ETL pipelines, and validation frameworks
- Built and owned the core data ingestion layer — ETL pipelines across **10+ external APIs** loading **500K+** daily records into AWS Redshift, forming the analytics foundation for the platform.
- Drove a **20% reduction** in user churn by translating raw behavioral data into a production ML pipeline (**78% accuracy**), enabling automated SQL/Python retention campaigns.
- Eliminated manual data intervention via **automated** Python scheduling workflows, achieving 24/7 pipeline availability

Associate Data Engineer, FutureSense Technologies Pvt. Ltd. 10/2021 – 07/2022 | Bengaluru, India

- Spearheaded a full-scale ETL modernisation — migrating **50+ legacy SAS batch jobs** to Apache Spark on Azure Databricks, transforming manual workflows into governed, cloud-native data processes.
- Cut critical batch processing time by **86%** (6+ hrs → under 50 mins) via partitioning, caching, broadcast joins, and data skew resolution.
- Maintained **99.5% data accuracy** post-migration with PySpark/SQL quality validation frameworks, preserving client trust across all production datasets.

Data Analyst, Koron Projects Pvt Ltd 10/2018 – 07/2021 | India

- Built 15 Power BI dashboards providing real-time visibility across \$50M+ in annual construction projects
- Automated monthly reporting, replacing manual Excel workflows and saving 20 hours/month for finance stakeholders
- Designed KPI dashboards tracking 50+ active projects, enabling leadership to identify delays and make faster decisions

EDUCATION

M.S. Software Engineering, Arizona State University 08/2024 – 05/2026 | Tempe, AZ

B.Tech., Computer Science Engineering, Guru Gobind Singh Indraprastha University 08/2014 – 06/2018 | New Delhi

SKILLS

Databases

MSSQL, Snowflake, PostgreSQL, MySQL, MongoDB, AWS Athena, Google Spanner, Redshift, Graph Databases

Languages

Python, SQL, PySpark, Java, Shell Scripting

ML & QA

TensorFlow, PyTorch, Scikit-learn, XGBoost, Great Expectations, Pytest, unittest, DBT

Big Data

Apache Spark, Databricks, Fabric, AWS EMR, Azure Databricks, Pandas, NumPy

Cloud Platform

AWS (S3, Glue, EMR, Athena, Lambda, SageMaker, EC2), Azure (ADF), Fabric, Kinesis

PROJECTS

Serverless Face Recognition Pipeline 08/2025 – 12/2025

Implemented a serverless, multistage face recognition pipeline using **AWS Lambda, SQS, and ECR** to process video frames from IoT-like clients. Containerized Lambda functions with Docker and Amazon ECR, bundling PyTorch, facenetpytorch, and OpenCV to perform MTCNNbased face detection and Inception Resnet V1based face recognition

[MeetFlow — DEVLABS Hackathon, 1st Place · Mar 2026](#) 🔗

Built an LLM-powered pipeline that analyzes meeting transcripts, extracts actionable tasks, checks team capacity via Taiga, and reassigns overloaded members automatically → Integrated end-to-end orchestration: GPT-4o-mini analysis → capacity engine → Taiga ticket creation → Slack team notification

[FairCharge — HackASU Claude AI Builder Hackathon · Mar 2026](#) 🔗

Built a medical bill audit pipeline that parses bills via Claude Vision, benchmarks charges against CMS Medicare pricing data (19K+ CPT codes), and generates ready-to-send dispute letters. Implemented hybrid matching using SQLite for exact CPT lookups and SapBERT + ChromaDB for fuzzy biomedical entity linking to handle OCR errors and abbreviations

AWARDS

DEVLABS Hackathon — 1st Place · Mar 2026

SPOT Recognition Award · Super Six Sports Gaming, 2023