

Anuj Kumar

📍 Boston, MA | ✉ kumar.anuj@northeastern.edu | 🔗 LinkedIn | 🐙 Github | 📁 Portfolio

PROFESSIONAL EXPERIENCE

Sacramento Kings, Data Engineering Intern - Sacramento, CA **January 2024 - April 2024**

- Automated post-game report testing using R and SQL data pipelines, ensuring data accuracy, and reducing report generation time by 50%
- Engineered R unit testing and linting with GitHub Actions, increasing code coverage and streamlining deployment
- Automated AWS EC2 operations using autoscaling (CloudWatch, ALB), lowering costs and maintaining 99.5% up time during peak loads

Barabasi Lab, Northeastern University, Graduate Research Assistant- Boston, MA **June 2023 - January 2022**

- Extracted TBs of data from research APIs (Wikidata, Google Ngrams, OpenAlex) for analysis and research
- Automated data acquisition using Python batch processing for efficient handling of large datasets from research APIs
- Leveraged LLaMA LLM with NLP techniques to extract named entities and relationships from network science research papers, boosting analysis efficiency by 50%
- Architected a high-performance star schema on Google BigQuery, increasing query speeds by 20% and empowering data-driven decisions

Shell India Markets, Upstream Operations, Data Engineer- India **February 2022 - July 2022**

- Develop a Marketing Mix Model (MMM) in collaboration with a cross-functional agile team resulting in forecasting marketing budget spends
- Unified 15+ disparate datasets into an Azure Data Lake, reducing data silos by 50% and enabling integration within the master data warehouse
- Implemented a SARIMA (2,1,1) model with seasonal differencing (D=1) to analyze weekly seasonality in historical ad campaign data resulting in a statistically significant 20% increase in conversion rate for the next campaign

Shell India Markets, Downstream Operations, Data Engineer- India **September 2020 - February 2022**

- Migrated key SAP S4 HANA ERP reports and data to Azure Data Lake, MS SQL Server, Databricks and PowerBI, enabling a 30% cost reduction in Shell's Gulf of Mexico platform operations
- Built oil platform safety pipelines in Azure Data Factory & PySpark achieving 20% cost reduction through improved anomaly detection
- Optimized PL/SQL queries optimization through index, partitioning and query plan analysis resulting in a significant 30%-time reduction
- Crafted 10 real-time Power BI dashboards with advanced DAX, boosting KPI visibility and reporting by 30% compared to SAP ERP systems

TECHNICAL SKILLS

Python - R - GCP - PostgreSQL - git - Linux - docker - Kubernetes - Airflow - BigQuery - AWS - Terraform

EDUCATION

| | | |
|--|--------------------------------|--------------------|
| Northeastern University | M.S Information Systems | 2022 - 2024 |
| Visvesvaraya Technological University | B.E. Computer Science | 2016 - 2020 |

ACADEMIC PROJECTS

Meeting Intelligence Application, Github **May 2023**

- Designed and implemented a scalable audio search system using OpenAI Whisper, GPT-3.5 LLM, and Pinecone DB, achieving a 40% increase in transcription efficiency
- Optimized application performance by 30% by engineering a FastAPI backend with Pydantic data validation, JWT authentication, and rate limiting
- Streamlined deployment and data pipelines by containerizing the Streamlit app and FastAPI backend with Docker and implementing Airflow for data orchestration

Cloud Services Application, Github **March 2023**

- Modeled secure and scalable RESTful APIs using Python and FastAPI, leveraging SQLAlchemy for efficient database interaction and data integrity
- Implemented infrastructure as code (IaC) using Terraform, Packer and GitHub Actions for efficient provisioning and deployment of the application on Amazon EC2
- Established a single-node Kubernetes cluster on Amazon EKS for efficient initial testing and evaluating application
- Utilized Boto3 SDK to interact with AWS S3, enabling seamless image management and storage system
- Integrated CloudWatch for an automated monitoring and scaling system, ensuring application health, responsiveness, and cost optimization

RELATED COURSEWORK

| | |
|---|--------------------|
| Network Structure and Cloud Computing, Northeastern University | Spring 2023 |
| Big Data and Intelligence Analytics, Northeastern University | Spring 2023 |
| Data Management and Database Design, Northeastern University | Fall 2023 |