



SUBASH KONAR



subashkonar13@gmail.com | +918249939823 | Bengaluru, India

Education

IGIT, SARANG

Bachelor of Technology (Computer Science & Engineering)

Jun 2008 - May 2012

8.43 CGPA

Skills

- Python
- SQL
- Databricks
- Grafana
- ETL
- Big Query
- Apache Kafka
- Azure Devops
- Cassandra
- Apache Airflow
- dbt
- Kubernetes
- GCP
- Apache Beam
- Langchain
- Data Modelling
- Azure Cloud
- Docker
- Azure Synapse
- Power BI
- Spark Streaming
- A/B testing
- Data Governance
- Large language Model-Open AI
- Scala

Professional Experience

Microsoft

Senior Data Engineer, Cloud Supply Chain and Provisioning Team

Mar 2020 – Present

Bengaluru, India

- Engineered an event streaming solution using PySpark on Azure Synapse, pulling data from various streaming sources, processing 10M+ daily events from Azure Event Hub to Delta Lake, and improving data processing efficiency by 40%
- Innovated a data quality solution with Langchain and OpenAI, automating the generation of Pyspark code from English prompts, resulting in a 60% reduction in data validation time.
- Orchestrated 15+ CI/CD pipelines via Azure DevOps, ensuring seamless deployment across development, testing, and production environments.
- Drove the creation and provisioning of essential Azure resources including Synapse workspaces, Key Vaults, and ADLS using ARM templates.
- Led the implementation of comprehensive logging practices using Azure Log Analytics and Grafana, significantly improving issue tracking and resolution speed, resulting in a 25% decrease in incident response time.
- Collaborated with a cross-functional team to for requirements gathering and optimize the overall data architecture.
- Mentored and managed 8 data engineers, enhancing their skills in Pyspark, and cloud technologies.
- Oversaw sprint planning and project roadmap formulation using the Agile/Scrum methodology.
- Developed and executed Data Models, working with extensive data sets to construct comprehensive data pipelines for both batch and streaming processes using TOGAF.
- Implemented robust Data Lineage and Data Governance protocols leveraging Apache Atlas.
- Led Data Wrangling and Data Cleansing efforts through CDC and SCD II techniques, leading to the removal of 97% of redundant data.

Woolworths

DWH Designer

Jan 2015 – Jan 2020

Sydney, Australia

- Migrated ETL/ELT workflows from RDBMS to Google cloud using NiFi and airflow thereby reducing the cost of operation for clients.
- Spearheaded the development of a comprehensive data processing platform from the ground up. Utilized Scala, Python, Google Dataproc, Apache Beam, and Big Query to manage the seamless migration of 800 databases from SQL Server to Big Query
- Optimized processing speed with Pyspark distributed computing, improving processing time by 80%.
- Managed and supervised a team of 5 and implemented the Ringelmann Effect rule that resulted in a 30% increase in productivity.
- Conducted successful negotiations with stakeholders and successfully onboarded 3 new projects.
- Played a crucial role in building and analyzing data samples used for training machine learning models. Leveraged Google Cloud AutoML for streamlined data preparation, contributing to more accurate and efficient machine learning outcomes.

Tata Consultancy Services

Hadoop Developer

Nov 2012 – Jan 2015

Chennai, India

- Migrated transactional Data to Hadoop on-Prem cluster, leading to a 30% reduction in data processing time and 25% cost savings in storage of 1 TB.
- Built an automated flow for data archival in Apache NiFi, reducing manual intervention by 90% and minimizing data errors.
- Ingested data from RDBMS like Oracle and MySQL, optimizing data transfer with a 15% reduction in latency.
- Built an automation framework using shell scripting for Ingestion, enhancing data workflow efficiency by 20%.
- Developed a web UI for clients using Bootstrap and HTML 5, increasing client satisfaction by providing user-friendly access to data and reports.
- Developed an application in Java to create XSDs and WSDLs as input in an SOA framework, reducing development time by 15%.
- Created RESTful and microservices in Java, improving system interoperability and enabling cross-platform data sharing.
- Developed highly maintainable Hadoop code and followed all best practices regarding coding, resulting in a 25% reduction in code maintenance efforts.

PERSONAL PROJECTS

Udemy

Fraud Detection in Financial Transactions

Dec 2020 – Jan 2021

- Built a fraud detection system for financial transactions using random forests.
- Achieved an 85% accuracy rate in identifying fraudulent transactions.

CERTIFICATIONS

- **Cloudera CCA 175**
- **Azure Data Engineer DP 203**
- **GCP Coursera**
- **Azure Data Scientist Associate DP100**
- **Fabric Analytics Engineer DP600**