Amit
Senior Expert Data Engineer

SUMMARY

- 8+ year experience in building data engineering and analytics products (Big data, BI, and Cloud products) - Expertise in building Artificial intelligence and Machine learning applications. - Extensive design and development experience in AZURE, Google, and AWS Clouds. - Extensive experience in loading and analyzing large datasets with Hadoop framework (Map Reduce, HDFS, PIG and HIVE, Flume, Sqoop, SPARK, Impala), No SQL databases like Cassandra. - Extensive experience in migrating on-premise infrastructure to AWS and GCP clouds. - Intermediate English - Available ASAP

TECHNICAL SKILLS

Main Technical Skills	Apache Hadoop, Kafka, GCP, AWS
Programming Languages	JavaScript, PL, Python, Scala
AI & Machine Learning	artificial intelligence, AWS ML (Amazon Machine learning services), Machine Learning
.NET Platform	Azure
Java Libraries and Tools	JSON
Data Analysis and Visualization Technologies	Apache Hive, Apache Pig, Attunity, AWS Athena, Databricks, Domo, Flume, Hunk, Impala, Map Reduce, Oozie, Presto S3, Snaplogic, Sqoop
Databases & Management Systems / ORM	Apache Hadoop, Apache Hive, AWS Redshift, Cassandra, MySQL, Neteeza, Oracle Database, Snowflake, SQL
Cloud Platforms, Services & Computing	AWS, Azure, GCP
Amazon Web Services	AWS Athena, AWS EMR, AWS Kinesis, AWS ML (Amazon Machine learning services), AWS Quicksight, AWS Redshift, AWS SQS
Azure Cloud Services	Databricks
Google Cloud Platform	Google BigQuery, Google Cloud Pub/Sub
Platforms	Apache Solr
Deployment, CI/CD & Administration	Bamboo
Version Control	BitBucket, Git
Collaboration, Task & Issue Tracking	IBM Rational ClearCase

Message/Queue/Task Brokers	Kafka
Operating Systems	Linux, Windows
Scripting and Command Line Interfaces	*nix Shell Scripts
Logging and Monitoring	Splunk
Other Technical Skills	Cloudera search, Lex, Polly, VSS

PROJECTS

Architect , EMPLOYER– RAIDON CLOUD SOLUTIONS

JAN'19– PRESENT
Responsibility:

- Advanced analytics platform for one of the largest US banks on Google cloud infrastructure using Bigquery / Dataproc (Hive, Spark).
- Analytical pipeline on AWS- EMR, Kinesis, Athena, Amazon-ML, Lex, Polly.
- Data lake design and development for an Australian insurance company based on GCP.
- Data lake strategy and implementation for European AMC based on AWS-EMR (Spark, Hive, Sqoop, Quicksight).
- Contribution to Wipro BDRE- opens source platform for Blockchain analytics.
- Real-time streaming platform on Kafka stream/KSQL.
- Real-time ingestion pipeline with Kafka, Spark streaming.
- Design and development of Snowflake data warehouse and integration with AWS lambda. Development of Stored procedures, tasks, and python client for data ingestion into Snowflake.
- Pyspark design and development using Databricks.
- Stakeholder management.
- Strategic directives for the advanced analytics capabilities.

Technologies: Azure, GCP and AWS, Advanced Analytics (AI and ML), Big data and Hadoop

Senior Data Developer, Lead Online business for Tesco, EMPLOYER– TESCO

MAR'16– DEC'18
Responsibility:

- Design and development of a machine learning platform for the marketplace using Kafka, Spark streaming, Spark ML modules.
- Design and development of batch and real-time analytics systems based on Hive, Spark
  and Kafka, AWS and Azure Cloud.
- Design and development of a visualization layer based on Domo.
- Created Athena pipeline in sync with Hive Metastore to directly read S3 buckets.
- PySpark development on Databricks.
- Migration of PIG scripts from Cloudera hadoop to Databricks pySpark layer.
- Built automation framework for Spark and Hive jobs.
- Developed JSIVE utility to automatically create hive ddl for complex JSON schemas.
- Developed JSON data generator for creating test data on research clusters to help data scientists.

Senior big data developer, EMPLOYER – BLACKROCK INC.

APRL'14– MAR'16

Description: Worked as a senior big data developer in the web product research team at Blackrock in Gurgaon /Bangalore, designing and developing big data applications.
Environment: Hadoop streaming (Python), HIVE, Sqoop, Shell scripting, TWS scheduler, Impala
The purpose of the project is to replace existing RUFF calculations from conventional ETL platforms with Hadoop. This would ensure faster and on-time delivery of Loss and Policy models
to clients.

Responsibility:

- Involved in project design and creation of technical specifications.
- Developed Sqoop based ETL systems to bring data from EDW data warehouse.
- Created Hive tables to store and transform files from ADW data warehouse.
- Written MR steaming jobs in Python.
- Involved in creating TWS workflow to automate data transformation and presentation processes.
- Developed the process for downstream systems using IMPALA.
- Participated in deployment, system testing, UAT.
- Prepared implementation plans for moving a code to production.