ADITYA BANSAL

E-Mail: aditya.bansal1234@gmail.com / **Phone:** +91-9910774870
LinkedIn Id: https://www.linkedin.com/in/aditya -bansal-39606339/
GitHub Id: https://github.com/devilaadi

PROFILE SUMMARY

• Over **9 years** of experience in designing, developing and delivering enterprise-scale Data Engineering and Big Data solutions across **Azure, AWS and GCP** for IT & Services domain

• Strong expertise in translating complex business requirements into scalable architecture, data pipelines, functional specifications and reusable components that drive measurable business value

• Experienced in **Azure Data Engineering ecosystem: Azure Data Factory, Azure Data Lake, Azure Databricks, Azure Synapse Analytics, Azure Functions, Unity Catalog governance**

• Skilled in Big Data processing using **Spark/PySpark, Databricks Delta Lake, Hive, NiFi, HDFS** and other Apache ecosystem tools

• Proven ability in DevOps oriented delivery using Bitbucket / Git, Jenkins, Jira, Confluence and Azure DevOps

• Highly motivated team player and module lead with strong communication, ownership mindset, stakeholder management capability and continuous learning approach

IT SKILLS

• **Apache Hadoop:** Apache Hadoop, Apache Nifi, Apache , Oozie, Hive, HDFS, Kafka and Sqoop
• **Machine Learning:** Linear Regression, Naïve Bayes ,Logistic Regression, Decision Trees and KNN.
• **Languages :** Python ,Scala
• **Spark/PySpark:** Data Frames, Spark SQL , Databricks Delta
• **Azure Cloud :** Azure Data flow , Azure Data Factory , Azure Databricks ,Azure Function , Azure Devops , Azure Syanpse,Cosmo DB,Azure Functions.
• **Cloud Platforms :** Azure, AWS and GCP

CERTIFICATIONS

• **Azure Data Engineer (DP -200 , DP – 201)**
• **AZ - 900** : Microsoft Certified Azure Fundamentals
• **AD2- E551** : Adobe Experience Platform Qualified
• Hortonworks Certified Hadoop Developer (HDPCD) from Hortonworks

ORGANISATIONAL EXPERIENCE

Nov 2023 - Present	Metlife	Unit Manager

Projects Undertaken:

Client: Metlife

Period: Nov 2023 – Present

Technology Azure Databricks , PySpark , Azure DataLake , Databricks Delta lake , SQL, Azure Synapse Analytics, SQL Server, Azure Data Factory , Azure DataFlow ,Azure Functions , Unit Catalog, Cosmos DB

Leading a module with 2–3 data engineers, delivering enterprise data engineering solutions in the Insurance domain (Cases & Enrollment data).

• Designing and implementing end-to-end data pipelines and orchestration using Azure Data Factory.

• Building and operating a multi-layered Data Lake architecture (Raw → Silver → Gold) on Azure Data Lake Gen2 to enable scalable analytical processing and governed consumption.

• Developing batch and near-real-time processing pipelines in Azure Databricks (PySpark) for large-scale data transformation and business-rule application.

• Unity Catalog (listed separately) for metadata/cataloging, access control, lineage and governance across Databricks assets.

• Creating analytical models, materialized views and performance-optimized queries in Azure Synapse for downstream analytics and reporting.

• Designing and implementing Azure Event Hubs streaming architecture and real-time processing pipelines in Databricks to handle streaming case/enrollment events.

• Defining solution and coding architecture standards for Databricks & PySpark projects — modular code patterns, testing, data quality checks and deployment best practices.

• Building CI/CD pipelines and automated deployment flows using Azure DevOps, including environment parameterization and release promotion.

• Active participant and contributor in Agile ceremonies — sprint planning, backlog grooming, estimations, demos and retrospectives; stakeholder collaboration for requirement capture and acceptance criteria.

July 2021 – Oct 2023	Nagarro	Associate Staff Engineer – Cloud and Big Data Engineer

Projects Undertaken:

Client: Multinational Medical Device Company

Period: July 2021 – Present

Technology Azure Databricks , PySpark , Azure DataLake , Databricks Delta lake , SQL, Azure Synapse Analytics , SQL Server, Azure Data Factory , Azure DataFlow , Azure Logic Apps , Azure Functions

Role :

- Building Pipelines using Azure Data Factory , Azure Databricks , Azure Dataflows, SQL Server ,Azure Dataflows, which transform data from the application like LiveConnect , D365FNO , Axapta etc.
- Building Incremental Pipelines for Ingestion and Transformation.
- Helping in Creating Architecture for Data Mesh.
- Creating views and Procedure for exposing Data for downstream applications.
- Gathering requirement from the Business Stakeholders.
- Scheduling and Monitoring Pipeline in Azure Data Factory.
- Creating views and table in Azure Synapse for Analytics Purpose.

Projects Undertaken:

Client: **Multinational Automotive Corporation**

Period: May 20 – July 21

Technology   **Azure Data Platform and Apache Spark**

Role:

- Creating Azure Data Factory Pipeline for ingestion and processing of data
- Contributed in gathering requirements from the client
- Pre Processing and processing of Data using Azure Databricks using Pyspark and Python pandas .
- Insert Processed Data using Azure Databricks and Azure Data Factory into Azure SQL DW.
- Create views in the Azure SQL DW which are used in Power BI .
- Worked in POC for different Analytics Use case for Machine learning and other.
- Worked in Creating Power BI chart for end users.
- Exploring Data sets which can we used for Analytics.

Client: **U.S.A (RailRoad Company)**

Period: Jan'20 – April '20

Technology: **AWS Cloud Data Platform** and Adobe Experience Platform (Customer Data Platform)

Role:

- Ingest Realtime Data from User using AWS Kinesis and AWS S3
- Process Realtime Data using AWS Lambda and dump the Data in AWS S3
- Worked on Processing of Batch data using AWS Glue and dump the Data in AWS S3
- Realtime and Batch data is dumped in AWS S3 which is used in Adobe Experience Platform Tool
- Created Schemas and Table in Adobe Experience Platform Tool
- Created views in Adobe Experience Platform Tool which is used for further Analytics.

Projects Undertaken:

Client :-   **Sapient**

Period: July'19 – Jan '20

Technology: **Google Cloud Platform (GCP)** and other techonlogies

Role:

- Working as Developer Trained in Google Cloud Platform and terraform
- Worked in creating framework in Azure Spark and GCP for processing and ingestion of Data.

- Developed Terraform code for infrastructure provisioning in GCP.
- Implementing Terraform in Azure Cloud and working on Spark Streaming
- Learning Azure , azure data factory and azure data lake gen2 for future needs and doing a POC for the same .
- Worked in different POV for the Sapient like Azure Data Catalog , Azure Synapse vs Deta Lake , Azure Delta Tables.

Since May'18-July 19	NCR Corporation India Pvt. Ltd., Gurgaon as Big Data Engineer

Project Undertaken:

Title:	**Supply Chain & HR Analytics COE**
Period:	May'18 – Present
Technology:	Supply Chain & HR Analytics COE and building data lake & data pipeline for the Supply Chain domain. Building Big Data Platform for Supply Chain using Hortonworks Data Platform. **Role:**

- Developing Big Data Platform using Stg -> Trans -> Hub Layered Architecture on Supply Chain •   Working on:
  - o Different module of Supply Chain and creating Spark Jobs & Hive Scripts for Inventory, Purchase Order, Work in Process
    Modules, and so on o   Automation of Hive Scripts using PySpark o   Oozie
  Workflows & Oozie Coordinators o   Survey Data Analysis by using Spark, Hive & Nifi for
  HR Domain o   Replacement of Kinaxis and implementation of Oracle Cloud Planning
  - o Migration of Race and implementation of Oracle Cloud Planning in Big Data Platform o
    Data grooming in spark
- Creating Data Repository which will help in Machine Learning and Reporting Purpose
- Acted as a member of a team working in automating the end-to-end process of data lifecycle across the layered architecture
- Using Spark & Hive for transformation and working on Hortonworks Data Platform for the same
- Collaborating with Business Users for the Requirement Gathering
- Preparing model for the classification problem, data gathering, cleaning, validation, quality, exploratory data analysis, missing value and outlier treatment
- Performing data pre-processing using NLTK (Natural Language Toolkit)
- Contributing in requirement analysis and solution discussion

Accomplishments:
- Steered efforts in creating Data Pipeline and Data Repository which will help business users using SAFe Agile Framework throughout the life cycle of projects
- Developed an Ingestion Module to ingest data from various sources
- Designed Nifi Workflow and Nifi Custom Processor in Java and working on data pipeline
- Created Machine Learning Model using PySpark and Big Data platforms
- Played a vital role in developing Spark and Hive activities for HR Domain
- Used Naïve Bayes and Logistic Regression for the Machine Learning Model   o   Different phases of Machine Learning, Data Cleaning and Feature Engineering o   POC of a NLP Based Machine Learning Model to predict the remote resolution

Oct'16 – May'18	SDG Global Technology Center, NOIDA as Associate Software Engineer

Projects Undertaken:

Title:	**Company Product – POC (Big Data)**
Period:	Mar'18 – May'18

Technlogies: Key objective of project was to implement Big Data Analysis for internal product and work on Real-time Analytics using Spark, Kafka & Cassandra.

Role:
- Analysed log data from different sockets & files and provided insights about the risk associated with the user

- Contributed in gathering requirementsfrom the client
- Implemented Logs Analysis using Spark Streaming and Kafka
- Used analytics to improve the business processes which would help in security of the applications
- Implemented the solution with the existing product and presented the same to client
- Identified the solution for Machine Learning use case
- Maintained record of analysed data in Cassandra; attained accuracy for the Model

Title: **World Largest American Hospitality Company (Big Data)**

Period: Oct'16 – Feb'18

Description: The purpose of project was to store information and extract meaningful information out of the same. The solution was based on the open source Big Data s/w Hadoop and Spark. The data wouldbe stored in Hadoop File System and processed using Map Reduce, Spark & Hive. This in turn includes getting the structured data of users which was used for giving insights using Hive.

- **Packages:** Mining of data regarding the traveller paths via social media to create relevant tour packages. Like if they see a rise in demand for a traveller path X-Y-A-B. Create Specific Packages for that to boost their customer acquisition & profitability.
- **Customer Acquisition:** Creation of personalized experiences results in greater customer acquisition. It provides trip itineraries, add-on services and discount coupon based on the taste & preference patterns. They also utilized Big Data to create extremely targeted marketing.
- **Add-on Services:** Revenues from add-on services could be boosted through the integration of discount in the services like Customer A has a history of purchasing luxury products from the hotel boutique due to this, their profile was given a Luxury Tag based on which they provide discounts and add-on services. **Role:**
- Developed simple to complex Spark jobs using Scala; contributed in the requirement and analysis phase
- Managed the importing of data from various data sources; performed transformations using Spark, Hive & Map Reduce
- Engaged in collecting the data from different data sources using SQOOP
- Migrated Map Reduce jobs to Spark
- Wrote Hive queries to analyse data in Hive Warehouse using Hive Query Language (HQL)
- Developed Hive and Spark SQL for the business logic
- Transformed structured data using Dataframes and HiveQL

ACADEMIC DETAIL

2016   **B.Tech. (IT)** from HMR Institute of Technology & Management, New Delhi, GGISPU

PERSONAL DETAILS

Date of Birth: 16th July 1994

Mailing Address: J-108 Ground Floor ashok vihar phase -1 new delhi

Languages Known: English and Hindi

Email   Aditya.bansal1234@gmail.com

Phone No.   +91-9910774870