



MANISHA SINGH

Lead Data Engineer



Current Organization: Albert Heijn, Netherlands
Previous Organization: Wipro Limited & TCS

PROFESSIONAL SUMMARY

A passionate, highly motivated and detail-oriented Senior Big Data Engineer with 7.5 years of experience in designing and developing big data solutions by leveraging excellent understanding of both technical and business process skills. Effective communication skills and delivery focused with a proven track record of success using Agile and scrum. Enjoys developing and coaching others and quickly builds good network of working relationships.

"Always willing to go beyond what may be expected to meet the needs of my employee and my Client"

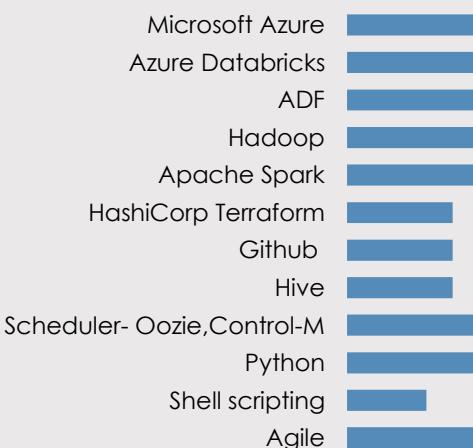
CONTACT

PHONE: +31-685242443
+91-7797770472

LinkedIn [in](#):
<https://www.linkedin.com/in/manisha-singh-78127462/>

EMAIL:
singh.manisha1125@gmail.com

SKILLS



WORK EXPERIENCE

- 7.5 Years of Professional Experience as a Big Data Engineer .
- Hands-On experience in Data Engineering for Batch and Real-time streaming Data.
- Hands on experience in Microsoft Azure - Azure Data Factory, Databricks, Azure Storage.
- Working experience in Pyspark/Spark Scala, Spark Streaming.
- Experience in various Hadoop technologies such as Hive/Hive(UDF), Pig, Sqoop, Oozie ,Cloudera Impala.
- Involved in migration projects.
- Proficient knowledge of SQL.
- Experience in IAAC tool - HashiCorp Terraform
- Experience in various build, Continuous Integration and Code quality check tools such as Jenkins, Git, Gradle, Maven.
- Exposure to major Big data platforms such as Cloudera, MapR.
- Involved in designing technical architecture ,data Integration and processing data pipeline.
- Worked on Shell scripts ,Python programs
- Exposure of working on client's location for end customers.
- Experience in scheduling tool - Oozie and Control M.
- Exposure of working on client's location for end customers.

PROJECTS

Albert Heijn

August 2022 – Till date

- As a core data platform team, we run our data platform on Azure, and use technologies like ADLS, Databricks, Azure Data Factory , EventHub, DevOps, Terraform, Git , Python, Bash and much more
- Involved in building self service Product/Data Platform for Ahold that creates the self-service meta-data driven data platform.
- Responsible for creating and improving our homemade pySpark library which ingests and transforms raw data into clean data.
- Migrating away Customized Legacy Data Lakes
- Building self service Integration framework using Azure Databricks, Azure Data Factory.
- Experience in working with both real time and batch data
- Experience in working with different file formats – json, avro, csv, parquet and others.

Wipro

OPTUS (Oct 2021 – April 2022)

Medallia Enhancements

Medallia captures customer feedback for Optus created through digital channels, over video and social media and IoT interactions and sends it to Optus as feed files (CSV). This data helps Optus to calculate NPS(Net Promoter score) which helps them to analyze their area of improvement.

Roles and Responsibilities:

AWARDS & RECOGNITIONS

Nielsen CTO Appreciation Award
CLP Faculty Award
Start of the month award
On the spot award
Applause Award
Star Team Award

EDUCATION BACKGROUND

2012-2015	MCA
2009-2012	BCA
2007-2009	Higher Secondary
2007	Matriculation

LANGUAGES

English
Hindi
Bengali

HOBBIES



PERSONAL SKILLS



PERSONAL INFORMATION

Date of birth: 11-JAN-1991
Marital status: Married
Nationality: Indian
Passport Number: Valid till 2024

- Ingestion of CSV files from landing path to Hive raw data layer using Spark Scala.
- Applying validation and transformations logic as per the business requirement.
- Performing Sqoop export to Teradata

Previous Organization (TCS) Projects

STELLAR – CAPITAL ADEQUACY PLATFORM @CLIENT – NOMURA, JAPAN (AUG 2020 to SEP 2021)

Stellar Project would develop an ELT platform for Capital Adequacy Reporting platform in **Spark Scala** for large scale dataset in different phases i.e. Ingestion, Data Enrichment, Calculation Engine, and User adjustment. It was primarily focused on migration of legacy ETL data workflows from third party system to Spark Scala workflows

Roles and Responsibilities:

- Building and maintenance of the ETL process in Spark Scala involving data staging, data transformation and processing data pipeline.
- Involved in design, code, reconciliation, Data Gap Analysis for components that manage data ingestion, data extraction, transformation, and loading across a broad portion of the existing Hadoop ecosystems , primarily Spark Scala.
- Involved in performance optimization of code across certain modules written in Spark Scala.

FRAUD RISK MONITORING SYSTEM@ CLIENT – STANDARD CHARTERED BANK, MALAYSIA (JUL 2018 to JUL 2020)

SCB's Private Banking sector implements a selected Fraud Surveillance solution for PvB Singapore, Hong Kong, UK and Jersey booking centres. This project was initiated to implement at Fraud Solution for Private Banking using Hadoop , primarily Pyspark.

Roles and Responsibilities:

- Designed and developed various modules for implementing various business validation and transformation rules to clean the sourced
- Designed and developed a reconciliation framework in Pyspark to put checkpoints of data progress to report any discrepancies seen in data.
- Involved in setting up Control M drafts for managing and scheduling Hadoop jobs.

ECOMMERCE SOLUTION@ CLIENT - NIELSEN USA (NOV 2015 to JUNE 2018)

Nielsen's U.S. e-commerce solution provides its clients a holistic e-commerce measurement solution to help them gain a stronger foothold within the digital retail marketplace. This solution offers the most complete and integrated view of online and in-store metrics. It offers weekly, monthly, quarterly, yearly sales volume and market share of US ecommerce retailer based on category, subcategory, geographical location.

Roles and responsibilities:

- Designed technical architecture and data Ingestion and processing pipeline using Pyspark.
- Developed Pyspark code to validate, transform, reconcile, derived attributes, merge data and ensure consistency of data-sets.
- Involved in performance tuning on few Modules written in Pyspark.
- Involved in setting up workflow using Apache Oozie workflow engine for managing and scheduling Hadoop jobs.
- Support business decisions with ad hoc analysis as needed.