SURENDRA KUMAR MAHAPATRA

Senior Data Engineer

Profile (Executive Summary)

- **Multi skilled professional with overall 10.6 years of experience in IT Industry including 4+ years of experience in Data Engineering Skills Knowledge on multiple domains like Banking, Media & Education with rich Expertise in pySpark, Azure in building Data Pipeline, Optimizing Performance and writing complex queries.**
- **Knowledge of Relational DB like Oracle, SQL, Sybase, Presto**
- **Experience in Azure Databricks, Azure Data Factory and other Azure Services**
- Excellent interpersonal, communication and mentoring skills. Communicating with people across the organizational hierarchy and consistently delivering results in high-pressure environment.
- Ability to handle/work on multiple projects and work in teams across geographic locations with seamless integration with other systems and available services

Professional Experience & Certifications

Top Skills (Technical/Nontechnical skills)

- Big Data Processing Framework (Apache Spark)
- Product, Service Development
- DB Modelling, DB architectures
- Programming Languages like Python, C#
- Relational DB like Sql, sybase and Oracle

Languages	Python, C#
Technologies	Spark 2.0 & above
Cloud Services	MS Azure, Azure Data Factory, Azure Databricks, Azure Synapse
Databases	Oracle,SQL Server, Sybase, Presto
Development	PyCharm, Git, Jira/Confluence, Databricks Notebook
Orchestration	Airflow, Jenkins, Crontab
Source Code Management	GITLAB, GITHUB, Azure Devops
Monitoring	Spark UI, Airflow UI, Grafana
Scripting Language	Shell Script, Java Script, Power Shell Script
OS	Windows, Unix and Linux
Certification	**Databricks Certified Associate Developer for Apache Spark 3.0 Databricks Lakehouse Fundamentals Microsoft certified Azure Data Engineer (DP-203)**

Project Experience – Detail

Senior Data Engineer
Jan, 2021 – Till Date
TCS, Bhubaneswar, INDIA

<u>Key Responsibilities:</u>
- Developed and maintained scalable data processing solutions using **Azure Databricks** with PySpark for large-scale batch and streaming workloads.
- Designed and built complex **ETL pipelines** to transform raw data from Azure Data Lake Storage into curated datasets for analytics and reporting.
- Designed and implemented data ingestion pipelines from Sybase to Azure Data Lake using PySpark and **Azure Data Factory**.
- Built scalable ETL pipelines for processing 50M+ records daily using **PySpark transformations** and Spark SQL.
- Implemented **Slowly Changing Dimension Type 2 (SCD2)** logic for historical data tracking from HUB to curated zone.
- Scheduled workflows using **Apache Airflow** for daily batch processing and monitored failures via Teams notifications.
- Contributed to performance tuning **of Spark jobs, optimizing partitioning** strategies and cluster resource usage.
- Writing Complex queries for generating reports based on the **processed data in HDFS**
- Managing ETL pipelines on Databricks through **notebooks and Spark**
- Worked closely with business teams to understand data requirements and deliver custom datasets via Spark.

Project	Precima Migration
Role	Senior Data Engineer
Duration	Nov-2023 to till Date
Client	**Nielsen**
Technology	Spark, Python, Azure Databricks, Azure Datalake, SQL,
Description	Nielsen buys Precima to expand retail, CPG analytics. Precima has a software-as-a-service (SaaS) platform that relies on artificial intelligence (AI) and machine learning tech to help retailers improve planning and operations for marketing, promotions, pricing, inventory and collaborations with suppliers.As a result of the deal, Nielsen Global Connect clients have access to Precima's services such as consumer loyalty and retailer analytics software. Precima legacy model previously use Control M and on-prem server for data proessing which is migrated to Azure services (Azure datalake, Azure databricks, Azure data factory).
Responsibilities	• Designed, developed, and maintained **modular PySpark-based ETL pipelines** in Azure Databricks to process structured and semi-structured data. • Created **parameterized notebooks** and job clusters for dynamic pipeline execution across multiple datasets or environments. • Implemented **Delta Lake** architecture for versioned data storage, ACID transactions, and time travel capabilities. • Integrated Databricks jobs with **Azure Data Factory** and **Apache Airflow** for scheduled, event-driven, or dependency-based workflow orchestration. • Ensured **data quality and integrity** by applying validation checks, schema enforcement, and exception handling in PySpark code. • Used **Unity Catalog** for secure access control and metadata management in the Databricks workspace. • Tuned Spark jobs using best practices such as **caching**, **broadcast joins**, **coalescing**, and **partitioning** to reduce cluster costs and improve runtime.

	• Analyzed Spark UI metrics to identify and resolve performance bottlenecks in long-running jobs • Handling a team of 5 member
Project	DD Station Reference
Role	Data Engineer
Duration	Jan-2021 to Oct-2023
Client	**Nielsen**
Technology	Spark, Python, SQL,Sybase DB, Azure, Airflow, Hadoop
Description	Distributor Directory is the system of record for all media Distributor and related Distribution Source data as well as Crediting definitions. The DD application create distributor, distribution source and assign call letter, station code, DMA (Designated Market Area), SID(Source Identification), different characteristics to distribution source. Maintain the crediting information. Few information's are coming from upstream and other information are entered into the DD system. The data are consumed by different downstream for business purpose. The data send to MDL(Media Data Lake) by spark jobs
Responsibilities	• Designed and developed robust ETL pipelines using Apache spark for efficient data processing. • Developed and maintained scalable data integration workflows using python and spark. • Transformed complex data sets into structured formats to support business analytics and reporting. • Utilized Hive for data warehousing solutions, optimizing query performance through partioning and indexing • Managed and optimized data storage in HDFS, ensuring data integrity and availability • Conduct performance tuning and optimization of spark jobs to reduce processing time and resource utilization • Collaborated with data scientists, analysts and other stakeholders to understand data requirement and deliver solutions. • Provided technical support and troubleshooting for data pipeline issues, ensuring minimal downtime.
Project	IMS (Interactive Market System)
Role	Senior SQL Developer
Duration	Mar-2019 to Dec-2020
Client	Nielsen
Technology	SQL Server, C#, Asp .Net, Web Service
Description	The business IMS (Interactive Market System) is a set of various reporting tools to analyze the data for a better market reach. IMS has a suite of applications. The applications are used by Data analysts to analyze the media survey data for a better market reach. Clear Decisions and IMS Admin is the important applications in IMS suite used to define a target population and analyze their various demographics, product usage, behavior and media and lifestyles choice.
Responsibilities	• Designing databases and ensuring their stability, reliability, and performance. • Design, create, and implement database systems based on the end user's requirements. • Improve application's performances. • Prepare documentations for database applications. • Develop database schemas, tables. • Ensure the data quality and integrity in databases.

	• Fix any issues related to database performance and provide corrective measures. • Create complex functions, scripts, stored procedures and triggers to support application development. • Develop best practices for database design and development activities.
Project	Environment Booking
Role	Senior Developer
Duration	Mar-2017 to Feb-2019
Client	Canadian Imperial Bank of Commerce
Technology	C#.Net, SQL SERVER, JAVA SCRIPT
Description	Environment Booking is a CIBC internal application. Using this application user can raise request for test environment. It goes through approval process and after taking action on the request user notified with an email. User can view the application wise use of environment for a particular time period through chart and reports.
Responsibilities	• Analyzing the applications requirement • Quickly produce well-organized, optimized, and documented source code • Writing SQL queries • Optimizing the queries • Interface with clients and gather business requirements and objectives. • Design, develop and implement new integration. • Review, to the extent possible, changes in code and the environment that will affect system performance. • Responsible for analysis of current programs including performance, diagnosis and troubleshooting of problem programs, and designing solutions to problematic programming.
Project	HRIT Staffing System
Role	Developer
Duration	Feb-2016 to Jan-2017
Client	Microsoft
Technology	MS SQL Server, C#.Net
Description	HRIT Staffing system is the application used for Microsoft recruitment process. It integrates data from different application like Global careers, internal Careers, Candidate Portal, Recruitment web etc. For new version of HRIT staffing data need to migrate to new system. This project is deal with data migration where all recruitment process data will move to other platform. So the data need to optimize, processed as per the new system.
Responsibilities	• Identifying the objects and attributes to be migrated in sync with 3rd party tool • Producing detailed specifications and writing the program codes • Writing SQL queries (SPs, Views, Tables, Jobs) • Debug existing source code and polish feature sets. • Interface with clients and gather business requirements and objectives. • Work independently when required • Continuously learn and improve skills • Testing the product in controlled, real situations before going live

Software Developer
Aug, 2014 – Dec, 2015, Tech
Mahindra

Key Responsibilities:

- Understand business needs and know how to create the database applications.
- Analyzing the applications.
- Writing SQL queries, triggers and functions and SQL jobs.
- Provide innovative solutions to complex business problems.
- Work collaboratively with others to achieve goals
- Be a persistent, creative problem solver
- Maintaining the systems once they are up and running
- Maintain positive client interactions.

Project	FATCA(Foreign Account Tax Compliance Act)
Role	Software Developer
Duration	Aug-2014 to Dec-2015
Client	Scotia Bank, Canada
Technology	SQL Server, C# .Net, Asp .Net, Ado .Net
Description	FATCA stands for Foreign Account Tax Compliance Act. FATCA represents the Treasury Department's efforts to prevent U.S. taxpayers who hold financial assets in non-U.S. financial institutions (foreign financial institutions or FFIs) and other offshore vehicles from avoiding their U.S. tax obligations. FATCA requires Foreign Financial Institutions (FFIs) to report directly to the US Internal Revenue Service (IRS) certain information about the financial accounts held by US Taxpayers, or by foreign entities in which US taxpayers hold a substantial ownership interest. These reports are in xml format and generated from DION application. We developed a DIT application to check the report validation.
Responsibilities	- Analysis and Design of Database - Developed and Maintaining SQL Server objects (Stored Procedures, Views and Functions) - Preparing unit test cases. - Execute system development and maintenance activities.. - Understand and evaluate complex data models. - Tests new programs to ensure that logic and syntax are correct, and that program results are accurate; assists lower-level programmers with programming assignments.

General Information

Education	Degree: Bachelor of Technologies– Computer Science College/University: BPUT –DRIEMS College
Citizenship	Indian
Languages	English, HIndi and Odia