ANAY PALANDURKAR

BITWISE GLOBAL, PUNE

Email	**:**	anaypalandurkar@gmail.com
Mobile No. :		+91-8805317588
LinkedIn	**:**	https://www.linkedin.com/in/anay-palandurkar-bb2b4b105/

OBJECTIVE

Seeking an environment that encourages learning, creativity, provides exposer to new ideas and to hold a growth-oriented position in a reputed organization, all this to utilize and enhance my knowledge for career building.

PROFESSIONAL PROFILE

- Results-oriented individual having 3+ year's Industrial experience where 2+ yrs of experience in different domains in IT industry that includes Hadoop Ecosystem, Spark API's, Cloud Services (AWS, Databricks) and SQL Databases.
- Hands on experience in delivering appropriate technology solutions for Big Data core components and Ecosystem including Data Ingestion and Data Processing (Spark, Hive, SQOOP, HDFS).
- Processing large amounts of structured data, including integrating data from multiple sources.
- Importing and exporting data from different Relational databases into HDFS using Sqoop.
- Developed hive scripts with partitioning and bucketing for the hive tables for better performance.
- Designed Both Managed and External tables in hive for optimization performance.
- Experience with building the Spark application, RDD, Data frames.
- Experience on working with cloud services like Amazon Web Services (AWS) and Databricks.
- Excellent communication, collaboration and team interactive skills with proficiency at grasping new concepts quickly.
- Aspiring for a challenging career in Software Development as to use my skills and experience for best results.

TECHNOLOGY   PROFICIENCY

Operating System	Windows, UNIX
Hadoop Ecosystem	HDFS, Map-Reduce, YARN, Hive, Sqoop, Spark
Relational Database	Oracle, MySQL
Programming Language	Python
Cloud Platforms	Amazon Web Services (AWS), Databricks
SDLC Methodology	Agile
Ticketing Tool	JIRA
Distribution Package	Cloudera
Tools / IDE	PyCharm, Putty, SQL Workbench, MSOffice

Project 1

BITWISE GLOBAL

[Aug 2021 – Current Date]

Retail Store Data Analysis

Designation: Data Engineer
Team Size: 9

Description:

Retail Store Data Analysis is Data Hub were analysis for a range of products like groceries, home decor, furniture, electronics data is done on Big Data platform. Worldwide this is estimated to be a multi-billion-dollar problem. These stores generate the sales data at their POS (Point of sale) & get pushed into the big data pipeline for cleaning, processing and Business analysis.

Need for Retail Store Data Analysis:
- Parallel Data Processing that is unconstrained.
- Provide storage for billions and trillions of unstructured data sets.
- Fault tolerance along with high availability of the system.

POSITIONS OF RESPONSIBILITY

- Attaining a daily scrum call with team members and discussing the task on a particular sprint.
- In Every sprint we have assigned tasks that we are completing as per timeline.
- Have responsibility to write efficient code as per required and requested
- Filter data and perform multiple operations in order to get meaningful insights that will help the client in decision making and analysis.
- Creating PySpark job for extract into S3.
- Basically we have created a End to End data pipeline for Retail stores data using HDFS/S3,Apache Spark, Oracle And PowerBi.
- Create spark jobs for data transformation and aggregation.
- Produce unit tests for Spark transformations and helper methods.
- Implement Data Validation, Quality Checks, Profiling.
- Engage with BA to understand the requirements clearly.
- Monitor Pre-Validation process to validate the client data.
- Monitor the jobs daily, fixing production bugs and raising ticket for production activity.
- Resolve JIRA tickets raised by Client.

Project 2

BITWISE GLOBAL

[Sept 2020 – July 2021]

Enterprise Datawarehouse Analysis
Designation: Data Engineer

The purpose of the project is to do the ingestion of the super market data that is present in Mainframe system to Data Warehouse . It's Mainframe to HDFS ingestion project in Real Time. Data migration utility to migrate existing system into MySQL . Later the extraction would be done in PowerBI.

Roles & Responsibilities:
- These stores generate the sales data at their POS (Point of sale) & get pushed into the big data pipeline for cleaning, processing and Business analysis.
- Basically, we have created an End-to-End data pipeline for Retail stores data using S3,Apache Pyspark,mySql And PowerBi. We have to filter data and perform multiple operations in order to get meaningful insights that will help the client in decision making and analysis.
- We have written logic in Pyspark using **Databricks**.
- Developed PySpark programs and created the data frames and worked on transformations. Involved in working on Spark SQL Code as an alternative approach for Faster Data Processing and better Performance. Experience in data ingestion, transformation and performance tuning. Engage with BA to understand the requirement clearly and comes up with missing scenarios and get them clarified with BA. Actively participated in daily stand-up, biweekly SCRUM and project meetings.

Technical environment: RDBMS, AWS S3, AWS EMR, PySpark, Databricks

Academic Qualifications:

Qualification	Year	Institute/School
Bachelor of Engineering	2019	Marathwada Institute ofTechnology, Aurangabad
HSC	2015	Deogiri College
SSC	2013	St. Lawrence English School

PERSONAL PROFILE

Name	:	Anay Anil Palandurkar
Date of birth	:	05-August-1997
Nationality	:	Indian
Sex	:	Male
Languages Known	:	English, Hindi, Marathi
Address	:	Mandar Bhavan, Plot No 111-A, Old Subhedar Layout, Behind Bank Colony, Nagpur
Contact	:	Mob : +91-8805317588
		Email : anaypalandurkar@gmail.com
		LinkedIn : https://www.linkedin.com/in/anay-palandurkar-bb2b4b105

Declaration:

I hereby declare that the above information is true to the best of my knowledge.

Anay Anil Palandurkar