

SANKET SANTOSH KANCHNAKAR

DATA ENGINEER

 kanchankarssk@gmail.com

 Aurangabad, MH, India

 +91 8329781816

TECHNICAL SKILLS

BIG DATA FRAMEWORKS

Hadoop, Spark, Sqoop, Hive, HBase, Phoenix

RELATIONAL DATABASE

Oracle, MySQL

CLOUD TECHNOLOGY

AWS (EC2, S3, EMR, RDS, IAM)

PROGRAMMING LANGUAGE

Python

Objective

To be associated with a progressive organization that provides an opportunity for a challenging and rewarding career by applying my knowledge, Skills, and Potential in the Profession.

Summary

- Data Engineer with overall 3.5 years of Industry Experience with 2.4 years of Experience as Data Engineer.
- Hands on experience with Big Data core components and Eco System including Data Processing like PySpark, Hive, Sqoop, HDFS, HBase.
- Implemented a framework for ETL (Extract, Transform, Load) using PySpark to improve data processing.
- Experience with working on cloud infrastructure Amazon Web Services (AWS).
- Work experience in the project requirements like Gathering, Performance optimization, Data extraction, Cleaning and Reporting.
- Experience in Data ingestion using, PySpark with data sources into HBase using Sqoop and perform transformations using Hive, Spark and then loading data into S3.
- Experience in using optimization techniques in Hive, Sqoop, Spark, SQL if required.
- Working experience in handling SQL queries for structured data bases.
- Working in an Agile/Scrum Methodology with JIRA as bug tracking tool.

Work Experience

Zielotech Infosolutions India LLP

JUNE 2019 - Present

Data Engineer

Project: MONETARY INTELLIGENCE UNIT

Nov 2021 – Present

Environment: AWS, S3, EMR, PySpark, Hive, HBase, Phoenix, Oracle, Jira.

Project Description:

In order to maximize the benefits, credit card providers need to start changing their views regarding transactional credit card data, and start looking at it more from a retailer's perspective. Tapping into the location, product, and competitive intelligence aspects of credit card data opens new revenue sources for both the providers and the merchants for this we have used AWS for cloud processing. In this project we have used for processing as PySpark, and cloud-based technology as AWS.

OTHER TOOLS

PyCharm, WinSCP,
Putty, PowerBI

EDUCATION

B.Tech

from

SGGSIT(Autonomous),
Nanded, MH.

PERSONAL DETAILS

Date of Birth: 17/10/1994

Marital Status: Single

Languages Known:

English, Hindi, Marathi

Role and Responsibilities:

- Data migration from various sources such as RDS to HBase, HBase to Hive, Hive to S3 using Pyspark.
- AWS S3 for Upstream and Downstream Data storage to handle historical, reference and delta data using PySpark.
- Different transformation on reference and delta data like data cleaning, deduplication, Join, etc. using PySpark.
- Analytical thinking about handling long running jobs to minimize time required for completion using spark optimization techniques.
- Set up different scenario to write data using PySpark job like if job run successful, failed, fetching errors.
- Working on HBase top on of its Phoenix for providing SQL interface.
- Using PySpark do different transformation like joins, de-duplication as prerequisites.
- Worked with Parquet and ORC file format for the various compression techniques to leverage the storage in HDFS.
- Working experience in handling optimize performance concepts in Hive and designing both Managed and External tables in Hive.
- Filtered data and performed multiple operations in-order to get meaningful insights that help our client in decision making and analysis.
- Working in Parallel with QA to solve bugs raised on particular modules.

Project 1: Merchandise Store Data Analysis

July-2020 – Oct2021

Environment: AWS, S3, EMR, PySpark, MySQL, Jira, PowerBI

Project Description:

Retail Store Data Analysis is an analysis of everything in your business, from your sales and inventory to your customer data. It gives you the ability to effectively track customer actions, like their purchases and foot traffic in your store.

Role and Responsibilities:

- XYZ stores are chain of retail stores, serving the customers worldwide with a range of products like groceries, home decor, furniture, electronics.
- 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 End to End data pipeline for Retail stores data using S3, Apache Spark, MySQL and PowerBI.
- We have to filter data and perform multiple operations in order to get meaningful insight that will help the client in decision making and analysis
- Have responsibility to Write the data extraction logic for the application.
- Raise JIRA for Infrastructure, Platform issues.
- Engage with BA to understand the requirements clearly.

- Define all the possible test cases along with the Test Data.
- Comes up with musings scenarios and get them clarified with BA.
- Implement Data Validation, Quality Check, Profiling.
- Resolve Jira Tickets raised by Client.

I hereby declare that all the information mentioned above is true and to the best of my Knowledge.

Sanket Santosh Kanchankar