Nitin Prakash Gadge

Add: Geeta Nagar Himayat Bagh,   Mob No: +91–9607267370
Plot No 8-28-243, Aurangabad- 431001   Email: gadgenitin097@gmail.com

CAREER OBJECTIVE

Looking forward to explore more in the big data domain, adapt as much as possible, build a better career for myself and be useful to the organization I am associated with.

CAREER SUMMARY

- **Good knpwledge of Hands-on Experience** in Hadoop Ecosystem components.
- Have Good knowledge about HDFS, YARN, and Map Reduce.
- Uploaded and processed terabytes of data from various structured and unstructured sources into HDFS.
- Good understanding of NoSQL databases and hands on work experience in writing applications on NoSQL databases like HBase.
- Collecting large amount of structured data using Apache Sqoop and aggregating using HIVE.
- Designed Internal and External tables in hive.
- Good understanding of Partitioning, bucketing concepts in Hive for performance optimization.
- Good exposure with Agile software development process.
- Experience in rebalancing the data in the cluster.
- Manage and review Hadoop log files.
- Experience in giving on-job training to the new members those have less experience.
- Strong problem solving & technical skills coupled with clear decision-making.
- Hands-on Experience in Hadoop eco-system. Good exposures in Requirement Analysis, Customization and user acceptance.
- Experienced in working with Amazon Web Services (AWS) using EC2 for computing and S3 as storage mechanism.

TECHNOLOGY PROFICIENCY

- **Platforms:**   Windows, Linux.
- **Programming Language:**   SQL, Python
- **Scripting Language:**   Shell Script.
- **Big Data Ecosystem:**   HDFS, Sqoop, Hive, Spark core, Spark-SQL.
- **Distribution Package:**   Cloudera
- **Management Tool**:   Zookeeper.
- **IDE**:   Eclipse, PyCharm
- **Cloud Technologies**:   AWS (S3, EC2, EMR)
- **Relational database**:   MySQL, Oracle.

ORGANIZATIONAL SCAN

- **Organization:** Appstallions Technologies Pvt Ltd.
- **Designation:** Data Engineer
- **Location:** Pune
- **Duration:** From June 2019 to Till Date.

WORK EXPERIENCE

Project #1: Data Injection Framework   **Duration:** June 2019 to June 20

Role: Data Engineer

In this framework, we fetching the data from different sources and ingest the data into data lake. For this we are directly connecting to their source system & we ingest the data into our Hive table. For some of the sources, we are using File formats for ingestion. For testing, we are taking some sample data & we ingest that data in to our Hive table with Data Migration application. Then we are loading historical data batchwise into the cluster.

Roles and Responsibilities:

- Import Data into the Hive from various Relational Databases (Oracle, MYSQL Server) into HDFS using Sqoop.
- Write Hive DDL and DML queries in Hive.
- Ingest Flat Files like Delimited, Fixed Length, etc. into Hive Warehouse.
- Engage with BA to understand the requirements clearly.
- Perform optimization techniques in order to enhance performance of Sqoop job.
- Worked with various File Formats like Text File, Sequence File, ORC Files, Parquet Files and various Compression Formats like Snappy, bzip2.
- Designed both Managed and External Hive Tables and Defined static and dynamic partitions as per requirement for optimized performance on production datasets.
- Written Hive Queries for Data Analysis to meet the business requirements.
- Monitoring Data Validation.

Environment: HDFS, Yarn, MR, Hive, Sqoop, Oracle, Shell Scripting.

Project #2:   Healthcare System Data Pipeline   **Duration:** June 2020 to Till Date
Role: Data Engineer

For this project we are fetching the data from various sources and processing the data in pyspark like data cleaning, data integration, data transformation.
After processing the data in pyspark, the output result we are storing on Hive as a Business View for further generating the reports as per client's requirements.

- Importing the data into pyspark from various sources.
- Data cleaning, data Integration, Data Transformation for data compatibility.
- Processing the data in pyspark as per the requirements.

- Raise JIRA for Infrastructure, Platform issues.
- Engage with BA to understand the requirements clearly.
- Monitor Spark jobs.
- Perform optimization techniques in order to enhance performance of Spark jobs.
- Storing the final processed data in to Hive table with optimum use of Hive Optimization techniques.
- Monitor Data Validation
- Written Hive Queries for Data Analysis to meet the business requirements.

EDUCATIONAL QUALIFICATION

- Bachelor of Computer Science from the Maharashtra University with the percentile 68.45 in the academic year 2014-15

PERSONAL DETAILS

Date of Birth   : 08 june 1992
Languages known   : English, Hindi and Marathi

Date:

Place: (Nitin Prakash Gadge)