Gaurav Soni

91+(8764001727 / 8005789566

gaurav.soni226@gmail.com   LinkedIn: gaurav-soni-007 GitHub: gsoni22

CAREER SUMMARY

Result oriented & passionate professional with 3 years in Distributed System expertise in implementing Batch and Stream Data Processing solutions. Vertical industry experience: retail, supply-chain, Media Solutions, social analytics.

IT Eminence

- Azure Databricks
- Azure Data Factory (v1, v2)
- Azure Data Engineer
- Azure SQL DB / DW
- Azure HDInsight
- Azure Machine Learning Services

- Apache Spark (Batch/Streaming)
- Spark Catalyst Optimizer
- Machine Learning (MLlib)
- Apache Flink, Pyspark
- Hive, SQL, No-SQL, Airflow
- Cluster Manager (Mesos /YARN)

- Scala*/Python*/Java
- Docker Containers / K8
- Nginx Web Server
- Lambda and Kappa
- Databricks Platform
- Redis Caching

ACCOMPLISHMENTS SUMMARY

- A Big Data and Data science professional with 3-year experience as Apache Hadoop/Spark Developer using Big Data Technologies like Hadoop Ecosystem, Spark Ecosystem with Azure cloud platform.
- **Expertise in Multiple Azure data platforms** e.g. Azure Databricks, Azure HDInsight
- **Tuning of Spark Catalyst Optimizer for different queries and user cases.**
- **Improved performance of ETL pipelines by tuning Spark executors and partitions size.**
- **Defined partitions schema and leveraging benefits of Scala Implicit functionality to improve performance.**
- **worked on code test libraries like Scala Test, Junit for unit testing of the code.**
- **worked on** Azure SQL Database and Data Warehouse, Azure blob Storage.
- **Domain expertise** in **Financial services, Media Industry** and **Supply Chain Analytics**

PROFESSIONAL EXPERIENCE

One97 Communication (Paytm Payments Bank)   **2020 April to Present**
- **Consumer Bureau Feature store:**
  - Analysis of user data coming from CIBIL and Experian in encrypted format
  - Conversion of records from XML to relational format so that other teams can consumer data for modeling.
  - Generated more then 1500 features for each user for daily basis.
  - Whitelisting of users for there credit cards and Postpaid accounts / loan accounts
  - Identifying the user behavior whether he will be loan defaulter or not if we provide him the loan.
- **Citi / IndusInd / SBI Banks DRE exercise:**
  - To provide credit cards outside of Paytm users who prefer specific banks this activity done on Paytm data by those banks.
  - Generation and consuming the features store for running the machine learning models which are compiled in C++ shared objects (.so) file on the secured infrastructure and environments.

Celebal Technologies Jaipur, India   **2017 to 2020 March**
- **Vedanta Aluminum Limited: Algorithm and System Design**
  - Scheduling BTAPs for Vedanta Alumina Ltd. Vedanta wants to use there BTAP/Train in efficient way such that their domestic import should be maximum and international import should be minimum. it's a constraint-based route optimization problem. which we solved using activity-based model.
  - **Language Tools and Frameworks**: Flask, **Azure Blob Store**, Git, JSON, **Azure SQL Database**
- **Viacom18 Media Pvt. Ltd: Data Pipeline and ETL Development and Migration**

- Developed Data pipelines of Viacom for more than 10 data sources and performed ETL on them the data size per day is 2 Terabytes which we compressed using snappy codec and saved in parquet file format.
- Implemented KPI logic and conversion of values into Vectors to improve overall performance of the system.
- Data extraction using Post API, SSIS Packages, SQL Database, SQL Data Warehouse etc.
- **Language Tools and Frameworks**: Thrift Server, **Azure Databricks, Azure HDInsight, Azure Blob**

- **Align Technology (Spark Streaming):**
  - Align Tech is a manufacturer of 3D digital scanners which are used in orthodontics.
  - The Data is coming in **streaming format** from salesforce for which we have used salesforce-streaming-client
  - The client requirement is to allocate reward points to all their registered doctors based on their patient sales and recommendation of devices to the patients.
  - sales records are submitted to salesforce by their retailers and which we consumed in Azure EventHub and sent back to Databricks for performing Analytics on them.
  - Reward points are allocated based on their rules which are decided by align team, so they store all this rule in SQL database, we are taking all those rules which are in form of sting e.g. 1 and 2 or 3.
  - **Language and Frameworks**: Apache Spark Streaming, JSON Parsing, **Azure Databricks, Data structures**
- **DvSum (Bethesda MD)**
  - It's a supply chain solution for coca cola in which we perform inventory optimization techniques, and demand-forecasting of their products.
  - Data has been stored in ZIP file on their Azure blob and inside each zip file there is a nested directories and files and we performed KPI calculation of the data and store back into sql database.
- **Yes Bank**
  - Yes, Bank want to identify whether their employees are in formal wear or in casual wear and it's a computer vision problem.
  - we have implemented it using Customvision.ai which is a platform by Microsoft to solve all types of vision based problems.
  - Feature extraction based on all above category, we trained a deep neural network and deploy it on the azure platform so that we consume it via REST API calls

ACHIEVEMENTS

- **Google Applied CS Facilitator** in Rajasthan (2017-2018)
- Conducted Android workshops on behalf of Google.
- Continue two times Winner of Inter College IT Quiz Competition
- Continue two times 2nd Position in Class Presentation
- **Teaching Assistant in** Google Developer Days
- First Runner-Up in Rajasthan IT Day **Relay Coding Event**.
- **Organize a State Level Event in MCA 4th Sem: Code Shuffle**
  - Got Sponsorship from Digital Ocean , DoSelect , HackerEarth , JetBrains , GitHub

CERTIFICATIONS

- Microsoft Certified:
  - Big Data for Data Analytics (MPN14349)
  - Data Platform & Analytics Foundational for Data Analytics Competency (MPN14356)

EDUCATION

- Master of Computer Application (MCA) from JECRC University in Computer Application, Jaipur
- Bachelor of Computer Application (BCA) from Vijay Singh Pathik College in Computer Application, Ajmer