Gaurav Sehrawat

https://gauravsehrawat19.wixsite.com/website

Email : gauravsehrawat19@gmail.com

Mobile : +91 9927993798

Education

- **Meerut Institute of Engineering Technology (UPTU)** — Meerut, UP
  Bachelor of Engineering in CSE; 75 percent — *Aug. 2013 – June. 2017*
- **Kendriya Vidyalaya No. 1, AFS-1, Jamnagar (CBSE)** — Gujarat, India
  Intermediate; 78 percent — *2011-12*
- **Kendriya Vidyalaya NAD Karanja (CBSE)** — Mumbai, India
  High School ; 81 percent — *2009-10*

Experience

- **IBM India Pvt Ltd** — Manyata Embassy Business Park, Bengaluru
  Software Engineer — *Feb 2018 - Present*
  - **Apache Spark**:
    - Used Python language and its Pyspark API to code the spark applications.
    - Have a good knowledge of SCALA language to code the Spark applications.
    - Hadoop Yarn is used as the scheduler, resource and cluster manager.
    - Structured API i.e. Data Frames are used extensively in spark application for the ETL process.
    - Use of Spark SQL along with data frame to query the data and efficiently building the ETL process.
    - Good knowledge of working with low level API RDD.
    - Have a solid understanding of sparks MLib and GraphX libs.
  - **Apache Hive**:
    - HQL queries are written and the scripts are called from bash shell scripts.
    - HQL is also used in the spark SQL alongside the SQL language.
    - Good hold on partitioning of table, dropping of partition, creating new tables and other manipulation queries.
    - Query hive from console, Ambari and Hue.
    - Creating UDFs in Java.
    - Optimising the hive queries using bucketing, partitioning, vectorization and skewed/temporary tables.
    - Data optimisation using different file format, compression and optimisation.
  - **Linux (shell scripts)**:
    - Basic commands in Linux bash shell.
    - Writing shell script to automate the ETL process.
    - Writing the config files, pashed along with .sh files while execution of the script.
    - Hive and the spark script are called from the shell scripts to accomplish the ETL task.
    - Shell programming which includes if else blocks, looping constructs, writing functions and many more.
  - **Python**:
    - Python used to code the Spark application using Pyspark API.
    - Using python Pandas library along with spark data frame for ELT process (i.e. Extract, transform and load).
    - Use of extensive Pandas functionality to achieve the ETL task
    - Using python matplotlib library to create visualisation from different datasets.
    - Array manipulation using NumPy library.
    - Using python to send the mails of the status of the script's execution using smtplib.
  - **SQL Server (Version 2008 and 2016)**:
    - Writing complex SQL queries
    - Writing views, using join conditions
    - Writing Stored procedures.
    - Have a good knowledge of T-SQL
  - **SQL server integration services (Version 2008 and 2016)**:
    - Creating SSIS packages in SSDT.
    - Deploying the package to SSISDB catalogue in SQL server using SSMS.

- Creating, managing and running SSIS jobs in SQL server using SSMS.
- Working with Flat file, OLE DB and Excel connection managers.
- Working with various transformations like Lookup, Merge, Scripts, Execute SQL Task, Data Flow Task, Execute Package Task, and various other transformations.
- Heaving knowledge of various containers in SS IS i.e. For loop, for each loop, sequential containers.
- Good knowledge of Control flow, Data flow, variables, parameter and event handler.

◦ **Microsoft PowerBI**:
- Core expertise in Power BI Desktop, requirement mapping and translating them into user stories.
- Expertise in designing and creating data model for Power BI reports.
- Proficient in Data Analysis Expressions (DAX) with ability to write complex DAX functions in Power BI and Power Pivot.
- Proficiency in writing complex SQL queries.
- Strong in writing power queries and applying advanced calculations using MS Power BI Desktop (Aggregate, Date, Logical, String, table).
- Very strong in different visualizations using Slicers, Lines, Pies, Maps, Waterfall, Heat Maps, Tree maps.
- Capable of optimizing Microsoft Power BI dashboards with a focus on usability, performance, flexibility, testability, and standardization.
- Analytics using R in Power BI Desktop and Services.
- Creating and deploying SSIS packages.
- T-SQL Programming and creating Stored Proc.
- Creating jobs in SQL server

◦ **Microsoft Excel**:
- Using Microsoft power pivot to perform the ETL tool
- Using advance excel formulas and concepts to create prototype of the application.

Projects

◦ **CET Dashboard [UPS, USA]** : Flat files (i.e. feed files) are received from different services that run in the mainframe. The files are placed in the network storage using a jump server. SSIS packages used to pick the files form this location and processes the data. The data is loaded to the data warehouse implemented in SQL server.

◦ **MetLife [USA]** : Spark, cluster computing framework, is used as ELT tool. I have worked on the structured API of Spark which includes spark Data frames. Spark SQL is used extensively to query the data. Spark SQL provides functionality to both SQL and HQL.