Harshit Gupta

📞 +91 8340558673 | ✉ harshitgupta.gh53@gmail.com | in harshit053 | ⚙ harshit053 | 📍 Bangalore, India

SKILLS

- **Core Skills**: Python3, Apache Spark and PySpark, SQL, Apache Airflow, BigQuery, Google Cloud Platform, PyTest.
- **Web Development Skills**: HTML, CSS, JavaScript, NodeJS, ExpressJS, MongoDB, Bootstrap.
- **Additional Skills**: Data Structures and Algorithms, Git, GitHub, MongoDB Atlas, Agile Methodology (Jira, Rally).

Certifications

- **Google Cloud Platform** – Data Engineer, Big Data

WORK EXPERIENCE

Infosys Ltd, Bangalore, India   **August 2022 – Present**
Client**: American Express - Engineer II**

- Developed PySpark scripts to extract data from GCP BigQuery and convert it to .txt files for transfer to mainframe systems, aiding the Modernization of the Account Receivables System at American Express.
- Utilized Airflow DAGs in GCP Cloud Composer to automate workflows and employed Spark Clusters for efficient job orchestration and cluster management.
- Used Dataproc to optimize job performance to achieve a 25% increase in overall job throughput and responsiveness, thus reducing execution time and enhancing efficiency.
- Implemented Pytest scripts and a Behaviour-Driven Development (BDD) Testing approach to ensure script reliability, code quality and robustness.
- Developed Python-based SFTP scripts and implemented PGP Encryption to securely transfer files from GCS buckets to mainframe, adhering to Amex's specific SFTP protocols. Also configured the file transfer parameters using Amex's SFTP tools.
- Utilized mainframe systems to verify the successful transfer of processed data files and validated data file accuracy after transfer by performing fundamental system handling tasks.
- Leveraged Agile methodologies and related tools (Rally and Jira) to collaborate with cross-functional teams and to optimize workflows and enhance system performance.

Skills Acquired: **Python3, PySpark, SQL, Airflow, PGP Encryption, GCP - BigQuery, Cloud Composer, Dataproc**

PROJECTS

Data Processing Pipeline| **Python3** , **PySpark** , **Apache Spark** , **SQLite**   Website | GitHub

- Demonstrated a simple data processing pipeline designed to clean, transform, and analyze data.
- Optimized Join Operation using broadcast join to efficiently join the resulting dataframes.
- Validated records by identifying suspect records and separated good records from bad ones.
- Created an appropriate data model in an SQLite database and ingested the cleaned data into it.

Skills Learnt: Data extraction, Data handling and manipulation and Data ingestion

YelpCamp | **MongoDB** , **ExpressJS** , **EJS** , **NodeJS** , **PassportJS** , **Bootstrap**   Website | GitHub

- Developed a full-stack campground review site with user accounts, campground listings, and review posting.
- Implemented authentication and authorization using Passport.js with session-based login.
- Built complete CRUD functionality for campgrounds and reviews, adhering to RESTful API principles.

Skills Learnt: Full-Stack Development, CRUD operations, Authentication and Authorization

DigiMenu | **HTML** , **CSS** , **JavaScript** , **MongoDB** , **NodeJS** , **ExpressJS** , **PassportJS**   Website | GitHub

- Built a digital menu application allowing users to browse and view available dishes and categories.
- Implemented core features for menu management, including item creation and display.
- Incorporated user authentication and login functionality using PassportJS.

Skills Learnt: CRUD operations, RESTful routing, Authentication and Authorization

EDUCATION

Visvesvaraya Technological University, Bangalore, Karnataka, IN   **August 2018 – July 2022**
B.E. in Electronics and Communication | Overall: **8.8 CGPA**