Shravan Kumar |Big Data Engineer|+1 (469) 458-0009©
Email ID: skumar210594@gmail.com
LinkedIn: www.linkedin.com/in/shravankumar999

PROFESSIONAL SUMMARY:

- IT Professional with 7+ years of experience in Information Technology with expertise in designing data intensive applications, Big Data Analytics, Data Warehouse/ Data Mart, Cloud Data engineering, Data Visualization, Reporting and Data Quality solutions.
- Excellent understanding of technologies on systems that include huge amounts of data & run in a highly distributed fashion in **Amazon web services** (**AWS**), **Azure**, **Cloudera**, **Hortonworks** & **Hadoop distributions**.
- Experience in creating & managing reporting & analytics infrastructure for internal business clients using AWS services including **Athena**, **Redshift**, **Spectrum**, **EMR**, & **Quick Sight**.
- Developed data pipelines using AWS services including **EC2**, **S3**, **Glue**, **RDS**, **IAM**, **Athena**, **Redshift**, **Lambda functions**, **CloudWatch**, **Step functions**, **SNS**, **DynamoDB**, and **SQS**.
- Experience in Analytics & cloud migration from on-premises to **AWS Cloud** with **AWS EMR**, **S3,** & **DynamoDB**.
- Extensively worked on Cloud services like AWS (**EC2**, **S3**, **EMR**, **Lambda**, **Cloud watch**, **RDS**, **Auto scaling**, **Cloud Formation**, **SQS**, **ECS**, **EFS**, **DynamoDB**, **Route53**, **Glue**, etc.)
- Experience with **Azure Cloud**, **Azure Data Factory**, **Azure Data Lake Storage**, **Azure Synapse Analytics**, **Azure Analytical services**, **Big Data Technologies** (**Apache Spark**), & **Azure Data Bricks**.
- Worked on Azure cloud components (**HDInsight**, **Data Bricks**, **Data Lake**, **Blob Storage**, **Data Factory**, **Storage** Explorer, **SQL DB**, **SQL DWH**).
- Profession Knowledge on Hadoop architecture and its components like **HDFS**, **YARN**, **Name Node**, **Data Node**, **Application Master**, **Job Tracker**, **Resource Manager**, and **Map Reduce programming paradigm**.
- Experienced in using various Python libraries like **NumPy**, **SciPy**, **Boto3**, **Pandas**.
- Worked on visualization tools like **Tableau**, **synapse**, **Power BI** for report creation and further analysis.
- Expertise with **SQL** and **NoSQL Databases**, **Data Modeling** and **Data Pipelines**.
- Involved in end-to-end development and automation of **ETL pipelines** using **SQL** and **Python**.
- Worked on **NoSQL** databases including **HBase**, and **DynamoDB**.
- Hand-on experience in Big Data analytics, Data manipulation, using **Hadoop Eco system** tools **HDFS**, **Map-Reduce**, **Yarn**, **Spark**, **Flume**, **kinesis**, **Sqoop**, **AWS**, **Spring Boot**, **Spark**, **Pig**, **Hive**, Avro, Solr and Zookeeper.
- Contributed towards building **Apache Spark** applications using **Python**.
- Developed ETL pipelines in & out of the data warehouse using a mix of **Python** & **Snowflakes**, **SnowSQL** Writing SQL queries against Snowflake.
- Worked on architecture and components of Spark, and efficient in working with **Hadoop Core**, **Spark SQL**, **Spark streaming** and expertise in building **PySpark** applications for interactive analysis, batch processing and stream processing.
- Worked on Airflow for orchestration and familiar with building custom Airflow operators and orchestration of workflows with dependencies in **Azure**.
- Configured Spark Streaming to receive real time data from the **Apache Kafka** and store the stream data to **HDFS** and expertise in using **Spark** with various data sources like **JSON**, **Parquet,** and **Hive**.
- Worked with the **Map Reduce** programming paradigm & the Hadoop **Distributed File System**.
- Integrated **Hadoop** with **Tableau** to generate visualizations like **Tableau Dashboards**.
- Expertise in all aspects of the **Software Development Life Cycle** (**SDLC**), including **Agile** & **Waterfall** techniques.
- Excellent Communication skills, Interpersonal skills, problem-solving skills & being a team player.
- Ability to quickly adapt to new environments & technologies.

TOOLS AND TECHNOLOGIES:

Big Data Ecosystem	HDFS, MapReduce, Hive, Pig, Sqoop, Flume, HBase, kinesis, Impala, Stream sets, Spark, Zookeeper, Airflow.
Hadoop Distributions	Apache Hadoop 1x/2x, Cloudera CDP, Hortonworks HDP
Languages	Python, Pig Latin, HiveQL, Shell Scripting.
Software Methodologies	Agile, SDLC Waterfall.
Design Patterns	Eclipse, Net Beans, Pyspark**,** IntelliJ, Spring Tool Suite, Jenkin's, Kubernetes, Docker
Databases	MySQL, Oracle, DB2, PostgreSQL, SSIS, DynamoDB, SQL SERVER, Teradata.
NoSQL	HBase, DynamoDB, MongoDB
ETL/BI	PowerBI, Tableau, Snowflake, Informatica, Dax, SSRS, SSAS, QlikView, Qlik Sense.
Version control	GIT, SVN, Bitbucket.
Operating Systems	Mac OS, Windows, Linux, Unix
Cloud Technologies	Amazon Web Services, Microsoft Azure

PROFESSIONAL EXPERIENCE:

MicroStrategy **Aug 2021 – Till Date**
AWS Data Engineer

Roles & Responsibilities:

- Working on importing the data from various data sources & applying various transformations using **AWS EMR** & then loading data into Hive tables or **AWS S3** buckets.
- Worked on **Data flow** diagrams and defined SLAs for batch provisioning of data across multiple applications.
- Used **AWS Athena** to import structured data from **S3** into other systems such as **RedShift** and to generate reports.
- Develop pipelines for migrating the Data from Oracle DB to **AWS Data Lake**, using the **Glue** and **Lambda** necessarily.
- Created **Apache presto** and **Apache drill** configurations on an **AWS EMR** (Elastic Map Reduce) cluster to integrate different databases.
- Managed various capacity planning graphical reports using **Python**.
- Developed ETL Processes in **AWS Glue** to migrate data from external sources including **S3**, **ORC**, **Parquet**, Text Files into **AWS Redshift**
- Wrote Lambda functions in **Python** for **AWS Lambda** and invoked python scripts for data transformations and analytics on large data sets in **EMR** clusters and **AWS kinesis** data streams.
- Proposed and implemented improvements to increase process efficiency and effectiveness, providing input to solution designs to ensure consistency, security, and fault tolerant AWS solutions.
- Designed both 2NF & 3NF data models for OLTP systems and dimensional data models using star and snowflake Schemas.
- Experience in implementing CI/CD processes using Jenkins, Bit bucket Pipelines, Elastic Beanstalk.
- Processed unstructured data in Hadoop big data platform & loaded data sets from various environments.

- Implemented advanced **AWS EMR** procedures like text analytics & processing using the in-memory computing capabilities.
- Designed and developed **airflow DAGs** in Python using different airflow operators.
- Developed ELT pipelines on **Apache Airflow** using **python**.
- Monitored the SQL scripts & modified them for improved performance using **PySpark SQL**.
- Implemented advanced **EMR** procedures like text analytics & processing using the in-memory computing capabilities.
- Implemented **AWS EMR** best practices like partitions, caching & checkpointing for faster.
- Wrote jobs for processing unstructured data into structured data for analysis, pre-processing, and ingesting data.
- Use **AWS Kibana** to create charts from the data derived.
- Experience on technical leadership and mentoring other engineers for best practices on requirements for project.
- Experience in working with business and developers to groom elaborate user stories.
- Experience of working in Agile scrum environment.

Environment: AWS EMR, AWS Glue, Redshift, S3, AWS Athena, MapReduce, Python, Pyspark, Shell scripting, Linux, MySQL, Oracle Enterprise DB, Jenkins, Git, Tableau, Soap, & Agile Methodologies.

Tetrasoft **May 2020– July 2021**
Azure Data Engineer

Roles & Responsibilities:

- Worked with data transfer from **on-premises SQL servers to cloud databases** (Azure Synapse Analytics (DW) & Azure SQL DB).
- Migrated On-Premises Databases namely SQL Server, Oracle to Azure SQL Databases using Various methods like **Azure Migration services** (Data migration Assistant), **Azure Data Sync, Azure Data Factory, SSIS, BACPAC and Restore**.
- Utilized Azure Logic Apps to schedule and automate batch jobs and to build workflows by integrating apps, ADF pipelines, and other services using email triggers, HTTP requests etc.
- Used Pipelines to extract, transform, & load data from a variety of sources including **Azure SQL**, **Blob storage**, **Azure SQL Data warehouse**, write-back tool, & reverse.
- Analyzed large amounts of unstructured data sets from various sources to determine the optimal way to aggregate & report on it.
- Ingestion of data into one or more Azure Services (**Azure Data Lake, Azure Storage, Azure SQL, Azure DW**) & processing of data in Azure Databricks.
- Built **Hive tables**, loading with data, and writing **Hive queries,** which will run internally in **MapReduce** way.
- Used Hive to analyze the partition & bucket data & compute various metrics for reporting.
- Designed & customized data models for Data warehouse supporting data from multiple sources in **real-time**
- Worked on building the ETL architecture & Source to Target mapping to load data into the Data warehouse.
- Worked on creating Pipelines in ADF using Linked Services/Pipelines/Datasets to Extract, Transform and load Data from various sources like Blob storage, **Azure SQL**, **Azure SQL Data warehouse**, write- back tool and backwards.
- Developed mapping parameters & variables to support SQL override.
- Created mapplets to use in different mappings.
- Developed mappings to load into staging tables & then to Dimensions & Facts.
- Coordinated with team member to resolve any technical issue, Troubleshooting, Project Risk & Issue identification, and management.

Environment: Azure Cloud, Azure HDInsight, DataBricks (ADBX), CosmosDB, Azure SQL Server, Azure Data Warehouse, MySQL, Azure DevOps, Azure AD, Azure Data Lake, Git, Blob Storage, Data Factory, Data Storage Explorer, Spark v2.0.2, Airflow, HBase

Sun Pharmaceutical Ltd. **April 2018 – April 2020**
Big Data Engineer/ Hadoop Developer

Roles & Responsibilities:

- Developed **Impala** queries for faster querying and perform data transformations on **Hive** tables.
- Worked on creating **Hive** tables and written **Hive queries** for data analysis to meet business requirements and experienced in **Sqoop** to import and export the data from **Oracle** & **MySQL**.
- Executed scripts and indexing strategy for a migration to Confidential **Redshift** from **SQL Server** and **MySQL** databases.
- Wrote MapReduce code to parse the data from various sources & stored parsed data into **Hbase** & **Hive**.
- Developed Hadoop cluster using Hortonworks with **Pig**, **Hive**, **HBase** and **Spark**.
- Executed scripts and indexing strategy for a migration to Confidential **Redshift** from **SQL Server** and **MySQL** databases.
- Wrote **MapReduce** code to parse the data from various sources & stored parsed data into **Hbase** & **Hive**.
- Used **Python** to extract, transform & load source data from transaction systems, generated reports, insights, & key conclusions.
- Created Managed, Unmanaged Tables, Views, Cache and Used **Hive** meta store to persist tables.
- Developed **Sqoop** occupations for information ingestion, steady information loads from RDBMS to Snowflake.
- Designed data models for dynamic & real-time data used for various applications with **OLAP** & **OLTP** needs
- Used **Hive** to analyze the partitioned and bucketed data and compute various metrics for reporting on the dashboard.
- Working on **SSIS**, **T-SQL**, **SSMS**, and **SQL** Server Database.
- Created **SSIS** packages to load data into Data Warehouse using **SSIS** Tasks like Data flow task, Execute SQL Task, file system task, script, send mail and XML tasks.
- Experienced in performing in memory batch processing using **Spark** streaming (**Spark**, **Spark-SQL**, and **Spark Shell**)
- Experienced in writing **Python** as **ETL** framework & **PySpark** to process massive amounts of data daily.
- Implemented Spark to migrate **MapReduce** jobs into Spark RDD transformations and Spark streaming.
- Tested **Apache Tez** an extensible framework for building high performance batch and interactive data processing applications, on Pig and Hive jobs.
- Responsible for continuous monitoring and managing the **Hadoop Cluster** using **Cloudera** Manager.
- Involved in building the runnable jars for module framework through **Maven** clean, Maven dependencies

Environment: Cloudera CDH, Linux, HDFS, MapReduce, Shell Scripting, Talend, Hive, Pig, Spark, Strom, Sqoop, Flume, Oozie, Kafka, Apache Tez, Talend, Yarn, Maven

Kotak Mahindra Bank **Aug. 2016 – Mar 2018**
Data Analyst

Roles & Responsibilities:

- Performed data analysis & developed complex **SQL** queries based on requirements to generate data for mock-up reports.
- Collaborated with Business analysts and architects to assemble the prerequisites and implement the designing work to make final data sets.
- Developed forms, reports, queries, macros, VBA code and tables to automate data importation and exportation to a system created in MS Access.
- Used SQL queries for organizing and abstracting data from **MS access** databases, created reports, forms on MS Access.
- Created Dashboards & reported deliverables in **Tableau**, utilized advanced features, capabilities, & designs.
- Created **SSIS** packages for Uploading of different formats of files (Excel, Access, dbf) and databases (**SQL server**, Flat files) into the **SQL Server** data warehouse using **SQL Server** Integration Services (**SSIS**).
- Worked on relational databases and writing SQL queries to extract data and compare to the test result to the expected result.

- Experience in **ETL** processes using various tools such as **SSIS**, **PowerCenter Informatica**, **MS Access** and **SAS Data management**.
- Implementing **ETL** (extract, transform & load) in **SAS** to import data from multiple sources like **Mainframe**, FLAT files, spreadsheets to perform data analysis, validations & build tabular reports.
- Created **Pivot tables** using Excel Pivot table, **VLOOKUP** & other excel functionalities are utilized to analyze & generate reports.
- Conducted GAP analysis, SWOT analysis, Cost-Benefit Analysis and ROI analysis.
- Developed **SAS** Program for Converting Large volume of Text File into Teradata Tables by importing the text file from Mainframes to Desktop.
- Experience in **Tableau visualization** & built interactive dashboards and generating reports with lines, bars, pies, heat maps, scatter plots, bubbles etc. according to the business requirements.

Environment: PLSQL, Python/R, MS Project, RDBMS, MS Assess, Informatica, Load Runner 8.x, Tableau, JIRA, Microsoft SQL Server, Reporting Requirements, Oracle, Agile & Scrum methodology, MS Office, MS Project.

ACT Fibernet **Sep. 2014 – Jul 2016**
SQL Developer

Roles & Responsibilities:
- Written complex **SQL queries** using **joins**, sub queries and correlated sub queries.
- Wrote **SQL** Queries for cross verification of customer order data.
- **Transformer**, **Lookup**, **Merge**, **Join**, **Aggregator**, **Sort**, **Filter** and Remove **duplicates** stages were extensively used to Cleanse/Transform the data extracted into staging.
- Developed data warehouse process models, including sourcing, loading, transformation, and extraction of source data.
- Worked on **SSIS** and extract transform load (**ETL**) tool of **SQL Server** to populate data from various data sources, creating packages for different data loading operations for applications.
- Developed and deployed **SSIS** packages, configuration files, and schedules job to run the packages to generate data in CSV files.
- Data Automation Using **SSIS** to automate data processes for input of information.
- Worked with subject matter experts to prepare source to target mapping document for the migration project.
- Involved in Installation, Configuration and Deployment of **SSRS** Reports using different topologies.

Environment: MySQL, PostgreSQL, SISS, SSRS, SQL Server, ETL

EDUCATION:

- Bachelor of Engineering, Computer Science, Jawaharlal Nehru Technological University, India