Manideep

Kabilan.Kempraj@mechlintech.com

442-278-5514

SUMMARY

- Senior Data Engineer with 9+ years of experience in building data intensive applications, tackling challenging architectural and scalability problems, managing data repos for efficient visualization, for a wide range of products.
- Experience on Migrating SQL database to Azure data Lake, Azure data lake Analytics, Azure SQL Database, Data Bricks and Azure SQL Data warehouse and controlling and granting database access and Migrating On premise databases to Azure Data Lake store using Azure Data factory.
- Experience in Developing Spark applications using Spark - SQL in Databricks for data extraction, transformation and aggregation from multiple file formats for analyzing & transforming the data to uncover insights into the customer usage patterns.
- Good understanding of Spark Architecture including Spark Core, Spark SQL, Data Frames, Spark Streaming, Driver Node, Worker Node, Stages, Executors and Tasks.
- Good understanding of Big Data Hadoop and Yarn architecture along with various Hadoop Demons such as Job Tracker, Task Tracker, Name Node, Data Node, Resource/Cluster Manager, and Kafka (distributed stream-processing).
- Experience in Database Design and development with Business Intelligence using SQL Server 2014/2016, Integration Services (SSIS), DTS Packages, SQL Server Analysis Services (SSAS), DAX, OLAP Cubes, Star Schema and Snowflake Schema.
- Excellent communication skills with excellent work ethics and a proactive team player with a positive attitude.
- Domain Knowledge of Finance, Logistics and Health insurance. Strong skills in visualization tools Power BI, Confidential Excel - formulas, Pivot Tables, Charts and DAX Commands.
- Expertise in various phases of project life cycles (Design, Analysis, Implementation and testing). Led database administration and database performance tuning efforts to provide scalability and accessibility in a timely fashion, provide 24/7 availability of data, and solve end-user reporting and accessibility problems.

Technologies:

- Azure SQL database, Neo4J/Azure Cosmos Graph, Cassandra, MongoDB/Azure Cosmos, SQL Analytics, Tableau, Docker
- Microsoft SQL server, MySQL, Cosmo DB, Azure Data Lake, Azure blob storage Gen 2, Azure Synapse, IoT hub, Event hub, data factory, Azure data bricks, Azure Monitor service, Machine Learning Studio
- Programming Scala, Python, Spark SQL
- MSBI (SSIS, SSAS, SSRS)
- Data Visualization
- Data Migration
- SQL Server programming
- Confidential Power BI
- Analytic Problem-Solving

PROFESSIONAL EXPERIENCE

Wells Fargo, Charlotte-NC   **Mar 2015– Current**

Role: Lead Data Engineer

Wealth & Investment Risk:

- Analyze, design and build Modern data solutions using Azure PaaS service to support visualization of data. Understand current Production state of application and determine the impact of new implementation on existing business processes.

- Extract Transform and Load data from Sources Systems to Azure Data Storage services using a combination of Azure Data Factory, T-SQL, Spark SQL and U-SQL Azure Data Lake Analytics. Data Ingestion to one or more Azure Services - (Azure Data Lake, Azure Storage, Azure SQL, Azure DW) and processing the data in In Azure Databricks.
- Created Pipelines in ADF using Linked Services/Datasets/Pipeline/ to Extract, Transform and load data from different sources like Azure SQL, Blob storage, Azure SQL Data warehouse, write-back tool and backwards.
- Developed Spark applications using Pyspark and Spark-SQL for data extraction, transformation and aggregation from multiple file formats for analyzing & transforming the data to uncover insights into the customer usage patterns.
- Responsible for estimating the cluster size, monitoring and troubleshooting of the Spark databricks cluster. Experienced in performance tuning of Spark Applications for setting right Batch Interval time, correct level of Parallelism and memory tuning.
- To meet specific business requirements wrote UDF's in Scala and Pyspark. Developed JSON Scripts for deploying the Pipeline in Azure Data Factory (ADF) that process the data using the Sql Activity.
- Hands-on experience on developing SQL Scripts for automation purpose. Created Build and Release for multiple projects (modules) in production environment using Visual Studio Team Services (VSTS).

Options Risk/P&L:
- Developed mapping document to map columns from source to target.
- Created azure data factory (ADF pipelines) using Azure polybase and Azure blob. Performed ETL using Azure Data Bricks. Migrated on-premise Oracle ETL process to Azure Synapse Analytics.
- Worked on python scripting to automate generation of scripts. Data curation done using azure data bricks.
- Worked on azure data bricks, pyspark, hdinsight, u-sql, t-sql, spark sql, Azure ADW and hive used to load and transform data.
- Used Azure Data Lake as Source and pulled data using Azure Polybase.
- Used stored procedure, lookup, execute pipeline, data flow, copy data, azure function features in ADF. Experienced in azure Data Lake storage gen2 to store excel files, parquet files and retrieve using data using Blob API.
- Used azure data bricks to do ETL (workspace, hive, rdd, map, reduce so on). Worked on creating star schema for drilling data. Created pyspark procedures, functions, packages to load data.

Agency Risk/P&L:
- Migrated big data applications from on-premises to Azure for scalability.
- Data transition to Azure Data Lake storage using secure REST calls.
- Transitioned real-time data using Kafka from on-premises to Azure and processed it using Spark in Databricks.
- Active participant in Business meetings, Requirement gathering, Architecture Design & Development.
- Cleansed and analyzed huge datasets in Azure using Scala and Spark SQL.
- Experience in setting up networking using VNET, subnet, NSG with Terraform and Power Shell.
- Extensive hands-on SQL experience working with huge data sets in Spark to analyze and visualize the data.
- Automated all the deployments using Terraform and DevOps tools like Git, Jenkins, JFrog, SonarQube and Azure Data Factory.

Credit Derivative:
- Data transition involving import and exports from multiple RDBMSs like Oracle and Teradata to Big Data technologies like Hive, HBase using Sqoop automation using Shell scripting.

- Data cleaning using Spark Scala and Spark SQL with RDDs and Data frames to analyze datasets for potential data issues.
- Loaded data into Hive as a presentation layer on top of HBase for Analysts/Analytics.
- Developed shell scripting for scheduling daily Sqoop jobs for data transitions.
- Automated direct hot fixes on Production using CI/CD pipeline (Git, Jenkins, JFrog and Fortify).
- Developed MapReduce jobs to parse raw data into tabular format and loaded it into HBase tables.
- Query optimization in Hive by modifying legacy table structure using Partitioning, Bucketing and Compression techniques.
- Redesigning legacy SQL functions into SparkSQL and HiveQL for fast and effective processing.
- Created Hive external tables to store JSON and XML files using Serializers/Deserializers (JsonSerDe, XPath)
- Active participation in daily scrum calls, business meetings, requirement gathering, test case development and job monitoring.

All Management Corporation, Los Angeles-CA   Aug 2013– Feb 2015
Role: Sr. Application Data Engineer
- Requirement gathering, business discussions and retrospectives meetings to provide quality output.
- Code optimization using Spark to reduce runtime and enhance performance.
- Job log transition through Kafka data pipelines; Analyzed the logs using Spark Sink connectors to Kafka and Visualized using Grafana to identify potential failures.
- Migrated legacy R programs to Spark framework using Scala for Data cleaning and Identifying fraud transactions.
- Scheduled and monitored daily jobs in Tivoli workspace for automating daily execution.
- Mentored and trained users to speed up and bring them aligned with modernization efforts.
- Tableau report generation for business review and validation.
- Documented and developed reusable Spark code templates and made them available in GitHub to initiate and the migration efforts across the organization.
- Implemented DevOps (CI/CD) model to automate traditional deployment process.

EDUCATIONAL QUALIFICATIONS:
- *Master of Science in Computer Science and Computer Engineering, University of Arkansas, Fayetteville.*