VIKASH KUMAR JANGID

DATA ENGINEER

Delhi, India 🏠
+91-998-219-2566 📞
vikash.kr.jangid@gmail.com ✉
www.linkedin.com/in/vikash-kr-jangid/ 🔗

PROFESSIONAL PROFILE

- A software professional with 2.3 years of experience in Design, Coding, Testing and Production management around azure big data space.
- Expertise in Multiple Bigdata Technologies i.e. Spark Scala, Kafka, Hortonworks Data platform, Databricks Unified Analytics Platform, Azure SQL DWH, Cassandra, Sqoop
- Extensively worked on Data extraction, Transformation, Cleansing, Loading, Real Time Analytics and Reporting the output data.
- Experience in tuning mappings, identifying and resolving performance bottlenecks in various levels like sources, targets, mappings and sessions.
- Better Understanding of Apache Spark and its managed version Azure Databricks
- Domain experience: Worked in the Media and E-Commerce domain

IT EMINENCY

- **Programming Languages:** Scala, Python
- **Big Data Framework:** Databricks (AZURE and AWS), Spark, Hadoop, HDInsights, Data Lake, Apache Beam
- **Data Pipeline Tool:** Azure Data Factory, Air Flow
- **Streaming Ingestion Tools:** Azure Event Hubs, Kafka
- **Visualization Tools:** Microsoft Power BI
- **Microservices:** Kubernetes, Docker, Azure Service Fabric
- **Azure Cloud Services:** Azure Logic App, Azure Functions

EXPERIENCE

Celebal Technologies-Jaipur
Client | Media Company
MAR,2019 – PRESENT
Developed and design master data management (MDM) by integrating TV viewership data from data source BARC with revenue generated by the advertisers from source Broadview.

Responsibilities:

- Data extraction from BARC priority application i.e Barc Work Station.
- Setting up of distributed data extraction. Automating data extraction by Mafras tool and windows batch script.
- Created Scala-Spark jobs for implementing uses case such as 'comparison with competition', 'channel performance', 'market overview supply and demand' and 'company performance'.
- Implemented Spark Dataframes and applying transformations on Azure Databricks Runtime 6.0, actions to implement business analysis.
- Analyzed the SQL scripts and designed the solution to implement using Scala on Azure Cloud.
- Involved in converting Hive/SQL queries into Spark transformations and actions using spark SQL (RDDs and Data frames) in python and Scala.

Client | Media Company
JAN,2019 – MAR,2019

Developed and design end to end data pipeline for analyzing user behavioral data of digital video-on-demand streaming service and AVOD OTT platform - with 100+ billion minutes of watch time and 100 million monthly active users within the fiscal year. Performed use cases user consumption, user split and user retention.

Responsibilities:
- Taken care of the initial system architecture design and worked at client-side(Mumbai) for 6 months
- Created Scala-Spark jobs for data ETL and aggregation
- Troubleshoot and debug any spark ecosystem run time issues.
- Data orchestration pipeline using azure data factory.

Celebal Technologies-Jaipur
Client | Media Company
MAY,2018 – DEC,2018

Performed data engineering and analysis for largest media company headquartered in Mumbai. My role was to develop data pipelines. we get data from many different data sources like Youbora, mixpanel.com, Simplify360.com , appsflyer.com, webdunia.com, sql-database, sql-data warehouse. we process 7 TB data on daily basis. and optimize all the queries in Apache spark with caching and tuning spark. also, we are using parquet format with gzip codec compression to compress the daily 2TB data. after processing the data, we are refreshing all the dashboard through Hive tables on power BI.

Responsibilities:
- Taken care of the initial system architecture design and worked at client-side(Mumbai) for 6 months
- Loading data from different datasets including APIs and Data warehouse and deciding on which file format is efficient for a task.
- Understanding the requirements of input to output transformations
- Enabled the reporting on Power BI Dashboard
- Created Scala-Spark jobs for data ETL and aggregation
- Troubleshoot and debug any spark ecosystem run time issues.
- Data orchestration pipeline using azure data factory.
- Create Serverless application using azure function for some use cases in project.

Celebal Technologies-Jaipur
Client | IT recruitment & staffing services agency
JAN,2018 – APRIL,2018

Developed a stateless and stateful web application to package, deploy, and manage scalable and reliable microservices and containers. Azure Service Fabric.

Responsibilities:
- Developed reliable Stateless frontend using angularjs and Stateful backend using C#
- Setting up a Service Fabric cluster (Linux or Windows) in Azure cloud.
- Monitoring your live web application an extensible Application Performance Management (APM) service by Azure Application Insights.
- Real time log Logs Analytics by integration Azure service fabric with Azure Log Analytics.

EDUCATION

B.Tech(Hons) in Computer Science & Engineering | Swami Keshvanand Institute of Technology, Jaipur
2014 – 2018

ACTIVITIES
- Delivered Databricks training in pyJaipur, Databricks Meet up and workshop for client.
- AZ-900: Microsoft Azure Fundamentals Certified.
- 1st in Asia to productize Databricks in large scale.
- 2nd Runners up in Area Level Humorous Speech Contest at Toastmasters International