Supriya Badgujar

+91-7709264432

badgujar.supriya6@gmail.com
https://www.linkedin.com/in/badgujarsupriya6/

SYNOPSIS

Google Cloud Certified Professional Data Engineer and **AWS certified Developer Associate** having nearly 8 years of industry and full life cycle experience in a Hadoop big data cloud-based environment, along with exceptional analytical, design and problem-solving capabilities. Looking for an ambitious company which will challenge my development and problem-solving skills and allow me to continue to develop my knowledge and potential.

EXPERIENCE SUMMARY

Nearly 8 years of overall IT experience in a variety of industries (CRM and Retail domain), 3+ years in Big Data, Hadoop and its components like MapReduce, Hive, Kafka, Pig, Sqoop, 2 years of experience in Spark (including Spark Core, Spark Streaming, Spark SQL, PySpark) and Scala, 3+ years of experience in Core Java, 3+ years in Google Cloud Platform (GCP), 2 years in python (pandas, numpy, flask), 2+ years in shell scripting, 6 months in ElasticSearch, 6+ months in AWS
Experience in analysing data using HIVEQL and extending Hive, BQ core functionality by using custom UDF's
Experience in Manipulating/Analysing structured and unstructured data by Using Spark-Scala
Hands on experience in data analysis using MapReduce programs in Java. Extensive experience on handling real time data using Kafka. Used Spark-SQL for processing RDBMS data on Hadoop. Hands on experience on Core Java and Python.

- Being a speaker, **delivered global virtual webinars** on Google Cloud topics in communities and groups.
- Won recognition in **'Commitment to Work'** category for April and July 2020 and **Teamwork Values'** category for May 2020 in Sears Holdings, India
- Won **'GCP Guru of the Month'** recognition for the month of Jan 2020 from acloudguru
- Won **'Rising Star'** award for Quarter IV- 2015 in Serene Corporation, Pune in Feb 2016
- Completed Google Cloud Certified Professional Data Engineer certification in Dec 2018
- Completed AWS certified Developer Associate certification in 2018
- Completed Microsoft Azure fundamentals az-900 certification in 2020
- Completed Scrum Certified Professional Certificate (SFPC) in Jun 2020
- Earned badges from IBM like 'Big Data Foundations', 'Big Data Hadoop Foundations', 'Big Data Spark Foundations', 'Big Data - Programming', 'Big Data - Analytics', 'Big Data Storage and Retrieval' by completing Big data, Hadoop, Spark, NoSQL courses in July 2016

PROJECTS SUMMARY

Project: Audience Currency and Measurement (ACM)
Company: Decision Culture
Client: The Home Depot
This project focuses on building and experimentation and measurement framework that aims at expanding the reach and engagement by identifying the most important customers. It is critical in reaching the right person at the right time with right message and enables understanding marketing measurement at multiple levels.
Technical Platforms included are Google Cloud Platform (BigQuery, Dataproc, GCS, Composer), Python, PySpark, SQL
Roles and Responsibilities:
- finding the pitfalls & limitations with the existing pipeline and suggesting the better approaches to improve the pipeline and implementing the same. updating the existing jobs based on the business requirements
- identifying the tasks that consume more time and then modifying the existing queries and code
- restructuring the existing pipeline as it takes long time to execute by breaking the pipeline into multiple pipelines or parallel tasks by creating/updating the run log/ driver table entries to boost the performance
- configuring regions, fiscal week, communities, run mode, tables in configuration yml file and using them
- perform measurement code clean up, modularization, automation wherever possible
- creating cost optimization pipeline to delete unnecessary output bigquery tables generated through the pipeline
- designing, developing pyspark and python jobs, testing the jobs with some different scenarios
- Identifying the production issues, debugging and resolving. Mentoring and documentation

Project: Targeted Interaction
Company: Sears
This project focuses on generating the real time scores of the members of Sears and Kmart stores in order to comprehend the interest of the member in the specified categories based on their browsing. It uses the clickstream data to get the browsing

Technical Skills:

Big Data:
- Hadoop
- Spark
- Hive
- Kafka
- MapReduce
- Azkaban

Cloud:
- GCP (Dataproc, GCS, BigQuery, GCE, Pubsub, BigTable, Dataflow, Airflow, Cloud Vision API, Cloud Speech To Text API, Cloud Shell)
- AWS (S3, EC2, Kinesis, EMR)

Languages:
- Core Java
- SQL
- Scala for Spark
- Python

Databases:
- MySQL
- Oracle

Other Tools/ Utils/ Technologies:
- Linux/Unix
- Maven
- SVN/ Git
- ElasticSearch (*)

Web Server:
- Tomcat

IDE/ Editors:
- PyCharm
- Eclipse
- IntelliJ
- STS

Note: *basics

patterns of the members and generates the score by feeding that data to RTS machine learning model. Later on, dynamic emails with offers are sent by marketing team to those members based on these scores, generating their interest in buying those category items thereby contributing to increasing revenue of the company through this analysis.

Technical Platforms included are Google Cloud Platform (BigQuery, BigTable, Pubsub, GCE, Dataflow, GCS, Airflow, Cloud Source Repositories), Python, Java, Shell scripting, Adobe Audience Manager, Adobe Campaign

Roles and Responsibilities:

- Creating/Updating batch and streaming pipelines
- Updating all GCP projects to support the latest google cloud sdk versions
- Analysing the clickstream data present in bigquery and generating the real time scores per members per category. Combining those scores and running these jobs weekly through Cloud Composer.
- Understanding the most expensive GCP components in the project and optimizing the cost of all the GCP projects involved during the period when stores were closed.
- Automating the process of cost optimization by writing shell scripts wherever necessary.
- Finding the PII data from all the bigquery tables and securing it. Creating a process to encrypt loyalty ids of members.
- As part of Adobe Audience Manager, creating signals, traits and segments for multiple categories in order to identify the browsing behaviour of customers.
- As part of Adobe Campaign, mapping the segments created in the Audience Manager to audiences in the campaign in order to find out actual customers based on the segment categories. For Adobe Campaign Mart, maintaining a data lake in BigQuery and coding the same with Google Cloud Source Repository. Creating import and export jobs from bigquery to Adobe Campaign ACC/ ACS servers and vice versa. Scheduling those jobs using airflow
- For Preference Centre, creating batch/streaming pipelines which accepts member preferences available in Pubsub topics, process them using Dataflow and store the results in the BigQuery tables.

Project: Cloud Enablement
Company: EY
This project focuses on developing or migrating available one or more of the services and infrastructures to create a public cloud computing environment.

Technical Platforms included are Amazon Web Services (AWS), Google Cloud Platform (Google App Engine GAE, BigQuery, Pubsub, GCS, Cloud Vision API, Cloud Speech to Text API, Cloud Shell), Twitter API, Python, Flask, HTML

Roles and Responsibilities:

- Developed platforms and applications that were hosted on cloud for analytics team using Google App Engine (GAE), BigQuery, Pubsub, Cloud Vision API, Cloud Speech to Text API, GCS, Python, Flask, HTML
- Created POVs, POCs and proposals for multiple clients
- Consulted and suggested the best cloud provider and migration approach fulfilling the requirement of client
- Designed the architectures for the different use-cases
- Conducted Google Cloud training in multiple Indian locations

Project: Saleshub (Client: Kohl's, USA)
Company: Datametica
Top e-commerce firm wanted to analyse their sales transactions in the form of xml coming from IBM message queue to gain a deep understanding of customer shopping habits, social media use, web advertisement effectiveness and other metrics that inform business decisions.

Technical Platforms included are GCP (Google Cloud Storage, Dataproc, BigQuery, Bigtable, Dataflow, Compute Engine, Pubsub, Composer), Kafka, Hadoop, Pig, Hive, Sqoop, Spark, MapReduce, Java, Maven, VMware Horizon Client, Putty, WinSCP, MobaXTerm

Roles and Responsibilities:

- Creating the pipeline to ingest the day-to-day basis transactions coming through IBM MQ in the form of XML files through Kafka into GCS by parsing the XML files and mapping the fields (primitive and complex data types) with the gold table using JSON schema file through Spark. Store the data frame in GCS, Hive, Big Query and Bigtable. Converting current pipeline written into MapReduce, pig, hive into spark using spark core, sql and streaming modules.
- Extended the current pipeline to store the data from kafka to Bigtable instance
- Converted the existing Teradata SQLs into BigQuery and scheduled them using Composer (Airflow)
- Used Spark-SQL for processing RDBMS data on Hadoop.Generated the Schema for the Unstructured and Structured Files in the JSON. Made the ingestion code generic so that it will be used in all other modules
- Did POC's in Dataflow and Pubsub for data cleansing and ingestion tasks as an alternative basis
- Identifying and fixing production issues in current pipeline and Performance Tuning of Hive and Hadoop

Project: Discover
Company: ITyX

Discover is a web-based business reporting and analytics product which helps in monitoring key topics and trends in real-time from any source, transaction, teams or applications by context-relevant analysis of content and connected activities. It creates predefined basic reports and user definable widgets for the visualization of connections and relationships which determines the problems and derive measures to improve customer satisfaction and process efficiency.

Technical Platforms included are Java, J2EE, Elastic Search, Web Services (REST), MySQL, Spark, WinSCP, Git, Docker, AWS (S3, EMR and Kinesis)

Roles and Responsibilities:

- Designing dashboards, widgets on basis of Client Requirement
- Dynamic Query builder for Customization of Reports
- Sharing & collaboration of Reports Among the Group in an Organisation
- Generating Report Multiple database connections of Products
- Integration with other applications Using Web Services

Project: Legacy System to Oracle Sales cloud Migration (Client: HDR, USA)
Company: Serene Corporation (now AST Corporation)
This project briefly can be described as implementing a set of projects that includes developing a customized web application, an integration (data flow) process between Client database and CRM, developing a web service to expose it on Oracle Sales Cloud for Quote object.
Technical Platforms included are Java, Spring (core and MVC), Hibernate, SOAP, MySQL, Oracle database, Oracle Sales Cloud, RightNow CRM, JSP, Servlet, Secure CRT, ANT, Citrix Receiver, XML, JavaScript, AngularJS.
Roles and Responsibilities:

- Understanding, analysing and reviewing business requirement
- In customized application, developing a functionality to generate xml on click on 'single download' button as well as 'Export All' button as per the XSD and the fields selected in the JSP
- Designing and developing an integration process between JDE and OSC, clarify database, staging MySQL database and RightNow CRM, developing a job to create CSV files for all CRM objects for data warehousing, implementing a job to send out emails to Sales representative for failed records, expose a web service to invoke OSC Business Intelligence web service to fetch Quote TC reports and sending those reports as an attachment in pdf format to the respective sales representative.
- Designing of simplified & classic JSP pages, developing payment history module for providing details about the payment history to end users, exporting the records to excel & pdf.
- Building and packaging of application, SIT, UAT, Support and enhancement, Issue analysis, Bug Fixing, Documentation, Testing, deployment on Java Cloud Services (JCS)

WORK EXPERIENCE

Sr. No.	Company Name	Years of Experience	Duration
1.	Decision Culture, Bangalore	3+ months	Sep 2021 to till date
2.	Sears Holdings, India, Pune	1.8 years	Jan 2020 to Sep 2021
3.	Ernst & Young LLP, Pune	5+ months	Aug 2019 to Jan 2020
4.	Datametica, Pune	1.5+ year	Feb 2018 to July 2019
5.	ITyX India Pvt. Ltd, Pune	1.4 year	Oct 2016 to Feb 2018
6.	Serene Corporation, Pune (AST Corporation)	2.8+ years	Jan 2014 to Sept 2016

EDUCATION

Degree	Institute	Grade
B E	Govt. College of Engineering, Jalgaon	77.73% (73.1% aggregate)
HSC	M.J. College, Jalgaon (Nasik Board)	83.00%
SSC	P.N.L. High School, Jalgaon (Nasik Board)	81.20%