Shantanu Parag Bugadi

Email : shantanubugadi007@gmail.com
Mobile No : +91 9860699883
: +919404861439

Professional Summary:-

- **GCP Certified Professional Data Engineer.**
- Having 3.1 years of IT experience.
- Work experience in Cloud and Big Data Analytics, Technologies
- **GCP technologies : Bigquery ,GCS ,VM instance,Dataproc,Cloud shell,CloudSql,Compute Engine. GCP, AWS ,Azure.**
- **Hadoop technologies : PySpark, Hadoop, Hive, Sqoop, HCatalog, Pig, Core Java,Python.**
- **Received "Hulk Award" for gigantic contribution to the project.**
- **Received "Spot Award" for Technical Analysis Skill.**
- **Received "Long Term Service Award" for 3 year completion in same company.**
- Worked extensively over semi-structured data (fixed length & delimited files), for data sanitation and standardization.
- Experience on Cloudera & Hortonworks Hadoop Distributions.
- Experience in pulling data from **Oracle, Netezza and Teradata** to Hadoop using Sqoop.
- Good knowledge on Java and Python programming languages.
- Sound knowledge in **Data Structure and Algorithm Design**
- Effective team member skills, inspired and motivated towards work.
- Quick learner and bendable enough to adapt any technology

Experience:-

- Working as Engineer 1 for DataMetica Solutions Pvt Ltd Pune from DEC 2017 to till date.

Technical Skills:-

Cloud Platform	GCP, AWS ,Azure, Bigquery, Dataproc, Compute Instance,Composer
Hadoop Ecosystems	PySpark,HDFS, Hive,Sqoop,MapReduce, Yarn, Pig
Big Data Framework	Hadoop, Spark
Scheduler	Azkaban, Autosys,Airflow,Composer,Control M
Programming Languages	Java (Core)
Scripting	Python,Shell
Operating Systems	Linux, Windows 7, Windows 10 ,Centos, Mac OS X
Databases	Netezza,Sql Server , Teradata,Oracle , MySQL
IDE & Editors	IntelliJ Idea, Eclipse, Notepad++

Education:-

DEGREE : B.E. E&TC
COLLEGE : GOVERNMENT COLLEGE OF ENGINEERING AND RESEARCH , AWASARI (PUNE)
YEAR : 2017
PERCENTAGE : 66.6% (Distinction)

Projects:-

5. **Project Name**	LIFEPOINT
Role	Engineer 1
Organization	Datametica Solutions
Team Size	20
Technologies	Azure Synapse, Azure DataFactory ,SQL Server

Project Description

Lifepoint being a health based client wanted to build Analytics over Microsoft Azure using Azure DataFactory
and Azure Synapse by pulling data from Sql Server

Contribution
- Complete ownership of new EMR development setting up pipeline ,created set up for 54 staging tables , developed 22 Stored Procedures , Written Business Validation Queries.
- Analysed the existing system in SQL Server and figured out the logic in Stored Procedure.
- Identified the reusability of the present SP while referring to the STM given by client .
- Written Stored Procedures for the new module in project in Synapse.

4. **Project Name**	SKY
Role	Engineer 1
Organization	Datametica Solutions
Team Size	15
Technologies	BigQuery,Python ,Shell script, Compute Instance

Project Description

This project is to build analytics over GCP cloud using Bigquery

Contribution
- Completed end to end Development (from STM to Data Validation) of Tuner Event Module.
- Sql scripts parser to detect the source and target tables from the sql script for initial analysis.
- Python script to validate file load in bq table comparing the file vs table records.
- Written generic code to load data from gs bucket to respective tables
- BQ auditing and housekeeping scripts

3. **Project Name**	RMG
Role	Engineer 1
Organization	Datametica Solutions
Team Size	12
Technologies	Datastage , BigQuery , Shell script, Compute Instance

Project Description

This project is regarding migration of datastage (Teradata) to GCP cloud (Big Query)

Contribution
- Written a generic code to load data from gs bucket to bq table
- Converted bteq scripts to BQ compatible and wrapper for auditing and error handling

● Written python script for merging XSD files

2. Project Name	Catalina
Role	Engineer 1
Organization	Datametica Solutions
Team Size	9
Technologies	Spark,Hadoop, Python ,Hive, Sqoop, Shell script

Project Description

This project is regarding ingestion of workload from Netezza to Microsoft Azure Datalake.It also consisted of file ingestion of several format like CSV ,Fixed length, Json .Creation of Elastic Cluster for the balancing the workload accordingly and allocating the nodes at runtime.

Contribution
● Written sqoop jobs (Ingestion wrappers).
● Written a python script for automatic DDL creation by hitting Netezza through sqoop eval.
● Written Json file ingestion scripts with validation(Pyspark).
● Hive Scripts for transformation from one layer to another.
● Automated sql queries run from hadoop to netezza with proper auditing

1. Project Name	GCP Cloud Migration
Role	*Trainee Engineer*
Organization	Datametica Solutions
Team Size	22
Technologies	Spark,Java, Pig, Hive, Shell Script

Project Description

The project was a lift and shift from On-Premises to GCP Cloud. It was a migration project with the refactoring of the whole code structure according to security standards.Making the code to be generic . Separating the variables in a config and the main scripts separately.

Contribution
● Migrated PII subject areas to GCP Cloud (DataProc Cluster)
● Parameterized the scripts to make it complete generic
● Written a python validation script which would validate the data after migration

Courses and Achievements: -

Course Name	Organizer
GCP Data Engineer Certification	GOOGLE Certification
Coursera GCP certification	Coursera Certification
Data Science Essentials	Cloudera University

Declaration

The information provided above is the best of my knowledge and contains no faltered, modified or fabricated aspects about my profile and qualifications

Place: Pune, India.   Shantanu Bugadi