Gaurab Dawn

Lead Consultant (Big Data, Cloud, Devops)

9 years of IT experience in Big Data and Cloud (Azure,AWS), and Devops currently working in ITC INFOTECH INDIA PVT LTD . Expert in Spark components like Spark Core Programming, Spark SQL, Spark Streaming and Big Data components like Hive, Pig, HBase, Kafka, Yarn, Sqoop, Flume. Expert in programming languages like Scala, Python, Unix Shell Scripting. Experienced in Microsoft Azure and AWS.

gaurabishere@gmail.com

7063276040

SKILLS

- Apache Spark
- Scala
- Python
- Azure Databricks / Azure HDInsight
- Azure Data Factory
- Kubernetes/Docker/Git
- Amazon WebServices
- SQL Server/ Oracle
- Unix Shell Script
- Hive
- SQL
- Java
- Azure DataWarehouse

LANGUAGES

English
Full Professional Proficiency

Bengali
Native or Bilingual Proficiency

Hindi
Full Professional Proficiency

WORK EXPERIENCE

Lead Consultant
ITC Infotech Pvt LTD
07/2018 – Present — *Bangalore*

Senior consultant
PricewaterhouseCoopers Pvt Ltd
11/2016 – 06/2018 — *Kolkata*

Technology Lead
Cognizant Technology Solution
05/2015 – 10/2016 — *Kolkata*

Sr. Software Engineer
Tech Mahindra Limited
04/2010 – 05/2015 — *Pune/London*

PROJECTS

Premier Insurance Brokers (UK) (ITC Infotech Ltd) (01/2019 – Present)

- Project is about building an ETL framework using Scala in Azure Databricks and Azure Data Factory. Azure Data Factory is used to bring in ten of thousands of table from source to Azure Datalake. Spark transformations using Databricks Cluster are used to transform data residing in Datalake.

British American Tobacco (ITC Infotech Ltd) (07/2018 – 12/2018)

- Petra application was built to enable Effective, Efficient Trade Marketing & Distribution (TM&D) with Speed to Market Solution for BAT. Spark on Azure HDInsight was used for data transformation. Spark transformations were developed using Scala in intellij idea IDE. Azure Data Factory was used for integration and data transfer.

Airlines Predictive Maintenance Tool (PwC Limited) (11/2016 – 06/2018)

- This project is aimed at making a product for predictive maintenance of Aircraft components. It would generate predictive alerts based on sensor data gathered from various Aircraft parts. The product also helps in Diagnostic of aircraft components using machine learning techniques.
- Extraction of machine log from airlines to azure datalake was done by Azure Data Factory. Transformation logics were written in scala over spark. Predictive model was developed in spark ML.

NXG Gen Device Analytics (Cognizant) (05/2015 – 10/2016)

- The project is to develop one utility for analysing any type of device log and predicting future errors based on the error/events occurred in the device. Following are major components of the solution a) Ingesting semi-structured Data: - For analysis and predicting errors, ingesting semistructure data generated from any device to Hadoop or UNIX file systems. b) Dynamic datamart generation.: - Loading of the ingested data to the dynamic datamart, generated at run time is one of the crucial aspect of the solution
- Pig script for reading the semi-structured data and loading them into HDFS. Hive script for extracting valid data from the Pig script output files. Integrating Pig and Hive scripts with HBase tables. Using Yarn Distributed shell and capacity scheduler for parallel execution of nonmapreduce jobs like shell scripts etc.

EDUCATION

Electronics & Communication Engineering
West Bengal University of Technology
07/2005 – 07/2009 — *76.6*