Gaurab Dawn

Email: gaurabishere@gmail.com **Ph: +917063276040**

PROFESSIONAL SUMMARY:

- Gaurab Dawn is a Sr. Consultant working in ITC INFOTECH INDIA PVT LTD with 8 years of IT experience in Big Data and Cloud based technologies.
- Expert in Spark components like Spark Core Programming, Spark SQL, Spark Streaming and Big Data components like Hive, Pig, HBase, Kafka, Yarn, Sqoop, Flume.
- Good Exposure in working in Microsoft Azure(HDInsight).
- Expert in programming languages like Scala, Unix Shell Scripting.
- Proficient in Databases like Azure SQL DW, Sql Server and Oracle.
- Expert in ETL tool Syncsort DMExpress and DMExpress Hadoop version.
- He has worked in Telecom and Media & Entertainment and Airlines Domain.
- He has 7 months of onsite experience in UK, working with Clients to mitigate issues and understanding new client requirements.

SKILLS SUMMARY:

Operating System	UNIX(HP UX 11, Solaris 10, Ubuntu 12.x, AIX), Windows
Languages	Scala, Python, Core Java, UNIX BASH Script, SQL, PL/ SQL
Databases	Oracle, Azure SQL DB, Azure DW
ETL Tools	DMX, DMX-h(Hadoop Version)
Hadoop Utilities	Spark, Kafka, Hive, Pig, HBase, Sqoop, Flume
Hadoop Environment	Cloudera, Hortonworks, Azure HDInsight
Cloud Environment	Microsoft Azure, Amazon Web Services

EDUCATION:

Degree	University	Year	Percentage
B.Tech	WBUT(B.P.Poddar Institute of Management & Technology)	2009	76.6 DGPA
HSC	WBCHSE	2004	76.9%
SSC	WBBSE	2002	81.1%

Employment History:

Organization	Location	Designation	Duration
ITC Infotech Pvt LTD	Bangalore	Lead Consultant	July 2018-Till Date
PricewaterhouseCoopers Pvt Ltd	Kolkata	Senior consultant	Nov 2016-June 2018
Cognizant Technology Solutions	Kolkata	Technology Lead	May 2015-Oct 2016
Tech Mahindra Limited	Pune/London	Sr. Software Engineer	April 2010-May 2015

Project Details:

Project Name: BAT Petra (ITC Infotech Limited)

Client: British American Tobacco

BAT has started the transformation programme, PeTra, to enable Effective, Efficient Trade Marketing & Distribution (TM&D) with Speed to Market. The PeTra programme will focus on TM&D Processes, Capabilities and KPIs as well as IT Solutions with below objectives.

- Clear and embedded Global TM&D Line of Sight: capabilities, competencies, dashboard, process (TM&D-X: CHECK / PLAN / ACT)
- Enable sustainability of TM&D model
- Evolve from transactional to a performance driven organization
- Back office complexities are removed from TM&D Systems – build and deployment will be easier
- Technology maturity i.e. SaaS, Cloud services, agile development, modern devices

Technology Used: Spark, Scala, Hive, Unix Shell scripting, Azure, ADF, Azure Powershell.

Contributions:

- Developing spark modules in scala using intellij idea tool based on business requirement.
- Writing Spark-SQL and spark core scripts to transform data.
- Improving application performance by tuning spark code and efficient use of spark framework through various performance improving techniques.
- Writing Hive scripts for transform, validation and analysis of data.
- Developing Azure ADF pipelines for integration and creation of ETL flow involving various application components.

Project Name: Airlines Predictive Maintenance Tool (PwC Limited)
Client: Famous US Airlines Company
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.
Product Video: https://www.youtube.com/watch?v=2gVEI_6a2tk&feature=youtu.be

Contributions:

- Design entire process flow with Azure PaaS/SaaS components.
- Extraction of machine log from airlines to azure datalake and writing transformation logics in scala over spark.
- Write Spark jobs in scala to prepare data for predictive model.
- Tuning scala scripts to enhance spark job performance.
- Creation of Hive table on top of csv and JSON data residing on Azure Datalake.
- Schedule Hive, Spark and SQL Store procedure activity via ADF.
- Create predictive model in spark ML with data science team.

Project Name: Patent data analysis (PwC Limited)
Client: Advisory service
The objective of this project was to analyze patent data in Azure cluster and create predictive mode in Azure ML with Patent data and financial data from credit rating agency. This predictive model will help organization to create road map for their future investment in Research and development.

Contributions:

- Extraction of patent data, pre-processing using scala scripts in spark.
- Upload the processed XML and JSON file using scala in Azure Blob storage.

- Creation of Hive table on top of XML and JSON data residing on Azure Blob.
- Performance benchmarking using Spark in HDInsight as well as HDP VM in Azure.
- Create predictive model in Spark ML

Project Name: NXG Gen Device Analytics (Cognizant Technology Solutions)
Client: US Media Giant
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 semi-structure 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

One crucial aspect of the solution is to utilize the Hadoop cluster nodes for data integration.

Contributions:

- Designing the overall process flow along with multi-node data ingestion design
- Pig script for reading the semi-structured data and loading them into HDFS.
- Tuning Hive for performance benefit. 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 non-mapreduce jobs like shell scripts etc.
- Shell script for Data parsing and integrating with DMexpress ETL Tool.

Project Name: Log Analytics System (Tech Mahindra Limited)
Client Name: British Telecom
The application enables users to perform analytics over huge volumes of log data with optimum performance. The data is stored in HDFS in distributed manner over a cluster of nodes. The raw data is in the form of xml, .txt etc. Processed data is loaded into Hive table, and is processed by running HQL queries (SQL like), hive internally converts the HQL into optimized Map-Reduce job.

Contributions:

- Writing shell scripts to develop modules and integration with hive and Pig.
- Pig Latin script for loading data and implementing filters while loading.
- Hive script for data extraction and validations.

PROJECT NAME: 21CN NEO T2R (TECH MAHINDRA LIMITED)
CLIENT NAME: BRITISH TELECOM LOCATION: ONSHORE(UNITED KINGDOM)
NEO is heart of the Service Management capability within the OSS stack of BT. It is based on Clarify workflow package. NEO provides both Service Fulfillment (Order Management) and Service Assurance (Fault Management) capabilities/process within BT. NEO platform is responsible for Service fulfillment and Service assurance of all service rollouts by BT on 21CN. NEO manages the work flow in the Order Fulfillment as well Service Assurance, it interacts with various upstream/downstream systems like OneSeibel, GTC and SUMO etc. This project covers the design, development & testing part of all the deliverables of NEO. Both manual and automated testing process is followed in the project.
Contributions:

- Writing shell scripts to develop various functional modules.
- Performed SQL Query Tuning and database parameter tuning performance enhancement.
- Configuring and administering UDPL's, Queues, Users, and Groups in Amdocs ClarifyCRM Client 13.1 sales module.
- Offer Consulting to Service, Operations, Business Improvement Managers.

PROJECT NAME: AGORA (TECH MAHINDRA LIMITED)
CLIENT NAME: BRITISH TELECOM LOCATION: OFFSHORE(PUNE)
Agora is a **ETL**(Extraction-Transformation Loading) data distribution hub that gathers information from many source systems and provides aggregated data in any form or shape desired, but with a strong preference for Web services and incremental feeds.It migrates different source system (upstream systems) data into downstream systems i.e. target system.
Contributions:
- Writing back end stored procedures, packages in Oracle PL/SQL to build ETL logic.
- Performed SQL Query Tuning and database parameter tuning performance enhancement.
- Witten Unix shell scripts to integrate with PLSQL code and set up alarms at database and server level.

Awards/Recognition:

- Received Tech Mahindra Valuable Team Player Award and Pat on the Back Award for my performance in 2012 and 2014 consecutively.
- Received 2 BT E-cards from BT users for handling critical issues and escalation management.

Personal Details:

Name: Gaurab Dawn.
Fathers Name: Sabyasachi Dawn
DOB: 13/03/1987
Permanent Address: N/226/1, Hatudewan, Katwa Road, Burdwan, West Bengal – 713101.
Passport No: H9160067.