

OBJECTIVE

“Transforming large, unruly data sets into competitive advantages”

AREA OF EXPERTISE

- **Big Data Ecosystems:** Hadoop, **Spark**, **Pyspark**, HDFS, Hive, Sqoop.
- **Scripting Languages:** Unix and Python
- **Databases:** Oracle, DB2, MySql
- **Tools:** TWS, Service Now, **ETL Tool** Datastage, **Airflow**
- **Testing Tools:** Eclipse, Datastage Director, Jupyter Notebook
- **Methodologies:** Waterfall, Kanban (Jira)
- **Cloud:** GCP

PROFESSIONAL EXPERIENCE / INTERNSHIPS**Senior Data Developer- Dunnhumby India***January 2021-Present*

Project	Data Engineering GCP System
Description	Business needs processes on GCP based Big-Data Application.
Project Sub Modules	<ul style="list-style-type: none">• Building Data pipeline to populate the data on 3 layers.• Hive tables will be loaded with spark output files.• Further products will be published using top-most Hive Data-Layer.• Build segments for first-last shop on the basis of dimensions• Channels based segmentations developed.• Automation of the jobs to minimize the manual intervention.• Implementation of the products which extracts the data from Hadoop environment to Oracle system.• Migration of the codes from PY2 to PY3.• Prepare pipeline for CI/CD.
Role(Developer) & Responsibilities	<ul style="list-style-type: none">• Design, Implement and supports development of the jobs.• Airflow code setup in the production environment for automation.• Implementation of the product which extracts data from Big-Data environment to Oracle system for further reporting tools.• Mentor L3 teams for their issues during the current processing.• Documentation on the processes been developed.• Codes are migrated to py3 from py2.• Git Pipelines are re-established for migrated py3 code.• Implements solution which reads data and makes it available on GCP platform.• GCP based operators being used for automation through Airflow.

Senior Data Developer- Dunnhumby India*April 2019-January 2021*

Project	Data Engineering Premise System
Description	Business needs migration of SAS based application processes to Big Data Application.

Project Sub Modules	<ul style="list-style-type: none"> • Building Data pipeline to populate the data on 3 layers. • Sqoop the data from RDBMS to HDFS for base lining with old system. • From Datalake, Pyspark code picks the files and perform jobs on the file. • Hive tables will be loaded with spark output files. • Further products will be published using top-most Hive Data-Layer.
Role(Developer) & Responsibilities	<ul style="list-style-type: none"> • Design, Implement and supports development of the jobs. • Setup the test environment for testing the system for further development. • Extracted data from Oracle into Hive using Sqoop for base data creation. • Airflow code setup in the production environment for automation. • Implementation of the product which extracts data from Big-Data environment to Oracle system for further reporting tools. • Mentor L3 teams for their issues during the current processing. • Working knowledge on GIT. • Documentation on the processes been developed. • Optimized the process to reduce the data publish time by ~60% as compared to legacy system. • Developed a process for Data segmentations – Customer, Product and Stores level to help client understand the trends and insights • Developed an automated process to generate insights on sales and transaction data.

PROFESSIONAL EXPERIENCE / INTERNSHIPS

Software Engineer-United Health Group

December 2017-April 2019

Project	Claims Highway
Description	Business needs detailed reports on the claims and its related interest incurred on company.
Project Sub Modules	<ul style="list-style-type: none"> • Data is initially present on RDBMS. • Through Sqoop we need to move it to HDFS. • From HDFS, spark code picks the files and perform jobs on the file. • Hive tables will be loaded with spark output files. • Hive will generate report to the business.
Role(Developer) & Responsibilities	<ul style="list-style-type: none"> • Partners with Business to understand the challenges and pain points, desired business outcomes and provides technical feasibility insight and solution options. • Design, Implement and supports development of the mentioned jobs. • Worked with Admin team for the setup.. • Extracted data from RDBMS into Hive using Sqoop. • Very good understanding of Partitions, Bucketing concepts in Hive and designed both Internal and External tables in Hive to optimize performance. • Primary support project responsibilities like regular production issues (Batch Abend) and mentor downstream for their issues during Batch processing.

PROFESSIONAL EXPERIENCE / INTERNSHIPS

Software Engineer-United Health Group

February 2015-
November 2017

Project□	Claims Highway□
-----------------	------------------------

Description	Claims Highway is an internal UHG application which generates the process for the incoming claims. The purpose is to bring a strategic level focus on displaying capabilities and objectives by aligning all associated activities under a single program and vision. It consists of Spark , HBase, Hive and Sqoop. Claims are coming from upstream application as EBCDIC format. We convert the format into sequential file, extract the dates, records and apply business logic for manipulating and scheduling the claims to be displayed to UHG web-portal depending upon indicators and codes in Python . These claims are loaded to Backend schema tables and using utility these claims were replicated to Front-end tables through Qrep utility. After this, these claims were available for displaying onto web-portal according to indicators.
--------------------	--

Project Sub Modules	<ul style="list-style-type: none"> • Creation of step to sftp the files from Mainframe and make them available for HDFS for our processing. • Programs writing to process the files in python. • Initially auditing the file on the bases of volume, Date and total amount field. • File processing is performed using Spark Tool. • Updating records on the basis of record type business logic.
Role(Developer) & Responsibilities	<ul style="list-style-type: none"> • Partners with Business to understand the challenges and pain points, desired business outcomes and provides technical feasibility insight and solution options. • Design, Implement and supports development end to end data warehouses, data marts, ETLs development that provide structured and timely access to large datasets Worked with Admin team for Environment setup for 64 Node cluster. • Played vital role with Data Ingestion team and dealing with sequential files. • Extracted data from MySql into HDFS using Sqoop. • Knowledge of Oozie workflow for scheduling and orchestrating the ETL process. • Very good understanding of Partitions, Bucketing concepts in Hive and designed both Internal and External tables in Hive to optimize performance. • Primary support project responsibilities like regular production issues (Batch Abend) and mentor downstream for their issues during Batch processing. • Sound knowledge of using AWS Cloud services. • Script created to minimizing monitoring efforts.

TECHNICAL SKILLS

- **Primary Skills** : **PySpark**, Sqoop, HDFS, Python
- **Secondary Skills** : Hive, Shell scripting, **ETL processing**, GCP

OTHER SKILLS

- Self-motivated and diligent with good communication and analytical skills
- Exposure to multi-cultural environments of offshore experience in direct client interfacing
- Good inter-personal skills and experienced in working with teams of various sizes across geographies and time zones
- Flexibility and Adaptability in regards with new technologies and environments.
- Mentorship to L3 resources for the newly developed processes.

EXTRA CURRICULAR AND HOBBIES

- Saves 3rd Position in Dunnhumby Football Premier League.
- Saves 2nd position in Dunnhumby Cricket Premier League.
- Saves 2nd position in Optum Football League.
- Knowledge transition within team to new team-mates and recognized by each resource for support and training.
- Knowledge transition within organization for technologies for which the teams are new.

EDUCATION			
B. Tech- Electronics and Communications	2010-2014	Uttar Pradesh Technical University, UP	70.02%
12 th (Central Board of Secondary Education)	2010	Air Force Sr. Sec School, Delhi	74%
10 th (Central Board of Secondary Education)	2008	Air Force Sr. Sec School, Delhi	70%

PERSONAL DETAILS

Date of Birth	4th October 1991
Gender	Male
Languages known	English, Hindi
Nationality	Indian
Contact Address	G-314, Piyush Heights, Sector 89, Faridabad, Haryana-121002