

Toshit Kumar
Data Science Engineer
Dunnhumby IT Services India, Gurgaon
toshitgarg1993@gmail.com
Mob - +91 9996345607

Objective

Looking for a challenging career which demands the best of my professional ability in terms of technical and analytical skills, helps me broadening and enhancing my current skill and knowledge.

Professional Summary

- Total with 6 years. 2 years of experience in Dunnhumby as Data Science Engineer using Pyspark and Big data tools
i.e. HDFS, Hive, Python and SQL.
- 4 year of experience in TCS using Python, SQL, PLSQL and UNIX commands.
- Good knowledge of DevOps and Agile way of working.

Academics

- Bachelor of Technology in Electronics and communication Engineering
Mahrishi Dyanand University, Sonapat (83.5%) (2014).
- Higher Secondary of Certificate, RSS Sen. Sec School, Hnaumangarh (89.1%) (2010).
- Secondary School Certificate, National Public School, Hanumangarh (85%) (2008).

Technical & Non-Technical Skills

- **Big Data Ecosystems:** Hadoop, HDFS, Pyspark, Hive, MapReduce, Sqoop, Spark, Hue.
 - **Programming Languages:** Pyspark, Python, SQL, PL/SQL
 - **Databases:** Oracle, MySQL, NoSQL (HBase), Hive
 - **Operating Systems:** Windows 10, Linux, UNIX
 - **Tools:** Jupyter Notebook, Putty, Eclipse, NetBeans, Oracle Toad
 - **Methodologies:** Agile, Devops
 - **Familiar With:** Informatica (ETL), Docker , Airflow, Gitlab(Version Control)
 - **Non-Technical skills:** Problem solving & Analytical skills, Effective communication skills, Commitment & dedication towards work, Team building and interpersonal skills.
-

Professional Experience

Dunnhumby IT Services India, Gurgaon

Project : Contact Journey Development

Duration : June 2019 – Present

Role : Data Science Engineer

Skills : Pyspark, Python, HiveQL, UNIX

Tools : Spark 2.4.0, Hive 1.6, Hue, UNIX, Gitlab, Airflow

Summary : Developed project to analyze Retailers promotional offers given to end user and given predictability for the upcoming offers. Migrate codes from existing python2 to python3 standards.

Work Profile

- Developed projects for the analysis of Retailers performance in terms of their promotions.
- Performed advance optimization techniques to run codes in cluster mode to choose an optimum query execution plan.
- Worked on some migration projects to migrate code from python2 and SAS to python 3 using Pyspark.
- Work closely with the business and analytics team in gathering the system requirements. Ensure that the development is at par with the detailed design.

Tata Consultancy Services, Noida

Project : EDW – Application Development- Humana Inc.

Duration : April 2017 – May 2019

Role : Hadoop Developer

Skills : Spark, Python, HiveQL, UNIX

Tools : Hadoop 1.6, Hive 1.1, Hue, UNIX, Spark 2.0,

Summary : The Project is to develop Hadoop ecosystem for the existing Humana Enterprise data warehouse using Sqoop and Hive tools. Hadoop was introduced in this project to load all the data into HDFS and calculated the Item level cost and sent to data warehouse before the expected time. Facilitated insightful daily analyses of patient data collected by external sources.

Work Profile

- Developing the solution using various tools and languages like Hadoop, Hive, PL/SQL, Python and UNIX by following the Iterative Development Technology.
- Import and export data using Sqoop into Hive from existing Oracle Servers.
- Involved in creation of external Hive tables and process the data.
- Involve in creating Hive tables, loading with data and writing Hive queries which will run internally in MapReduce way.
- Work closely with the business and analytics team in gathering the system requirements. Ensure that the development is at par with the detailed design.

Tata Consultancy Services, Pune

Project : CFT – Application Development - Barclays

Duration : Sept 2016 – March 2017

Role : PLSQL Developer

Skills : PLSQL, Informatica, UNIX

Tools : Oracle TOAD, Informatica 9.1.1, Dollar Universe, MS-Excel

Summary : The project was to develop PLSQL Scripts and schedule jobs for Barclays Applications. Data is coming from various sources like SQL Server, Mainframe Systems, flat files which will be loaded in to RDBMS as per the requirement. The entire ETL process consists of source systems, staging area, Data warehouse and Datamart.

Work Profile

- Developed PLSQL scripts to create tables, views, stored procedures, triggers, functions and packages for enhancement in the application using TOAD.
 - Performed SQL performance tuning and hints to direct the optimizer to choose an optimum query execution plan.
 - Developed Mapping, Sessions and Workflow using Informatica to implement the business requirements.
 - Effectively used Informatica parameter files for defining mapping variables, workflow variables, FTP connections and relational connections
-

