

PUNIT KUMAR /# (Senior Data Engineer)

7838386469 , sharmapunit91@gmail.com

Chandigarh

Professional Summary

- Total of 6.5 years of professional experience in Bigdata projects, data warehouse modeling, developing and designing ETL solutions in data warehousing.
- Work experience in big data ecosystems with varying level of expertise around different Hadoop technologies like Hive, Spark, Python, Sqoop, HDFS, ETL, Batch Processing, DataStage, Informatica, Unix, Oracle SQL, MYSQL, Postgres, AWS
- Experience in Development of Py-spark code used for Extracting Facebook-Instagram ads statistics data, AWS S3 files & RDBMS Tables in Hadoop File system.
- Experience in Unit testing along with creation of test cases of Pyspark code, Hive and Sqoop.
- Experience in big data framework and tools like Eclipse, Atom, Docker, TWS, Oozie, Airflow, HPSM, Service now, Trello, Presto, Zeppelin, Aws S3
- Experience in Designing, Development of Hive, Spark and Python jobs
- Experience in all the activities related to requirement gathering, analyzing requirement, development, implementation and support of ETL processes for large scale data.
- Experience in end to end batch processing, operational Monitoring & Incident Management.
- Experience in Artifactory, XL deploy, Jenkins code version maintenance/tracking using SVN, Bitbucket, Git hub
- Experience in methodologies like Agile, Waterfall and Scrum

Technical Competencies

- Programming Languages : Python, Hive, SQL & C++
- Scripting Language : Unix, Python
- Databases : Oracle, MySQL, PostgreSQL, HDFS
- Technologies : DataStage, Informatica, ETL, Hadoop, Spark, Sqoop, Docker
- Application Tools : Airflow, IBM TWS, HPSM, Service Now, Trello, Rally, XL Deploy
- Repository tools : GitHub, Tortoise SVN, NAS cloud
- Project Tools & Methodology : Waterfall, Agile and Scrum

Trantor Software Pvt Ltd

Nov 2020 -Till now

- Client : Genpact
- Project : Roche Data Lake

A multinational healthcare company that operates worldwide in pharmaceutical research, drug discovery deals with a particularly large and diverse set of data, including clinical and biological data, biomarkers, genetic information and much more. The capture, storage and retrieval of these data require specialized technology solutions. **Roche Data lake** is a particular attention in devising systems that allow us to extract information from the data, share it within project teams and derive meaningful knowledge that drives the drug discovery process.

Technologies: Python, Spark, Sqoop, Hive, HDFS,
TOOLS: Atom, Airflow, Docker, Cloudbeaver, Dbeaver
Cloud: Aws, EMR, Ec2

IDS Infotech limited

Dec 2019 -Oct 2020

- Client : Stoxy (Sweden)
- Project : Stoxy Data hub

Stoxy-Eiko: Creation of a Hadoop Platform which will be more secure and have capability to load AWS file, database tables and Facebook-Instagram ads statistic data in HDFS and can be graphically used for analysis using Power BI & Zeppelin Notebooks.

Technologies: Python, Spark, Sqoop, Hive, HDFS,
TOOLS: Atom, Airflow, Docker, Bitbucket, Github, Trello
DATABASE: MySQL, PostgreSQL
Cloud: Aws S3,

Job roles and responsibilities:

- Analyzed client requirements and transformed them to technical design specifications
- Involved in Scrum and Sprint calls for implementation for cards and checkout modules
- Worked on designing the project to have framework tables which is used for ingestion of datasets.
- Worked on the cards of Trello by creating Pyspark Dags as per requirement.
- Creation of Py-spark dags along with sqoop, Hive and HDFS to discover datasets from multiple sources
- Working on spark transformation on extracted Dataframe and convert them in Hive external tables
- Working on creation of testcases along with Unit testing of the Pyspark dags

Genpact

Feb 2019 -Dec-2019

- Client : Macquarie (Australia)
- Project Duration : Feb 2019 -Dec-2019

This project is about developing Pyspark and Hive codes to Highlight the transactions Stored on AWS which are not following new business rules of banking domain. There were 880 critical & non-critical rules which got mixed up with the time, which makes them difficult to judge real cause of exception

Technologies: Python, Spark, Hive, HDFS,
TOOLS: Bitbucket, Github Rally
DATABASE: MySQL

Job roles and responsibilities:

- Analyzed client requirements and transformed them to technical design specifications
- Worked on designing the project to have framework tables which is used for ingestion of datasets.
- Worked on the User stories of rally by creating HQL scripts as per requirement.
- Creation of Hive jobs and Py-spark jobs implement the rules provided by Macquarie bank
- Working on spark transformation on extracted the exemption records which are not following the rules and create Hive external tables on these exempted records.

United Health Group**Aug 2014 -Jan-2019**

- Client : United Health Group (USA)
- Project Duration : Aug 2014 -Jan-2019

3R-RRR- This Project captures Healthcare data like Policy, Plan, Member, Medical Claims, Rx Claims, Premium data from source team (UDW) to perform shadow calculations for Reinsurance and Risk Adjustment which mimics HHS business rules. Additionally, the accruals are sent to downstream FSDB and Enrollee and Claims data are sent to HHS (Center of Medicare & Medicaid).

LANGUAGE : Hive, Linux, SQOOP

TOOLS : SVN tortoise, Datastage, Informatica, IBM TWS, Service now

DATABASE : Oracle, HDFS

Job roles and responsibilities:

- Analyzed client requirements and transformed them to technical design specifications
- Involved in Scrum and Sprint calls for implementation for cards and checkout modules
- Creation of Datastage, Informatica, Hive jobs to implement ETL operations.
- Worked on Unit testing along with creation of test cases.
- Worked on creation of change ticket along with Production deployment of code using xl deploy.
- Worked on Operation monitoring of daily jobs running in production.
- Worked on the resolution of any production issues as per ticket priorities.

Educational Qualifications

- B. Tech (E.C.E) from I.K. GUJRAL PUNJAB TECHNICAL UNIVERSITY, JALANDHAR with 72%
- GATE (E.C.E)-2014 Qualified