

Himanshu Jain

himanshu.jn20@gmail.com, (M) +91-7709277297

TECH SKILLSET AND STACK

- **SKILLSET** - ETL, Data Analysis, Reporting, Data Auditing and Reconciliation, BIG DATA Implementation
- **STACK** - Cloudera HADOOP, HIVE, SPARK (SCALA), IBM Big Integrate / Datastage, Shell Scripting, Teradata DB, Oracle DB, SQL, Ctrl M

EXPERIENCE SUMMARY

Sr. DATA ENGINEER

CAPGEMINI – Gurgaon (September 2018 – Present)

AMDOCS – Pune (December 2016 – August 2018)

ACCENTURE – Pune (March 2015 – December 2016)

- Experience on IBM Big Integrate/Datastage to design jobs to perform ETL processes on HADOOP.
- Used SPARK to design few modules of the ETL Batch Processes such as format conversion of file, processing of HIVE Tables data etc.
- Experience on HIVE/HQL for data extraction and ingestion for various storage formats - Text/Avro/Parquet.
- Proficient in designing DATASTAGE Parallel Jobs for extracting, cleansing, transforming and loading data into data marts/data warehouse.
- Proficient in designing DATASTAGE Sequence Jobs in designing Schedulers, optimized workflows.
- Experience in modifying/optimizing ETL jobs to improve their performance using various performance improvement techniques.
- Experience on Datastage Director to schedule/run the jobs, reading logs, debugging & monitoring performance statistics of the jobs.
- Experience in performance tuning of various long running complex and high resource consuming SQL Queries with significant improvement.
- Proficient in writing Shell Scripts and SQL Scripts for various development activities and overall enhancement of a Module/Project.
- Experience on BTEQ/MLOAD/FASTLOAD utilities of TERADATA.

VOLUNTEER WORK

- Wrote a shell script utility to generate avsc schema files for all the tables/single table of a project from their DDL's.
- Wrote a Batch Utility to automate the individual export of Datastage Jobs present in a Folder and its subfolders in a specified project. It got proposed as an Asset at Accenture which can be used by any Project having such requirement inside or outside Accenture.
- Wrote a Batch Utility to automate the import of Datastage Jobs placed in a specific directory to remove manual intervention, errors and reduce import process time.
- Wrote a Shell Script Utility to identify sequence jobs which has one or more activities/jobs inside it in 'Failed' or 'Validated OK' state after execution.
- Wrote a Shell Script Utility to generate a basic report of last execution of all the Jobs listed in a file.

WORK EXPERIENCE

Project 6

Domain Experience: BANKING

Client: SBI Card, India

Role: Senior ETL Developer (Jan 2020 – Present)

Project & Role Description –

- Designed Incremental ETL process to read data from huge files having CFD schema and large number of columns and load them into Staging layer (Hive) as Text file using BIG DATA File stage.
- Designed a Spark Module using SCALA to convert the text file created on HADOOP into 'Avro' File format and load them into HIVE External table(Using Avro Schema files) by creating external partitions.
- Designed generic jobs to load and transform data from Staging Layer 'Avro' tables to Processing layer 'Parquet' format tables as per the business requirement along with the Snapshot of the Data.
- Designed several shell scripts for various processes of the Incremental Batch Load –
 - File Watcher to check availability of expected files with checks including busyness of file to avoid partial file copy issue.
 - Validation of Source Files – Count Validation, Length validation of each record etc.
 - Reconciliation between Source Files ,Staging layer(Hive) and Processing Layer(Target)
 - Archiving of files, Data Purging etc.
- Designed Multiple Sequence Jobs for Scheduling of the Batch.

Project 5

Domain Experience: BANKING

Client: HSBC, UK & HK

Role: Senior ETL Developer (Nov 2018 – Dec 2019)

Project & Role Description –

- Designed end to end ETL flow of a new solution in the project - Resubmission of reports.
- Implemented various performance tuning techniques to improve the performance of Datastage Jobs and SQL Queries.
- Developed various Shell Scripts for performing different requirements for the execution of the Batch as per the end to end deliverables of the Project.
- Developed various Datastage Sequence/Parallel jobs to load different Client Institutions data into the Staging area table as per the required transformation.
- Developed jobs to create multiple detail and recon reports(CSV) as per the business requirements.
- Developed script to reconcile the data of multiple reports before their distribution.
- Have mentored and shared my expertise to multiple members within and outside the team on Datastage, SQL and Unix Shell Scripting.

Project 4

Domain Experience: TELECOM

Client: THREE, Ireland

Role: ETL Developer (Jan 2018 – Aug 2018)

Project & Role Description –

- End to end designing of the ETL flow from Source (files) to the Target(Teradata Database) for implementing Batch processing of data.
- Designed and Developed generic Datastage Multi-instance jobs to read files and load in Staging area.
- Developed Bteq scripts to load data from Staging area to Target.
- Designed and Developed generic sequence job to implement end to end data flow for an entity.
- Designed and Developed Sequence jobs to implement end to end scheduling of all the jobs of the Project for Daily Batch Execution.
- While designing scheduling through sequence jobs, implemented below FEATURES –
 - Skip a Job which got failed in the scheduling sequence without setting 'Validate Only' or touching the Job after failure so that the whole batch don't get stuck.
 - Reset a sub-sequence job in case of failure where by default in the flow it is 'Run to start from where it failed' as per the run-time requirement.
- Designed performance efficient jobs to process huge amount of data.

Project 3

Domain Experience: TELECOM

Client: TELEFONICA, Argentina

Role: ETL Developer (Dec 2016 – Jan 2018)

Project & Role Description –

- Developed various jobs to build and maintain Operational Data Store (ODS).
- Developed Mload Scripts to extract data from source files and load into Landing Area.
- Developed Datastage Jobs and Bteq scripts to transform the Landing Area data and load into the ODS.
- Optimized the existing jobs for better performance and implemented new efficient business logic.
- Involved in various performance improvement techniques to reduce time for processing of the huge data.

Project 2

Domain Experience: RETAIL

Client: TJX, USA

Role: Associate Software Engineer (Feb 2016 – Dec 2016)

Project & Role Description -

- Involved in system study and coordinating directly with client for requirements gathering and system design implementation.
- Understanding business requirements, identifying systems and getting it developed and implemented.
- Developed and Supported the project fully from its Development phase to System Integrated Testing phase with a very small Team (3 Members).

Project 1

Domain Experience: Health & Public Service

Client: RMG, London UK

Role: Associate Software Engineer (May 2015 – Jan 2016)

Project & Role Description -

- Developed various modules independently as per the Business Requirement.
- Supported various module's end to end development with Senior Team Members.
- Wrote shell script utilities to identify jobs and their respective connector stages in a project having more than 400 jobs which used a specific property/tables. This saved a lot of manual effort and time.

EDUCATION

Course	Institute	Percentage
B.E. – E &TC	Pune Institute of Computer Technology, Pune (PICT)	64.66
HSC – CBSE	Central Academy, Kota (Raj.)	78
SSC – CBSE	Central Academy, Chittorgarh (Raj.)	81