Prasant Kumar Senapati

Email: prasantsenapati.2020@gmail.com | **Phone:** 9591446478 | Bangalore, India

CCA cloudera

SPARK & HADOOP DEVELOPER

CCA-175 License#: - **100-022-453**

Skills

BigData Technologies :-
        **Spark, Scala,Python, Hadoop
        HIVE,** Airflow, Nifi, Sqoop,Oozie
        & HDFS

Cloud Technologies :-
        **AWS,Azure,Amazon redshift ,S3 ,
        Azure** BLOB, Azure SQL Data
        warehouse.

NoSQL :-   **Cassandra** & **HBASE**

ETL Tools: - IBM Data Stage (8.x)
        Informatica Power Center (8.x/9.x)
        SSIS
Databases:- Oracle , **SQL Server**
        IBM DB2 &, **MySQL**

Reporting Tool: - Power BI

Certification

- **Cloudera Spark and Hadoop Developer** (CCA-175) Certified**,** 2019.
- **INS21**-Property&Liability Insurance certified, 2010.
- **Informatica** mapping Design 8.0 certified, 2008.

Awards & Achievement

- Received "**BEST of IBM" award** for contribution to stabilize EBI and EPAS Auto application.
- Received "**Tavant Excellence Award**" for successfully delivery of NMI EDW project.

Profile Summary

- Over **13** Years of IT experiences in **Big Data, Cloud** , Enterprise Data Architecture, Data Integration platforms.
- Having **5+** years of extensive hands-on experience in **Big Data** technologies **-Hadoop ,Spark, Scala, Python**, **Hive**, **HDFS**, **Amazon redshift** ,**Airflow**, NoSQL ,Amazon S3 , AZURE SQL Datawarehouse.
- Having **2+** years of experience in **Cloud technology** (**AWS** & **Azure**).
- Having **2+** years of experience in **Python** scripting.
- Having **6+** years of Data Integration experience using **Informatica** & **IBM DataStage** ETL tool.
- Good experience in designing **Data pipeline** using Apache **Airflow.**
- Having experience in **Hadoop Cluster capacity** planning.
- Having strong experience in SQL/PL-SQL & **UNIX scripting**.
- Extensive experience with design, development, customizations, and implementation of software applications using BI and Big data technologies.
- Proficient in analysing and translating business requirements to technical requirements and architecture.
- Experienced with processing different file formats like, Parquet**,** ORC, Avro, XML, JSON and Sequence file formats using Spark.
- Experience in implementing Spark using Scala and Spark SQL for faster analysing and processing of data.
- Good experience working on big data platforms like Cloudera CDH.
- Experience in designing table partitioning, bucketing and optimized hive scripts using different performance utilities and techniques.
- Hands-on JIRA, Confluence, Kanban Board, BMC remedy, TFS, IBM ClearCase ,**Git**, **Azure DevOps** and SVN tool.
- Good domain experience on Insurance and Finance, Mortgage Industry.
- Experience in Agile methodology**.**

Working Experiences

- ❖ **Tavant Technologies |** Associate Technical Architect **|** Bangalore
  Date: - 01/2013 to till date

- ❖ **IBM India Pvt. Limited |** Data Specialist **|** Bangalore
  Date: - 03/2007 to 01/2013

On-Site Experiences

- Visited **Sweden** in 2011 (duration: -10 months).
- Visited **USA** in 2007 (duration: - 03 months).

Work Experience

Lead Data Engineer – Associate Technical Architect (Tavant Technologies)
July 2015 to **Present**
Skills:- Spark, Scala, Python, Hadoop, HDFS, Hive, Airflow, SQL Server, MySQL, Jira, Git, and Cloudera CDH.

Responsibilities:

➢ Hands on experience with scalable solutions on Big data ecosystem using Spark, Scala, Python, Hadoop, HDFS, Hive, Airflow, SQL Server, MySQL, Jira, GitHub, and Cloudera CDH.
➢ Responsible for Data Ingestion, Data Cleansing, Data Standardization and Data Transformation.
➢ Optimized Spark executor memory while submitting spark jobs to cluster.
➢ Developed Apache spark jobs to ingest data from various Source systems to HDFS.
➢ Read data from HDFS into Spark Dataframe and Implemented Spark Dataframe transformations to map business analysis and load the data into Hive table and MYSQL table.
➢ Performed Data enrichments such as Filtering, Grouping, Sorting, Aggregation etc. in spark dataframe.
➢ Incremental loading of records implements using Apache spark jobs.
➢ Worked on spark code performance optimization to handle data skew.
➢ Used Spark-SQL to process data and to run complex query on Spark SQL.
➢ Explored and implemented various spark optimization technique to improve spark job performance.
➢ Utilized Hive tables and HQL queries for weekly and monthly reports.
➢ Worked on creating Hive managed and external tables based on the requirement.
➢ Implemented Partitioning, Dynamic Partitions and bucketing in HIVE table for efficient data access.
➢ Worked on Airflow to develop DAG to automate data pipeline.
➢ Configured Airflow data pipeline to run multiple spark jobs in parallel
➢ Configured Airflow Data Pipeline to automate the spark jobs for uploading data into SQL Server and Hive tables.
➢ Collaborated with the application Team, Data science team and BI teams to ensure data quality and availability.
➢ Collaborating with business users/product owners/developers to contribute to the analysis of functional requirements.
➢ Imported data from various sources into Spark Dataframe for analysis.
➢ Supported code/design analysis, strategy development and project planning.
➢ Created reporting tables for BI team in Hive and exported aggregate data from HDFS to Hive with help of Apache Spark.
➢ Involvement in design, development, and testing phases of Software Development Life Cycle
➢ Define Data Architecture standards, best practices and participate in governance activities during Big Data projects.
➢ Define and implement standards and best practices for data ingestion, analytics, and integration in Data Lake.

ETL Lead (Tavant Technologies)
Feb 2013 to Jun 2015
Skills:- Cloud Technology , Amazon redshift , Amazon S3, Azure BLOB, Azure SQL Datawarehouse
        Informatica, MySQL, SQL Server, Unix, PowerBI, TFS ,

Responsibilities:

➢ Experience in developing Informatica mappings using transformations like Source Qualifier, Connected and Unconnected Lookup, Normalizer, Router, Filter, Expression, Aggregator, Stored Procedure, Sequence Generator, Sorter, Joiner, Update Strategy, Union Transformations.
➢ Exported data warehouse data to Amazon S3 using informatica.
➢ Uploaded aggregated data to Amazon redshift using AWS copy command.
➢ Strong experience in Extraction, Transformation, loading (ETL) data from various sources into Data Warehouses and Data Marts using Informatica Power Center (Designer, Workflow Manager, Workflow Monitor, Repository Manger)
➢ Understand business requirements and assist the BA team for technical documentation

- ➢ Creating End-to-End framework of the ETL development process.
- ➢ Participate in the planning and estimation meetings with clients
- ➢ Lead a team of 4 ETL Developers and extend technical help to meet the deadlines
- ➢ Build mappings to read files from several sources and generate fixed width output files and XMLs
- ➢ Create DB2 scripts to generate and load test data in the existing tables
- ➢ Exported Datawarehouse aggregated data to Azure BLOB using informatica.
- ➢ Uploaded aggregated data from Azure BLOB to Azure SQL data warehouse using copy command
- ➢ Conduct Performance Analysis and Perform Code optimization
- ➢ Coordinating with the Control-M team for scheduling jobs for testing and production environment
- ➢ Create Sequence jobs with Error Handling and Email-Triggers
- ➢ Extend support in Unit Testing and Integration testing
- ➢ Generating Automation jobs for Generating Bulk data and loading it into the DB
- ➢ Writing Wrapper Scripts in UNIX
- ➢ Participate in the Code Reviews with the Client Team
- ➢ Creating Deployment Packages and tickets for Production
- ➢ Analyzing any UAT issue reported by the QA team and providing a timely solution for closing it.
- ➢ Recruiting of new ETL Developers in team and organizing Knowledge Transfer sessions.
- ➢ Experience on developing analytic reports and dashboard using Power BI.

Data Specialist (IBM India Pvt Ltd)
Mar 2007 to Jan 2013
Skills: IBM DataStage, Informatica, DB2, SQL Server, Oracle, Unix, IBM ClearCase.

Responsibilities:

- ➢ Designed the ETL jobs using IBM Info sphere Data Stage to Extract, Transform and load the data into staging, ODS and EDW.
- ➢ Designed parallel jobs using stages such as Join, Merge, Lookup, Remove Duplicates, Copy, Filter, Funnel, Dataset, Lookup, Pivot, and Sort, Surrogate key Generator, Change Data Capture (CDC), Modify, Row Generator and Aggregator.
- ➢ Responsible for data analysis, requirements gathering, report analysis, source-to-target mapping, frequency analysis, process flow diagrams, and documentation. Handled Performance Tuning of Jobs to ensure faster Data Loads.
- ➢ Designed sequence jobs using the activities such as Job Activity, Nested Condition, Notification Activity, Sequencer Activity, Terminator Activity and Execute Command.
- ➢ Performed the Integration and System testing on the ETL jobs. Responsible for preparing ad hoc jobs.
- ➢ Created Job Parameters and Environment variables to run the same job for different sources and targets.
- ➢ Used the Director to schedule running the job, testing, and debugging its components, and monitoring.
- ➢ Created Shared Containers for Re-using the Business functionality.
- ➢ Experience working with Data Warehousing Concepts like OLAP, OLTP, Star Schema, Snowflake Schema, Logical/Physical/ Dimensional Data Modeling.
- ➢ Extensively used ETL methodology for performing Data Profiling, Data Migration, Extraction, Transformation and Loading using IBM Datastage and Informatica and designed data conversions from wide variety of source systems including Oracle, DB2, SQL server, MySQL, non-relational sources like flat files, XML and Mainframe Files.
- ➢ Extracted data from multiple operational sources for loading staging area, Data warehouse, Data Marts using SCDs (Type 1/Type 2/ Type 3/type 5) loads.
- ➢ Able to assess business rules, collaborate with stakeholders and perform source-to-target data mapping, design, and review.
- ➢ Experience in monitoring and scheduling using Tidal, Control M and using UNIX Scripting.
- ➢ Able to perform independently in complex troubleshooting, root-cause analysis, and solution development.
- ➢ Analyzed and created facts and dimension tables.
- ➢ Experience with dimensional modelling using star schema and snowflake models.
- ➢ Experienced in identifying and documenting data integration issues, challenges such as duplicate data, nonconfirmed data, and unclean data
- ➢ Applied optimization techniques in ETL job level and query level.
- ➢ Proven team player, good communication skills and quick learner.