

Prasant Kumar Senapati

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



SPARK & HADOOP
DEVELOPER

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

Skills

BigData Technologies :-

Spark, Scala, HIVE, Hadoop, Airflow,
Nifi, Sqoop & **HDFS**

Cloud Technologies :-

Amazon redshift, S3, Azure BLOB,
Azure HDInsight

NO-SQL :- **Cassandra**

ETL Tools: - IBM Data Stage (8.x)

Informatica Power Center (8.x/9.x)
SSIS

Databases:- Oracle 10g/12, SQL Server 2016

IBM DB2 &, MySQL 5.x

Reporting Tools: - Microsoft 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

- ◆ **13+** Years of industrial experience in **Big Data, Cloud** and **Data warehousing** technologies using **Apache Spark, Scala, Hive, HDFS, Airflow, IBM DataStage & Informatica** tool.
- ◆ Extensive experience with design, development, customizations and implementation of software applications using Big data technologies.
- ◆ Proficient in analysing and translating business requirements to technical requirements and architecture.
- ◆ Having 4+years of extensive experience in **Big Data** technologies (**Apache Spark, Scala, Hive, HDFS, Airflow, Cassandra** etc).
- ◆ Exposure to **Amazon AWS** and **Azure Cloud**.
- ◆ Having 7 +years of exp. using **DataStage & Informatica** ETL tool.
- ◆ Having good knowledge in **SQL & PL-SQL & UNIX scripting**.
- ◆ 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 and SVN tools.
- ◆ Good domain experience on Insurance and Finance, Mortgage Industry.
- ◆ Experience in Agile methodology.
- ◆ Good experience in designing **Data pipeline** using **Apache Airflow**.
- ◆ **Spark & Hadoop Certified** from **Cloudera corporation**.

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

Data Engineer (Tavant Technologies)

July 2016 to Present

Skills: Spark, Scala, Hadoop, HDFS, Hive, Airflow, SQL Server, MySQL, Jira, TFS, SVN and Cloudera CDH .

Responsibilities:

- Involvement in design, development, and testing phases of Software Development Life Cycle.
- 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 SQL server 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.
- Responsible for Data Ingestion, Data Cleansing, Data Standardization and Data Transformation.
- 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.
- 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.
- 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.

ETL Lead (Tavant Technologies)

Feb 2013 to Jun 2016

Skills: Informatica 9.x, MySQL, SQLServer, Oracle, Unix, PowerBI, SVN, 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.
- Experience on developing analytic reports and dashboard using Power BI.
- 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
- 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.

Data Specialist (IBM India Pvt Ltd)

Mar 2007 to Jan 2013

Skills: Informatica 9.x, IBM DataStage 8.x, DB2, SQLServer, Oracle, Unix, PowerBI, IBM ClearCase, TFS.

Responsibilities:

- Designed the ETL jobs using IBM Info sphere Data Stage to Extract, Transform and load the data into staging, ODS and EDW.
- Designed and developed the ETL jobs using Parallel Edition which distributed the incoming data concurrently across all the processors, to achieve the best performance.
- 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.
- Translated business rules and functionality requirements into ETL derivations.
- 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.
- Worked with Developers to troubleshoot and resolve issues in job logic as well as performance.
- 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 ESP, 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.
- Analyzed business requirements and worked closely with the various application teams and business teams to develop ETL procedures that are consistent across all applications and systems.
- Developed the functionality at the database level PL/SQL using Toad tool as well at the Unix OS level using shell scripting.
- 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
- Worked cooperatively with the team members to identify and resolve various issues relating to ETL tools and databases.
- Applied optimization techniques in ETL job level and query level.
- Proven team player, good communication skills and quick learner.