** E-Mail :** santosh.kumar32@tcs.com

**Santosh Kumar** **Mobile** : 8017700650

# **Career Objective**

Pursue a career in a creative and challenging environment, wherein my technical and intellectual skills will be optimally used for mutual benefits and be loyal to my services in the software sector.

# **Professional Summary**

* Over 8.5 Years of IT experience with emphasis on business requirements analysis design development, implementations and testing using **DW/BI tools**, **Hadoop Ecosystems and AWS Cloud Services**.
* Experienced on different Insurance, Financial and life science sectors using **ETL methodology**, **Hadoop Ecosystem and Cloud Technologies**.
* 5 years of experience in development and designing of ETL methodology for supporting data transformations using **TALEND 6.x, Informatica Power Center** 8**.x /9.x/10.x.**
* Worked with different data sources like **S3 Cloud Storage**, SharePoint Sources, XML files, **REST/SOAP API’s**, **Databases** and **HDFS**.
* Have exposure to different databases like **Oracle, MS SQL server, Snowflake and AWS Redshift.**
* **Have 1.5 years of exposure to AWS cloud technologies like S3, Redshift and EC2.**
* Experienced in analyzing business requirements and translating requirements into functional and technical design specifications.
* Created UNIX shell scripts to run the Informatica workflows and controlling the ETL flow.
* Having 2+ years of experience in writing **Stored Procedures, Functions, Exception Handling, Triggers, Index, Views** etc.

**TECHNICAL SKILLS:**

|  |  |
| --- | --- |
| **DBMS** | AWS REDSHIFT, SNOWFLAKE, Oracle 11g/10g/9i/8i, MS SQL 2014, MS SQL 2008, |
| **Big Data Technologies** | Hive, HDFS, HBase, Sqoop, Spark 1.6 |
| **ETL Tools** | TALEND 6.x, Informatica (Power Center 10.1/9.6.1/9.5.1/9.1.0/8.6.1) |
| **Languages** | SQL, PL/SQL, shell scripts, Java (Core), Python |
| **Platform** | Windows, Unix, AWS Cloud Platform |
| **Other Tools** | GitHub, Presto, Toad7.6/8.5, PL/SQL-Developer, MS-Office, Putty, SQL Workbench, Aginity |

**CERTIFICATION:**

|  |  |
| --- | --- |
| **Certification Authority** | **Certification Achieved** |
| **Amazon Web Services** | **AWS Certified Developer - Associate**  (Validation Number JQWLFNLC114Q1ECK) |

|  |
| --- |
| **WORK EXPERIENCE** |

* Tata Consultancy Services, Kolkata (June 2016 – Till Date)
* Cognizant Technology Solutions, Kolkata (Dec 2014 –June 2016)
* Infosys Limited, Pune (July 2011 – Nov 2014)

**PROJECT DETAILS:**

1. **Client: AIG**  
   **Duration: April 2019 to June 2019**

**Role: Talend Senior Developer/Snowflake Developer**

**Project Name: - Wholesaler Modernization**

**Description:**

The main purpose of this project is to modernization of existing salesforce and Wholesaler Comp legacy Java batch jobs and convert to Talend Data integration to address data gaps, inefficiency of legacy jobs, frequent jobs related issues and poor performance and achieve better quality and correctness of data and maintainability. Cloud Database **SNOWFLAKE** was used as a target system and the complete idea was to move data to digital/Cloud platform so that underlying downstream systems consumes data in robust fashion.

**Responsibilities:**

* Parsing high-level design specification to simple ETL coding along with Talend mapping standards and Snowflake standards.
* Lead the team in optimized way to ensure on/before time delivery of bug free application.
* Designed an incremental CDC job in optimized way to perform upsert in faster manner.
* Performance tuning on Snowflake and Talend level to increase the data throughput.
* Directly interact with business user IT team for requirement clarification and translating into optimized Talend code.
* Implement error capturing mechanism to minimize the debugging time in Talend CODE.
* Performed Unit testing and maintained test logs and test cases for all the mappings.
* Co-ordination between End business user and Development team.

**Environment:** Talend 6.4, **Snowflake**, Salesforce, Shell Script, Core Java

1. **Client: AIG**  
   **Duration: Feb 2018 to April 2019**

**Role: Talend/Big Data Developer**

**Project Name: - eStation-Connext**

**Description:**

**AIG (American International Group**), is an American multinational finance and insurance corporation with operations in more than 80 countries and jurisdictions. The company operates through three core businesses: General Insurance, Life & Retirement, and a standalone technology-enabled subsidiary. eStation is a part of its one of the insurance LOB and AIG wanted it to migrate eStation insurance data (old data) to connext systems (one of its new LOB). In parallel, old system data and newly formed system data moved to Hortonworks Platform where data analytics is performed.

Using Talend 6.4.1, Bigdata Batch jobs were created and ran under Spark environment to load full/delta data on daily basis to the targets table in HIVE for data analysis and reporting purposes.

Three intermediate layer databases were created in Hortonworks platform for incremental raw data loading from oracle systems (sqooping) and further processing in HDFS using Talend spark jobs.

**Responsibilities:**

* Parsing high-level design specification to simple ETL coding along with Talend mapping standards.
* Created Sqoop framework in Talend which will dynamically generate the sqoop script by sourcing configuration files and will take care of all error handling mechanism required.
* Used Spark engine in Talend in subsequent layer after RAW, to optimize the data load time. Implemented all the tunning configurations in Talend spark job.
* Performance tuning is performed at the Mapping level as well as the Database level to increase the data throughput.
* Directly interact with business user IT team for requirement clarification and translating into optimized Talend code.
* Implement error capturing mechanism to minimize the debugging time in Talend CODE.
* Performed Unit testing and maintained test logs and test cases for all the mappings.
* Co-ordination between End business user and Development team.

**Environment:** Talend 6.4, Spark 1.6, Hive, Sqoop, HBase, Shell Script, Java

1. **Client: JNJ**  
   **Duration: July 2017 to FEB 2018**

**Role: Talend/AWS Developer**

**Project Name: - Precision Marketing**

**Description:**

In this project, we capture user information on different product segment of J&J with the help of vendor provided log files having size of 150 GB on daily basis.

It captures how much user has shown Interest, browsed and seen videos of different segments and its sub-segments related to JNJ which helps it make different marketing strategies and solutions to boost its revenue forecast.

Using the log source data, it is segregated into categories like “Media type and browser used by user”, “Trend on user’s interest on different segments” and “Top segments on hit list” to provide the visualization to business user for marketing purposes.

**Responsibilities:**

* Parsing high-level design specification to simple ETL coding along with Talend mapping standards.
* Implemented robust Talend data ingestion design to handle **150 GB log** data coming from source at AWS S3 file landing location on daily basis.
* Directly interact with business user IT team for requirement clarification and translating into optimized Talend code.
* Using Batch script to Upload/Download files from S3 to AWS EC2 to minimize the processing time taken in TALEND S3 components.
* Implement error capturing mechanism to minimize the debugging time in Talend CODE.
* Performance tuning is performed at the Mapping level as well as the Database level to increase the data throughput.
* Performed Unit testing and maintained test logs and test cases for all the mappings.
* Co-ordination between End business user and Development team.

**Environment:** Talend 6.x, AWS Redshift, AWS RDS MSSQL Server, AWS S3, Hive, Impala, Shell Script

1. **Client:** [**Major**](https://en.wikipedia.org/wiki/AstraZeneca) **Pharma Company  
   Duration: December 2016 to June 2017**

**Role: Talend/AWS Developer**

**Project Name: CRM**

**Description: --**

The main aim of this project is to enable holistic understanding of J&J’s Consumers, Patients, Professionals and Investors by bringing together data from all relationship marketing touchpoints to create a single view of our Consumers. Profile the data across sources to define common data model for consumer data and Build tech solution and Integration to support single view of consumers. There are highly fragmented data with many market-based RM solutions in place across Consumer and evolution is needed to meet maturing direct marketing needs drive by eCommerce. In this project, different sources like, **SFMC, REST API’s, SOAP API’s** and Files on S3 were consumed and were fit into common structured database inside **AWS RDS MSsql server** to provide single platform to view the data based on incremental/Full load strategies. This project was completely driven using AGILE process framework and delivered with in highly optimized way.

**Responsibilities: --**

* Parsing high-level design specification to simple ETL coding along with Talend mapping along with AWS standards consideration.
* Directly interact with business user IT team for requirement clarification and translating into optimized Talend code and AWS optimized usage.
* Migrating and scheduling Talend developed code in higher environment in TAC with help of GST team.
* Using Batch script to Upload/Download files from AWS S3 to AWS EC2 to minimize the processing time taken in TALEND S3 components.
* Implement error capturing mechanism to minimize the debugging time in Talend CODE.
* Performance tuning is performed at the Mapping level as well as the Database level to increase the data throughput.
* Performed Unit testing and maintained test logs and test cases for all the mappings.
* Co-ordination between End business user and Development team.

**Environment: Talend 6.2.1, AWS Redshift, AWS RDS Sql Server, AWS S3**, **AWS EC2, HIVE.**

1. **Client**: **JNJ**  
   **Duration: September 2016 to November 2016**  
   **Role: ETL Informatica Developer**

**Project Name: -- JSCEDW CPV Fine Tuning**

**Description: --**

JnJ is an American Pharma and [healthcare](https://en.wikipedia.org/wiki/Health_care) company, headquartered in New Brunswick, New Jersey. Every medical product invented in JNJ Labs, goes through rigorous test processes in its labs inside its medicine production center. Only, after getting consent from Labs, any medical product is ok to be released in the market with particular batch id. The purpose of JSCEDW is to integrated data from various SAP, RDBMS and Flat File sources into an Enterprise Data Warehouse (EDW) on Teradata platform. The purpose of this is to integrated data from MDE and OSI\_PI source systems to into existing ELIMS Enterprise Data Warehouse (EDW) on Teradata platform. This will cater the need of enterprise level supply chain data for operational and KPI reporting. ETL processes the fetch the data from source and delivers into DIL/DBL layer by keeping all the existing constraints intact. The EDW solution delivered would suffice the reporting needs of CPV project.

**Responsibilities:**

* Prepared the design document like HLD, LLD, TSD by directly interacting with clients.
* Understanding the business requirements and translating it into robust informatica code with optimized throughput.
* Prepared the DDL scripts for creating the tables in Database.
* Migrating complete code from informatica 9.6.1 to informatica 10.1 big data version with no defect.
* Used Teradata utilities like BTEQ, fast load, fast export, multiload for data conversion.
* Reusable transformations and Mapplets are built wherever redundancy needed
* Performance tuning is performed at the Mapping level as well as the Database level to increase the data throughput.
* Created shell scripts to fine tune the ETL flow of the Informatica workflows.
* Used relational SQL wherever possible to minimize the data transfer over the network.
* Building of test cases and test throughout to eliminate any kind of defect before it gets caught.
* Actively involved in team code reviews and advising team wherever needed to improve it and make it tuned.

**Environment:** Informatica 10.1, Informatica 9.6.1, Teradata 15, UNIX shell Script, Hive

1. **Client**: **JNJ**  
   **Duration: June 2016 to September 2016**  
   **Role: ETL Informatica Developer**

**Project Name: -- JSCEDW CPV Blocking Issue**

**Description: --**

JnJ is an American Pharma and [healthcare](https://en.wikipedia.org/wiki/Health_care) company, headquartered in New Brunswick, New Jersey. Every medical product invented in JNJ Labs, goes through rigorous test processes in its labs inside its medicine production center. Only, after getting consent from Labs, any medical product is ok to be released in the market with particular batch id. The purpose of JSCEDW is to integrated data from various SAP, RDBMS and Flat File sources into and ELIMS Enterprise Data Warehouse (EDW) on Teradata platform. This will cater the need of enterprise level supply chain data for operational and KPI reporting. ETL/ELT processes the fetch the data from source and delivers into DIL/DBL layer. The EDW solution delivered, would suffice the reporting needs of CPV project.

**Responsibilities:**

* Understanding the business requirements and translating it into robust informatica code with optimized throughput.
* Prepared the design document like HLD, LLD, TSD by directly interacting with clients.
* Prepared the DDL scripts for creating the tables in Database.
* Used Teradata utilities like BTEQ, fast load, fast export, multiload for data conversion.
* Reusable transformations and Mapplets are built wherever redundancy is needed.
* Performance tuning is performed at the Mapping level as well as the Database level to increase the data throughput.
* Created shell scripts to fine tune the ETL flow of the Informatica workflows.
* Used relational SQL wherever possible to minimize the data transfer over the network.
* Building of test cases and test throughout to eliminate any kind of defect before it gets caught.
* Actively involved in team code reviews and advising team wherever needed to improve it and make it tuned.

**Environment:** Informatica 9.6.1, Teradata 15, UNIX shell Script

1. **Client:** [**Amgen**](https://en.wikipedia.org/wiki/AstraZeneca) **Duration: January 2015 to till April 2016.**  
   **Role: ETL Informatica developer**

**Project Name: ERA\_CANADA**

**Description: --**

Drug discovery and development in [AMGEN](https://en.wikipedia.org/wiki/AstraZeneca) is a complex process and involves mixed collaborations with external organizations, persons (HCP i.e. Health Care Professional, non- HCP) which includes scientific exchange, sharing of molecules or materials, scientific publications, industry consortiums, sponsorships, and grants. It is using MsSql 2014 to build its data warehouse. Previously it was using SSIS to implement its warehouse but it was taking longer time and report generation was also very slow. [AMGEN](https://en.wikipedia.org/wiki/AstraZeneca) wanted to migrate its SSIS code to Informatica code for the betterment of lagging factors and also include EDL part of data extraction.

**Responsibilities:**

* Parsing high-level design specification to simple ETL coding along with mapping standards.
* Understanding the requirement from customer (Business User).
* Created Source to Target mapping sheet corresponding to SSIS code.
* Prepared the DDL scripts for creating the tables in Database.
* Created mapping, session, Workflows using **Power center designer** as per requirement to perform ETL from source systems (**EDL/Databases**) to target system.
* Unit testing of all ETL components and submitted the results.
* Used relational SQL wherever possible to minimize the data transfer over the network.
* Developed ETL programs using Informatica 10.1to implement the complex SSIS logic.
* Effectively used Informatica parameter files for defining mapping variables, workflow variables, FTP connections and relational connections.
* Programmed stored procedures, queries whenever required to minimize the load time.
* Unit and System Testing.
* Involved in resolving production problems for the applications and Ensure all support service level agreements are met.
* **Environment: Informatica 9.6, SSIS,** MSSQL 2014, MSSQL 2008, **HDFS**, Hive, Sqoop, UNIX

1. **Client: JPMC**  
   **Duration: December 2013 to November 2014.**  
   **Role: ETL Informatica developer (Callidus)**

**Description: --**

**JPMorgan Chase & Co.** is an American multinational banking and financial services holding company headquartered in New York City. JPMC migrated its original compensation management tool to Callidus for one of its line of business. It required its old compensation system data to be migrated in the format acceptable for **Callidus** tool to work and generate the expected result. Data migration implemented using informatica ETL.  
**Responsibilities:**

* Developed ETL programs using Informatica to implement the business requirements.
* Communicated with business customers to discuss the issues and requirements.
* Created shell scripts to fine tune the ETL flow of the Informatica workflows.
* Performance tuning was done at the functional level and map level. Used relational SQL wherever possible to minimize the data transfer over the network.
* Effectively used Informatica parameter files for defining mapping variables, workflow variables, FTP connections and relational connections.
* Used debugger in identifying bugs in existing mappings by analyzing data flow, evaluating transformations.
* Pre- andpost-session assignment variables were used to pass the variable values from one session to other.
* Designed workflows with many sessions with decision, assignment task, event wait, and event raise tasks, used informatica scheduler to schedule jobs.
* Reviewed and analyzed functional requirements, mapping documents, problem solving and trouble shooting.
* Performed unit testing at various levels of the ETL and actively involved in team code reviews.
* Identified problems in existing production data and developed one-time scripts to correct them

**Environment:** Informatica 9.1.0**,** Oracle 11g/10g, UNIX

1. **Client: AIG Japan**  
   **Duration: September 2012 to October 2013.**  
   **Role: Pl/Sql/Informatica developer (Callidus)**

**Description: --**

This project is based on incentive management tool “Callidus” which calculates incentives for employees and keeps the data to its base tables. The objective of this project is to provide the client with a robust procedure that generates customized reports based on the data available in **Callidus** base tables, which will either be used by the insurance company for their internal purposes or be sent to their headquarters in Japan. These reports will be generated based on the data available in their several Departments. I have been involved in end to end implementation of ETL life cycle also.

As a developer, I was involved in the following: --

* Understanding the requirement from customer (Business User).
* Translating the requirement into Design (HLD, LLD)
* Prepared the design document like HLD, LLD, TSD etc.
* Created Source to Target mapping sheet.
* Prepared the DDL scripts for creating the tables in Database.
* Created mapping, session, Workflows as per requirement.
* Unit testing of all ETL components and submitted the results.
* Used relational SQL wherever possible to minimize the data transfer over the network.
* Developed ETL programs using Informatica to implement the business requirements.
* Effectively used Informatica parameter files for defining mapping variables, workflow variables, FTP connections and relational connections.
* Programmed Packages, stored procedures, queries, and triggers on Oracle 10g.
* Unit and System Testing.
* Plan and supervise the daily activities of the Callidus application support.
* Database Query tuning and Implementation.
* Involved in resolving production problems for the applications and Ensure all support service level agreements are met.

1. **Client: HSBC Bank**  
   **Duration: June 2011 to July2012.**  
   **Role: Pl/Sql Developer (Operational Support System**)

**Description: --**

The system has centralized database to maintain all relevant information regarding Credit Information. The transaction data will be ported to this centralized database from the branches either through data import or through manual entry. The system allows the user to generate various kinds of reports and queries.

As a Programmer was involved in the following: -

* Involved in the whole life cycle of the project.
* Interacted with clients to understand their requirements.
* Programmed Packages, stored procedures, queries, and triggers on Oracle 9i.
* Involved in database design, Unit and System Testing.
* Identified problems in existing production data and developed one-time scripts to correct them.
* Provided Database maintenance and support for the continuation of the business (BAU).

**Environment:** Oracle 11g/10g, Toad, UNIX, Informatica

# **Academic Qualification:**

|  |  |  |  |
| --- | --- | --- | --- |
| **Degree** | **Institute** | **Percentage** | **Year** |
| B.Tech (Information Technology) | Durgapur Institute of Advanced Technology and Management, Durgapur (West Bengal) | 7.8(CGPA) | 2006-2010 |
| XII(BIEC) | A.N.S College, Barh, Patna | 64 | 2005 |
| X (BSEB) | Balie High School, Barh, Patna | 79 | 2003 |

**Personal Details:**

**Name :** Santosh Kumar

**Permanent Address : Kazichak** Barh, Patna, Bihar, India

**Languages Known :** English & Hindi

I declare that the information given above is true to the best of my knowledge.

###### Current Place: --Kolkata Name: Santosh Kumar