# SWATHI KUMAR

Email Address  : swathiinfo21@gmail.com

Mobile  : +91 6362751462

Technical SUMMARY

* **16+** years of experience in systems Analysis, design and development of complex software systems in Telecom, Healthcare and Financial sectors including 5+ years in Big data space.
* **Having USA B1 Visa (2015 – 2025).**
* **6+** years of experience in Big Data Technologies with full project development, implementation and deployment.
* Deep understanding of Hadoop Architecture and various components such as HDFS, Job Tracker, Task Tracker, Name Node, Data Node and MapReduce concepts along with Hive, Pig, Sqoop, Oozie, Zookeeper, Map Reduce framework and NoSQL databases like HBase,MongoDB.
* Experienced in processing, analyzing large sets of structured, semi-structured and unstructured data.
* Used Hadoop’s Tool interface to reduce programming effort for running MapReduce jobs.
* Expertise in writing Hive and Pig scripts to perform data analysis on large data sets.
* Created Hive Internal/External Tables and views.
* Leveraged the feature of custom UDFs in Pig and Hive to process data.
* Transferred data from Twitter, Avro using Apache Flume to perform further data analysis.
* Imported data to HDFS and exported data to Relational Databases using Sqoop.
* Indexed data in Hive to perform faster data operations.
* Worked in data formats such as TextFile, Sequence File, Row Columnar and Optimized Row Columnar in Hive.
* **Direct client handling experience in USA and Dubai**.
* Experienced in requirement analysis, coding and testing various modules in a software development life cycle.
* Adapted Agile methodology in the projects undergone and used testing strategies like **TDD** and **BDD**, which in turn invokes **Continuous Delivery and Integration.**
* Worked as Team lead, and lead a team in three assignments and successfully implemented the project before given deadline.
* Worked closely with product management and feedback given for regional trends, any perceived gaps and assisted with product roadmap and strategy.
* Developed and managed relationships in strategic markets and accounts including building customer networks, especially with C-levels and senior executives.
* Perform hands on design work to prove software capabilities to customers during the evaluation (Proof of Concept) process.
* Implemented new ways of working, improve internal processes.
* Provided internal sparring, helped review outputs and provided a fresh perspective.
* Provided input to new product development.
* Excellent team player and proven track record working in various team sizes performing cross-functional roles.
* Having 8yrs of strong ETL and data integration experience in developing ETL mappings and scripts using Informatica Power Center 8x, DataStage 7.5 and Oracle 9i.

Work Experience – 16 years

**Senior Data Engineer in Decision Minds – VMWARE from Jan 2021 to March 2023**

**Lead Big Data Analyst in SPXFLOW from Mar 2018 to Dec 2020**

**Analytics advisory specialist/Senior Consultant** in **Accenture** from **Mar2017** to **Feb2018**

**Architect/Senior Consultant** in **TCS** from **Nov 2013**to **Feb 2017**

**Technical project Lead** in **WIPRO** from **Nov 2011**to **Oct2013**

**SeniorSoftwareEngineer** in **Virtusa** from**June2006** to **November 2011**

Technical expertise

|  |  |
| --- | --- |
| **ETL Tools** | InformaticaPowercenter 9.6, DataStage 7.5 |
| **Hadoop/Big Data** | HDFS, MapReduce, HBase, Pig, Hive, Sqoop, Oozie, Apache Spark, Flume, MongoDB |
| **Web Technologies** | HTML, ASP.net,JavaScript,XML, Tomcat 6.0,JBoss4.0 |
| **J2EE Technologies** | Servlets, JSP, JDBC Struts, Spring and Hibernate |
| **IDE / Tools** | Eclipse, MiStudio3.5 IDE |
| **Languages** | Core JAVA, R language , Python – Data Science |
| **Database** | ORACLE 11g/10g/91,MySQL |
| **Cloud technologies** | GCP, Snowflake, Airflow |

|  |  |
| --- | --- |
|  |  |

TRAININGS UNDERGONE

* Hadoop/Big Data (Apache, Cloudera) – Training in TCS
* **R and python language** for DataScience – Training in Accenture
* InformaticaPowercenter 7.1 training in VirtusaPvt Ltd.
* DataStage 7.5 training in VirtusaPvt Ltd.

projects in DM – vmware

|  |  |
| --- | --- |
| **Project 1** | Tesseract - Vanguard |
| **Description** | * This project deals with the customer satisfaction survey details. * Migration of workloads and data assets from EDW 1.0 to EDW 2.0 data hubs. * This works on a Data Ingestion framework- Marlin. * Tesseract is VMware’s Enterprise Data and Analytics platform, a natural extension to the existing Enterprise Data Warehouse(EDW) that is leveraged by 23000+ VMware users across all the Business functions (Marketing, Sales, Finance, HR, Partner, SaaS, Customer Experience) & Business Units. The Platform is designed to scale and support with Curated/Certified Data & business Insights to help accelerate VMware’s transition to a modern multi-cloud SaaS future. * Data are from different sources like – salesforce, SAP, My learn etc. * By using this Marlin – we are processing the data. Processing pipelines deals with spark and python. |
| **Team Size** | 5 |
| **Duration** | From Jan 2022 to March 2022 |
| **Reporting Tool** | Tableau, SAP BOBJ |
| **Environment** | Hadoop/Big Data - Hive, SAP HANA, python, SQL, Snowflake, Airflow,GCP |
| **Role** | * Preparation of Project plan and tracking to the plan. * Risk identification, analysis and response planning. * Stake holder Management (Internal, Client and Vendor Personnel). * Execution, Monitoring and controlling project level activities. * Resource Management.   **Technical role:**   * Enabled the objects in HIVE – by writing HQL scripts. * Had experience in – table creation and optimization techniques – **by using SQL** – for audit log reports. * Creating new **python scripts**/updating based upon the requirement for new objects. * Writing **HQL’s** for Target table load and staging load. * Had experience in Informatica – for creating workflow’s and scheduling the jobs. * Had experience in – Deployment and migration of code using GitLab commands * Participated in team discussions to develop useful insights from big data processing results. * Developed simple to complex etl pipelines In and out of Datawarehouse using combination of DBT and snowflake’s SnowSQL as well python wherever it pitches in. * Processed data loading in Big Query from Google cloud storage using Google Data-Proc * Designed and developed and delivered data integration/data extraction/data migration using to GCP. * Building Airflow pipelines. * Scheduled end to end using GCP cloud composer service. * Migrated ETL jobs to Google Platform. * Maintained BQ datasets for reporting requirements * Used most of the services like Google Big Query, Google Cloud Storage, Google Dataflow, Cloud SQL, Google Cloud Data Proc, Google Pub/Sub, Sqoop and Py-Spark for data loading and ingesting.   Built and maintained data catalog in GCP |

|  |  |
| --- | --- |
| **Project 2** | PFCU – Partner federal credits Union |
| **Description** | * Objective of this project is to move the data from the data lake by applying required transformations to the semantic layer, which in turn can be used by the reports. * This project followed YODA Architecture & ELT/ETL Framework. * Move Symitar data through the pipeline to enable Reporting on Share Balances, Products, Member Counts, Branch metrics and digital usage counts. * Deploy a secure-by-default and scalable Azure Data and Analytics Platform (ADAP) to enable Partners’ future data end state and support the immediate goals of: * Leverage Azure Data Factory to connect to and ingest data from the ARCU data source, and two flat file sources (Cenlar and Co-op) * Leverage a SQL managed instance as the data warehouse repository for data (ensuring PCI/PII information is encrypted and masked) * Explore ADF Framework for advanced scheduling in Azure Data Factory * Understand how Azure DevOps can be leveraged for source control for integration pipelines in conjunction with Github * ExploreDataCatalog as a tool to define and enrich the data available in ADAP * Explore SSRS and PowerBIconsumption options to define the optimal solution to meet Partners’ requirements. * **Architecture** Blueprints for **Current** State and **Target architecture** |
| **Team Size** | 6 |
| **Duration** | From Jan 2021 to December 2021 |
| **Reporting Tool** | Tableau, Power BI |
| **Environment** | SSMS, ADF , ER win |
| **Role** | **Technical role:**   * Creating views based on the tables used for a particular report. * Developing report in Power BI * Creating Data Model * Creating semantic layer in azure before accessing the data through power BI reports. * Creating DAX expressions, query and functions. * Modifying the data pipelines in Azure based on the functionality and limitations. |

projects in SPXFLOW

|  |  |
| --- | --- |
| **Project 1** | **Decommission of Legacy Applications** |
| **Description** | * The data from the different applications like SAP ECC, Regional ERPs and Hybris etc are moved to the data lake, where the HDP platform tools came into picture. * Which deals with structured and unstructured data by the tools like Sqoop, Hive etc. * The transformed data have been stored into HANA. The reporting tool Sap BO/Power BI is used here. * R – studio is used for predictive analysis. This fetches the data from the HANA for querying and analysis. |
| **Team Size** | 7 |
| **Duration** | From Mar2018 to till date |
| **Reporting Tool** | Power BI |
| **Environment** | Hadoop/Big Data – Hortonworks, Hive, SAP HANA, R – studio |
| **Role** | * Preparation of Project plan and tracking to the plan. * Risk identification, analysis and response planning. * Stake holder Management (Internal, Client and Vendor Personnel). * Execution, Monitoring and controlling project level activities. * Resource Management.   **Technical role:**   * Performed debugging and fine tuning in Hive for improving performance. * Developed/Tested/Implemented HDP nodes for data processing in a cluster environment. * Participated in team discussions to develop useful insights from big data processing results. * Handling Power BI Architecture & Components (Power BI Desktop, Power BI Service, Power BI Mobile, Power BI Embedded, Power BI Report Server, Power BI Visual Market Place) |

projects in Accenture

|  |  |
| --- | --- |
| **Project 2** | **GCW – HP and Hawkeye(Shipment Visibility)** |
| **Description** | This project helps the management to understand the clear picture of the items/products handled by the client - region wise as well and profit wise measures  In simple - it gives the details like - transaction profile, what was sold, value of what was sold, how it was sold and to whom it was sold.   * power BI is used as a front end tool and Talend as backend. Finally all the data after the transforamtions based on the client requirement will be moved to Azure storage. The EDL acts as a staging layer. * Python queries have been written for predictive analysis phase   **The project flow is given below:**   * Path -> file picking -> azure blob storage loc -> pipelines and data sets – the data will be copied from azure blob to SQL DW external tables * Azure data factory – storage location for extracts * GCW – team work will start after the source files got landed from external tables – to actual fact tables through store procedures * Fact table -> IUD filter table (final data) then views will be built. -> from the views we will build business logic – cube (SSAS, (MSBI)) . * processing the cube then final result will be sent to front end .   **Hawkeye:**   * This part deals with the shipment related process and tracking of the same for the three regions (EMEA, APJ, AMS). * Eg : Sense and response and Back to Back ordering . * **R** scripting is used in analysis for forecasting reports. |
| **Team Size** | 10 |
| **Duration** | From Mar2017 to till date |
| **Environment** | Hadoop/Big Data , HIVE, Talend, Python |
| **Role** | * Preparation of Project plan and tracking to the plan. * Risk identification, analysis and response planning. * Stake holder Management (Internal, Client and Vendor Personnel). * Execution, Monitoring and controlling project level activities. * Resource Management. * Status meetings with project stakeholders from customer as well as Accenture management   **Technical role:**   * Performed debugging and fine tuning in Hive for improving performance. * Developed/Tested/Implemented Talend jobs for data processing. * Participated in team discussions to develop useful insights from big data processing results. |

|  |  |
| --- | --- |
| **Project 3** | **Sensor Data Analytics Framework – Hadoop** |
| **Description** | With the number of sensor-embedded intelligent devices increasing exponentially, enterprises struggle to effectively manage the generated voluminous sensor data. Different sensors imply different formats of data, which are difficult to correlate. Creating analytical models for this varied Big Data to provide alerts to end users in real time is an increasingly challenging task. In today's globalized market, leveraging this sensor data to identify strategic insights is essential to sustaining competitive advantage.  The Accenture SDA framework addresses these challenges by enabling organizations to collect, process, store and analyze the voluminous sensor data. Its powerful analytics engine leverages the MapReduce paradigm of Big Data processing to analyze large volumes of data in parallel, generating actionable insights rapidly. With the capability to configure and integrate various data sources and formats, SDA ensures data quality and facilitates analysis of integrated high volume sensor streams. |
| **Team Size** | 7 |
| **Duration** | From Aug 2016 to Mar2017 |
| **Environment** | Hadoop/Big Data ,SPARK,HBase,Sqoop,Oozie, HIVE, PIG |
| **Role** | * Preparation of Project plan and tracking to the plan. * Risk identification, analysis and response planning. * Stake holder Management (Internal, Client and Vendor Personnel). * Execution, Monitoring and controlling project level activities. * Resource Management. * Status meetings with project stakeholders from customer as well as Accenture management   **Technical role:**   * Handled different time series data using HBase to perform store data and perform analytics based on time to improve queries retrieval time. * Developed MapReduce programs to parse the raw data and store the refined data in tables. * Performed debugging and fine tuning in Hive & Pig for improving performance. * Used Oozie operational services for batch processing and scheduling workflows dynamically. * Worked in Spark Streaming component and performs RDD transformations on those mini-batches of data * Exported the analyzed data to the relational databases using Sqoop for visualization and to generate reports for the BI team. * Participated in team discussions to develop useful insights from big data processing results. * Integrated Map Reduce with HBase to import bulk amount of data into HBase using Map Reduce Programs. |

projects in TCS

|  |  |
| --- | --- |
| **Project 4** | **PeriVista – 360 Customer Insights for Retail (HDP)** |
| **Description** | Using Big Data and advanced analytics, PeriVista Shopper helps connect the dots by bringing together disparate customer data across internal and external sources for easy analysis and interpretation. It not only helps discover and predict customer behavior from their profile, but also from purchases across channels and lines of business, formal and informal interactions and individual social media conversations. The solution empowers retailers to reimagine the customer's journey across channels, predict crosschannel behavior, and thereby drive deeper engagement and enhanced customer experience |
| **Team Size** | 5 |
| **Duration** | From Aug 2015 to Aug 2016 |
| **Environment** | Hadoop/Big Data ,SPARK,HBase,Sqoop,Oozie, Flume |
| **Role** | * Involved in developing the retail data model. * Worked in Predictive and contextual retail analytics data. * Involved in engaging solution providing team, which enables algorithmic matching of customer identity across multiple enterprise databases and social media. * Analyzed the web log data using the HiveQL to extract number of unique visitors per day * Collected log data and staging data using Apache Flume and stored in HDFS for analysis * Involved in forecast based on the present results and insights derived from data analysis. * Worked in Spark Streaming component and performs RDD transformations on those mini-batches of data * Integrated Map Reduce with HBase to import bulk amount of data into HBase using Map Reduce Programs. * Suggested trends to the higher management based on social media data. * Performed data analysis of Twitter by importing data to HDFS using Apache Flume. |

|  |  |
| --- | --- |
| **Project 5** | **Delta Airlines – Revenue Pipeline – Onsite** |
| **Description** | RPAD system deals with the finance/revenue related data. The monster report generated every week shows the revenue growth of DA account. This system comprises development as well as support environment. The RPAD system has been streamlined at 2007 to have a separate flow for the revenue process which includes set of DWH jobs. |
| **Team Size** | 10 |
| **Duration** | From feb 2014 to Aug 2015 |
| **Environment** | Informatica 9x, Teradata 13.0 |
| **Role** | * Travelled Onsite for Requirement gathering phase. * Ensure new development is flexible and scalable to support new initiatives such as modular solutions. * Acquiring the scoped application knowledge to provide required application support * Supporting the application in the production environment to ensure the smooth business operation * Designing, developing, testing, debugging and implementing any additional business requirement as required by the business user * Providing status report to the client based on the issues handled in the particular time frame window. |

|  |  |
| --- | --- |
| **Project 6** | **Silicon Valley Bank (SVB)** |
| **Team Size** | 12 |
| **Duration** | From Nov 2013 to Jan 2014 |
| **Environment** | Informatica 9x, Oracle 10g,BODS |
| **Role** | * Ensure new development is flexible and scalable to support new initiatives such as modular solutions. * Acquiring the scoped application knowledge to provide required application support * Supporting the application in the production environment to ensure the smooth business operation * Designing, developing, testing, debugging and implementing any additional business requirement as required by the business user * Providing status report to the client based on the issues handled in the particular time frame window. |

projects undergone in WIPRO

|  |  |
| --- | --- |
| **Project 7** | **Zurich - CDW Replacement Project** |
| **Description** | The Commercial Data Warehouse (CDW) is currently hosted an unsupported platform. There is an urgent need to secure it and thereby remove the operational risk of a critical failure hence there is a requirement for the development of commercial broker reporting and analysis tool on Zurich’s standard, supported infrastructure and software. |
| **Team Size** | 5 |
| **Duration** | From Nov 2011 to Oct 2013 |
| **Environment** | Informatica 7.1/8.1/9.1, Oracle 10g |
| **Role** | * Preparation of Project Plan and Tracking of the plan. * Stake holder Management (Internal, Client and Vendor Personnel). * Preparing and presenting proposal to client’s senior management. * Weekly status meetings with project stakeholders from customer   as well as Wipro management.   * Impact analysis, Effort and cost estimation. * Providing the technical solution for the production issues. * Onsite Offshore Coordination. * Quality Management. * Project Management activities. * Requirement Analysis & Preparation of FSD * Preparation of project plan. * IQA/EQA Activities.   **Technical role:**   * Decommissioning of the current CDW * Ensure new development is flexible and scalable to support new initiatives such as modular solutions. * Acquiring the scoped application knowledge to provide required application support * Supporting the application in the production environment to ensure the smooth business operation * Designing, developing, testing, debugging and implementing any additional business requirement as required by the business user |

projects undergone in virtusA

|  |  |
| --- | --- |
| **Project 8** | **PGATOUR- WAP Site** |
| **Description** | The aim of developing these web sites for the mobile lovers, to provide detailed information about games and notifications regarding scheduled as well as the live games.  Features worked on:   * Based on device operating system should popup soft upgrade or force upgrade. * Implemented to support Stable Ford scoring system. * Integrated new tour called "LATINO AMERICA" into the current application. |
| **Team Size** | 5 |
| **Duration** | From Jan 2007 to Sep 2007 |
| **Environment** | Java, Eclipse, Mi Platform (Mi Studio 3.5) |
| **Role** | * Developed custom components for WAP sites. * Fetch data from content provider. * Involved into code fix for the front end issues raised in the production environment for Test Phase III using the JSP and Java * Involved in to code fix for the Business Logic issues raised in the production environment. * Involved in updating the Technical Design Documents |
|  |  |
| **Project 9** | **Thomas PINK – API Helpers** |
| **Description** | * Thomas Pink is a retail luxury clothing business company that provides online shopping to offering life style products to its users. * This application serves as the entry point for the customer mobile site, to prepare data by using the external website. The application currently exposes Web services which can be consumed by any client application. |
| **Team Size** | 7 |
| **Duration** | From Oct2007 to May2008 |
| **Environment** | Java, Eclipse –(software- Core Java, Groovy, Servlet, REST Web Services, MiStudio 3.5) |
| **Role** | * Requirement Analysis for the project. * Involved in the design of the application. * Developing and coding the different modules of the application |

|  |  |
| --- | --- |
| **Project 10** | **Micro Finance solutions – MFX** |
| **Description** | Financial Inclusion isthe provision of a broad range of financial services such as deposits, loans, payment services, money transfers, insurance to poor and low-income households and their micro enterprises. About 90% of the 180 million poor households still lack access to institutional financial services. Poor and low-income households have a large demand for safe and convenient deposit services. This demand reflects the importance of savings for these households for a variety of reasons |
| **Team Size** | 5 |
| **Duration** | From June 2008 to Dec 2008 |
| **Environment** | Languages : : Java, J2ee  Technologies : Struts, Hibernate, Servlets, Jasper Reports, Ajax  Web / App Servers: Apache Tomcat and JBoss |
| **Role** | * Developed web application using Struts Framework. * Developed user interfaces using JSP, HTML and CSS. * Used Eclipse as IDE tool to develop the application. * Worked heavily with the Struts tags used struts as the front controller to the web application. * Implemented validation framework for creation of validation.xml. * Worked on Subversion for software configuration management and version control. |

|  |  |
| --- | --- |
| **Project 11** | **BTGS – MDM PRODUCT BUILD DEVELOPMENT** |
| **Description** | This I This module includes :   * System configuration * Price updates/product data model updates * Data cleanse * Bug fixes that are not handled by Bridge * BCS(Business Change system) * Contract rollouts |
| **Team Size** | 14 |
| **Duration** | FromDec 2008 to Feb2009 |
| **Environment** | LION (Logical Interface for Ordering Network) Database |
| **Role** | * Worked as a Data Base Engineer. * Involved in the development of new products. * Gained thorough understanding of the existing reporting system. |
| **Project 12** | **THOMSON MEDSTAT- DATASTAGE DEVELOPMENT** |
| **Description** | * Business problem includes cleansing of raw data arriving from different sources to a common standards set by Thomson Medstat. * Filtering the customers based on particular claim like Facility header claims, Facility detail claims, Professional claims, Drug claims and Dental claims. * Setting eligibility criteria for a particular customer based on the claims. * Comparing the client data, transforming the data using business logic * Reusability components like shared containers, used for different customers * Improves performance by using optimal techniques like Transformer stage, Aggregator and Sort stages |
| **Team Size** | 9 |
| **Duration** | From Mar 2009 to Feb 2010 |
| **Environment** | Informatica 7.1, Datastage 7.5, CORR , Capella and ONYX |
| **Role** | * Created client shell and custom template jobs. * Initial design of transformation design work sheet by gathering business requirements from the client. * Develop jobs using Transformer, Aggregator, Sort stage and shared containers to meet business requirement * Rigorous unit testing done for the business logic for all permutation and combinations. * Tracking the change requests from the customer after the job development * Test case/data preparation, execution and verification of the testresults. |

|  |  |
| --- | --- |
| **Project 13** | **Cardinal Health – Medical Business Transformation** |
| **Description** | * Business problem includes cleansing of raw data arriving from different sources to a common standards set by Cardinal Health. * Filtering the customers based on particular claim like Medical claims and drug claims. * Comparing the client data, transforming the data using business logic * Reusability components like shared containers, used for different customers * Improves performance by using optimal techniques like Transformer stage, Aggregator and Sort stages |
| **Team Size** | 30 |
| **Duration** | From feb 2010 to Nov 2011 |
| **Environment** | InformaticaPowercenter8x |
| **Role** | * Worked as a Module Lead. * Conduct Technical Reviews and Audits of existing systems and Developed strategy document for new implementation. * Involved in Performance tuning of complex queries. * Created client shell and custom template jobs. * Initial design of transformation design work sheet by gathering business requirements from the client. * Allocate Jobs to the team members based on the priority. * Involved in designing project plan. * Tracking and sending the weekly report to project lead. |

Academic profile

|  |  |  |  |
| --- | --- | --- | --- |
| **Course** | **Institution/University** | **Percentage** | **Year** |
| B.Tech – IT  (Distinction) | AnnaUniversity. | 78% | 2006 |
| HSC | St.Mary’s convent HSC. | 83.5% | 2002 |
| SSLC | E.V.R.N. HSC. | 92.2% | 2000 |

ACHIEVEMENTS

* School Topper in SSLC (State level 5th Place)
* Topper (IT dept) in B. Tech with distinction (no history of arrears).

PERSONAL DETAILS

Name: Swathi Kumar

Gender: female

[Swathi K]