

SUDHAKAR BOMMISSETTY

sudhakar505@yahoo.in

Ph. +91-893-914-2520/850-034-3487

GitHub- <https://github.com/sudhakar505/Big-data>

LinkedIn-www.linkedin.com/in/sudhakar505



OBJECTIVE:

Actively looking for **Data Engineer** opportunity in EDW, BI, Analytics with **Challenging Environment** and can have a Chance to **learn and implement** optimized solutions using new technologies in the field of **Data Science**.

CERTIFICATIONS & COURSES:

- Cloudera Certified **CCA Apache Spark and Hadoop developer (CCA-175)** with License **100-019-318**
- Certified Training from **Datastax for Cassandra** NoSQL-database.

SUMMARY:

- Passionate about **Data Science** with an experience of **4 years** in building Enterprise data warehouse Applications.
- **Self-motivated** and passionate to **learn new technologies**.
- Bachelor's degree in **Computer Science & Engineering** from **Sree Vidyanikethan Engineering collage., Tirupathi, Andhara Pradesh**.
- Good Knowledge about **Health Care Domain**.
- Expert in ETL tools like **informatica** and **wherescape Red**.
- Expert in building data warehouse Applications using **informatica Power Center** and **Teradata db**.
- Strong knowledge in **SQL, PL/SQL-Procedures**.
- Good knowledge & Experience in **Unix Shell Scripts**.
- Good Experience in **Performance Tuning** in Informatica Power Center, SQL Queries in Teradata.
- Strong knowledge in **data warehousing concepts, Architectures and OLAP** functionalities.
- Had created **Shell Scripts** to reduce the manual tasks wherever Required.

Hadoop Eco System:

- Good knowledge in **Hadoop Eco System-HDFS, Apache Spark, Scala, Hive, Sqoop, Flume, Kafka, Impala & Zeppelin**.
- Good hands-on experience on Hadoop Eco system tools with **Cloudera distribution**.
- Strong understanding of **Apache Spark-RDD, Spark-Sql, Spark- Data Frames**.
- Hands-on Experience in building **Web-log Streaming Analytics** using **Spark, flume**.
- Hands-on Experience in **Integrating Spark with Cassandra, Hive, ingesting Tools Flume, Kafka**
- Good Knowledge about **Performance Tuning Techniques for Spark Applications**.
- Hands-on Experience in **Import/Export** data from **hdfs /hive to RDBMS** Using **Sqoop**.

TECHNICAL SKILLS & TOOLS:

Big Data Eco System : HDFS, Apache Spark using Scala, Hive, Sqoop, Flume, Kafka, Impala.
NoSQL Data base : Apache Cassandra
ETL Tools : Informatica Power Center 9.6.1, Wherescape Red 6.8.5
Data Bases & Utilities : Teradata 15.10, BTEQ, Fast Load, Multi Load, TPT, Fast Export
Scripting languages : UNIX Shell Scripting
Programming Skills : SQL, Basic PL/SQL

MY MAJOR ASSIGNMENTS:

Employer : Cognizant Technology Solutions Pvt Limited
Client : Blue Cross Blue Shield of Minnesota, U.S.
Role : Team Lead/Data Engineer/ETL Developer
Duration : 2014-April to Present

TECHNOLOGY: Hadoop Eco –system Tools, Apache Spark-sql, Hive, Hdfs, Shell Scripts

TREO-Data-Quality- Treo is a vendor to BCBSMN. As Part of the Project, have to load all the Treo data files created by Traditional tools into Hadoop in incremental fashion and Apply the Business Rules to data as per Business needs and send the Pass/ fail Report to Vendors for analysis Purpose.

BHI-Mandate- Blue Health Intelligence is a vendor to BCBSMN and High Mark. As Part of the Project, High Mark can send Monthly files to us which has High Mark Members information and can merge with BCBSMN data extracted by Traditional Tools to a Pipe delimited Flat files and Zip the files and send to Vendors for their Further Analysis.

TECHNOLOGY: ETL Tools –Informatica Power Center, Wherescape Red, Teradata, Shell Scripts

BHI- Blue Health Intelligence is a vendor to BCBSMN where we have to send the Monthly Healthcare Members related reports like claims, benefits, Products, Groups, client, etc. to a Pipe delimited Flat files and Zip the files and send to Vendors from BCBSMN to BHI for their Further Analysis

MOP- Member Out of Pocket (MOP) is a vendor to BCBSMN where we have to send the Weekly Healthcare Members related reports like claims, benefits, Products, Groups, client, etc. to a Fixed Width Flat files and Zip the files and send to Vendors from BCBSMN to MOP for their Further Analysis

Verisk- Verisk is a vendor to BCBSMN where we have to send the Monthly Healthcare Members from BCBSMN to Verisk for their Further Analysis

RESONSIBILITIES:

- Understand the Requirements from Business Users/Onsite Co-coordinators.
- Distribute the Requirements to the team according to their skills and make them to give Optimized solutions and work as a team to solve any challenging task.
- Translating the Requirements into design documents HLD, LLD.
- Have developed a Solution to Load the data from Source to Target systems.
- Have Extracted all the relevant data from data warehouse like claims, benefits, Products, Groups, client, etc. to a Fixed Width Flat file

- Preparing the DDL's Scripts, Unit Testing, TPR, Migration templates, Unix Shell scripts and Collect Stats, TPT insert statements SQL's Scripts
- Created Mappings, Sessions, Workflows as per Requirements.
- Handover the Code to ETL QA team to test and do follow up to fix the bugs if any.
- Deployed Code to PROD and Run, share the Result to the Client for BAT.
- Zip the files and send to Vendors.
- Prepare the KT document and give KT to support team.

POC PROJECTS:

Being Passionate about Data Science Have Explored below data sets as POC Projects and Uploaded Codes to **GitHub**.

- Have done Analytics on Below data sets from **Kaggle**.
 - H1 B Visa Petitions 2011-2016
 - 515k Hotels Reviews data in Europe.
 - Amazon Unlocked Mobile
- Have done POC on Spark- Cassandra Integration On **Movies By Actors** Dataset from **KillrVideos**.
- Have done Analytics on **Chicago Crimes** open datasets Using Spark and store back results to hdfs/ hive.

AUTOMATIONS or VALUE ADDS:

- Received appreciation from client for delivering the code on time even though requirements changes at last minute for BHI Project.
- Created Innovation-**Automation Shell Scripts which loads the multiple tables**, Data from Higher environments to lower environment using Fast load and Fast Export scripts in seconds which reduces the manual intervention to create tables or refresh the tables could Save Human effort, Cost and Money for the Client.
- Created the **Visual Basic Scripts** which can do the multiple Find & Replace tasks in Excel sheets (mapping docs) in a single shot which mostly helps for Data Modelers to change the requirements in mapping doc.
- Involved in Creating an **Automation Shell Script** could fetch the Table Names, Column Names, and Alias Names form a **SQL Scripts** which could save Cost and uses for Query Analysis.

Place:

Date:

(Sudhakar B)