**Hepzibha S Email:** [hepssibah@gmail.com](mailto:hepssibah@gmail.com)

**Big Data Analyst** **Mobile**: +1 608-709-6213

Linkedin : [https://www.linkedin.com/in/hepzibha](https://www.linkedin.com/in/hepsi-s-6abb4994)

## **Professional Summary:**

* Currently working as a Big Data Analyst with DSS Advanced Business Intelligent and Infrastructure Analyst Data Management Team in TDS.
* Good experience in data modeling and data analysis with very strong backend sql.
* Experience with Hortonworks, Cloudera, cloud and Amazon web services AWS and cloud9.
* Good understanding with experience and exposure in big data concepts and hive, pig queries.
* Advanced skills in Excel and Access.
* Experience in hive queries, pig scripts, python with Hadoop eco system and R scrips, programming and tableau visualization and analytics.
* Experience in Data modeling process, conceptual, logical and physical models, ERD and data warehouse design-Dimensional-Star, snowflake schema, Data Integrity and OLAP and facts, indexing, data dictionary
* Very good professional experience in all phases of **Software Development Life Cycle (SDLC)** including Design, Implementation and testing during the development of software applications along with **Hadoop, Big data, Oracle Database(11g, 12c), Sql Server 2012, MySQL, Derby etc. Oracle database Warehouse (DBW), Informatica Power Center, IBM Data Manager** for BI, Data Analysis, Data Extraction, Data Transformation and Data loading, **Data Mart & IBM Cognos Reporting tool** to create cubes and reports.
* Experience in create reports using Microsoft Power BI
* Strong experience in Writing **Complex HIVE & SQL Queries and Scheduling jobs using windows job scheduler and Oozie,** testing and implementing **triggers, Stored procedures, functions, Packages and Pl/Sql,** building reports using **Tableau, Statistical and pattern, algorithmic analysis using R-Programming, excel and pivot tables and Hadoop eco system.**
* **Hands-On** experience in coding with Java (Map-Reduce), python scripts and php, html, xhtml, xml, pl/sql and Informatica ETL.
* **Full stack experience in J2EE programming with different web services (RESTful, Restless) different Architectures (MVC Architecture) and different frameworks (Struts & Spring Frameworks), Java design patterns (Singleton, Factory etc.).**
* Experiencein **Front-end programming, web designing and analysis with JavaScript, JSP, PHP,HTML, CSS**
* Experiencein **Back-end programming EJB and JVM, develop, debug and unit testing with Eclipse, IBM Web Sphere**
* **E-Commerce** experience with the supply chain, retail website.
* Experience in data mining extremely large data sets, high proficiency in **SQL-Oracle**, **MySQL,** **MS SQL server 2012, oracle** relational database systems, JDBC connection and design.
* Experience in Data migration and upgrade in different databases & Cloud and Amazon Web Services (AWS).
* Sound Knowledge in **RDBMS, SQL and NOSQL databases (MongoDB) and Open Sources(Apache Couch DB) and the Storage Systems(Hadoop Distributed File System(HDFS)**
* Very good understanding in **Object Oriented Programming (OOPs) concept and** System Designing **with Microsoft Visio and Framework Manager.**
* Experience in **Enterprise Java Programming with IBM Web Sphere**
* Experience in **UNIX shell scripting and unix commands**
* Very good understanding and Experience in **Web Logic and JBOSS**.
* Experience in **data warehouse and OLAP** data
* Experience in SQL **Query tuning** and **instance tuning** as the part of **performance tuning**.
* Excellent Analytical, problem solving skills and a motivated team player and interpersonal skills.
* Functional expertise includes Telecom, Manufacturing, Health care, Insurance, Banking, Retail Store, supply chain, Insurance, Stock market, automotive and Financial Accounting.
* Experience in UNIX shell scripting.
* Good knowledge and hands-on experience in Hadoop, big data technologies such as Hive, Pig map reduce, Yarn, oozi, sqoop and cloud environment such as AWS
* Experience in Kafka, Spark and machine learning
* Good experience on installing Hadoop in AWS environment and worked on POC’s to extract data from traditional database to HDFS environment, data transformation and data ingestion.

**Education**:

* **Database and Data Analysis**

**University of California Santa Cruz (UCSC)**

* **Data Mining** **Master of Philosophy (MPhil)**

**Madurai Kamaraj University (MK University)**

* **Master of Computer Application (MCA)**

**Madurai Kamaraj University (MK University)**

* **Bachelor of Science (BSc) – Physics**

**Manonmaniam Sundaranar University (MS University)**

**CERTIFICATION**:

* Database and Big data Analysis – professional Certification – UCSC (Tableau, Hadoop, R-programming, Sql & NoSql databases, cloud computing and AWS)
* **Oracle Certified Professional (OCP)** in Oracle 11g, Advanced PL/SQL programming- 98%.
* **Oracle Certified Associate (OCA)** in oracle 11g Database administrator I -93%.
* **SAP Business Objects Web Intellignence: Report Design II**
* **Medical Billing & Coding**

**Workshops**:

* **IBM Cognos 11 Analytics Workshop – Nov 2016**
* **Hadoop 1.1 Workshop – May 2014 - Setup for single node for Word count**
* **Hadoop eco system Training– Hadoop 2.0 – March 2017 – setup nodes in cloud using Cloudera (Hive, Pig, Hbase).**
* **Big data Analysis with medical devices-Hadoop 2.0 – Real-time Analytics on Medical Device Data with Kudu and Stream sets**

**Courses**:

* **Big data, Hadoop Eco System, Map Reduce, Data Modelling, Java, Python, Database Management & Data Warehouse, Google Analytics, R-Programming, Tableau, Pig, Hive.**
* **Php, UNIX server, Object Oriented Programing Analysis and Design, XML, XHTML.**

**Technical Skills**:

|  |  |
| --- | --- |
| **Hadoop Tools & Concepts** | Hive, Oozie, Hbase, Flume, Hcatalog, Pig & HDFS, Map-Reduce, YARN, Hadoop Eco System, zoo keeper, sqoop, Mahout (Machine learning library), Hive, Pig, R connectors |
| **Programming & Scripting Languages** | Java, Python, php, JavaScript, PL/SQL, SQL, HTML, XHTML, XML,UNIX shell scripting |
| **Java, J2ee specific skills** | J2EE Common Services APIs, Web Services, Java JSP, Servlets, Struts driven web sites, JDBC connections, Hibernate, Applet, SOAP, RESTful, Restless JUNIT, Eclipse, My Eclipse, IBM WebSphere. |
| **Database** | Oracle 11g, Oracle 12c, MS Access, MySQL, SQL Server 2012,  NoSQL- Mongo DB, Couch Base, PostgreSQL DB, Oracle Siebel CRM. |
| **Data Ware house** | Kimball-DW/BI Life cycle methodology |
| **Data Modeling Tools** | Erwin, IBM Infosphere, DIA, Microsoft Visio |
| **Web Server** | Apache Tomcat 7.0, Derby, |
| **ETL Tools-BI, BO** | SAP Webi, Crystal Report, IBM Data Manager & Informatica power center |
| **Designing and modeling Tools** | OOAD- DIA, MS-Visio, Star UML, IBM framework Manager |
| **Tools & Utilities** | Eclipse IDE, Net Scape,JDK1.6,SQL\*Plus, SQL & PL/SQL Developer, SQL \* Loader, CVS, SVN, JIRA, JAMA, IBM RAD, WebSphere, Golden 6, Erwin |
| **Domain** | Manufacturing, Health care, Medical Insurance, Telecom, Automotive, finance |
| **Internet Technologies** | Oracle Web Tool kit, web services |
| **Analytics, Visualization tools** | IBM Cognos, R Studio, R-Programming, Tableau8.0, Pentaho, Advanced Excel VBA, Macro, Microsoft Power BI |
| **Methodologies & Frameworks** | Agile, Scrum methodology, LD-Lean Software Development, XP-Extreme Programming, RAD-Rapid application development and Water fall methodology |
| **Operating system** | Windows Vista/ XP/2000/7/8, UNIX, Ubuntu |

## **Professional Experience:**

**TDS Telecom, Madison, WI Mar 17 – Till Date**

**Big Data Analyst**

**Responsibilities:**

Working in the Business Intelligence and BigData Analysis team

* Working with telecom billing and financial data and discussions with stack holders to decide the design and migrations.
* Working with statistical analysis patterns and create the dashboards for quick references by self interest.
* Worked with streaming and Dataware housing projects
* Worked in Json scripts, mongodb and unix environment to non-sql data clean-up grouping and create the analysis reports
* Involved in Stack development plan discussions and decision making for the Recovery management, backup and migrations.
* Experienced with move partial data to cloud for the long run.
* Working on building the reports using SAP Webi and Crystal Reports
* Working with the data modeling team to create the data models using Erwin
* Working and expertise in data profiling.
* Rescheduling the jobs with various prompts and parameters
* Experience in CA Erwin Data Modeler- build models using CA Erwin’s Data Modelers Design Layer Architecture
* Good understanding in MetaSolv, vantage, Sibel and sabre data
* KPI dashboard, score card and visualization charts and reports using tableau
* ETL operations to create the report from different data sources.
* Create Ad hoc report and dynamic data visualization
* Used Kimball DW/BI Data ware house Life cycle
* Involved DBA and BI activities
* Knowledge with LSR data aggregation changes from data center exa-data to Hadoop big data process
* Working on weekly on-call process and work Assign and scheduling through Microsoft Access and BI launch pad.
* Working experience in CMC, BI launch pad and SAP BI, BO client and billing applications.
* Analyze, develop and deploy the SAP Crystal Report and Webi Reports.
* Working with queries, building universes and creating the reports based on the requirement, filtering, grouping and modifying the data.
* Write the python script for data warehousing project.
* Write Hive query for ad-hoc analysis report to analyze the customer feedback.
* On-call, triage and server management with query tuning.
* Create reports using Power BI, Tableau, SAP business objects- Crystal Report, Web Intelligence(Webi)
* Closely working with Infrastructure Architecture team and DBA, interact and co-ordinate with customers
* Familiar with Golden 6, sql developer and different analysis tools.
* Build the model and using Machine learning algorithms for predictive analysis.

Hadoop eco system, HDFS,Hive QL, HBase, hortonworks & cloudera, Mongodb, BI Launch pad, SAP Business Object Webi, Crystal Report, MySql, Oracle Database, SQL Server 2013, python,Sql Developer, CMC, SAP Business Intelligent client, Golden 6, erwin

**Madison Area Technical College (MATC), Madison, WI Oct 16 – Mar 17**

**Data Analyst**

**Responsibilities:**

Working in the IT services with Data Management team for different customer teams (enrollment, student help and HR, payroll etc.) to secure the data based on FERPA and create different reports based on the requirement.

* Working on building the cubes and reports
* Create the Business metric KPI (Key Performance Indicator) to evaluate factors for different modules.
* Create the catalog in data manager, fact build, dimension build, reference dimension, customize create and deploy ETL jobs from the transaction data
* Working on database migration, upgrade and maintenance.
* Cleansing, mapping and transforming data, create the job stream, add and delete the components to the job stream on data manager based on the requirement
* Use Service Now to incident management.
* Job schedule through the windows job scheduler.
* Migrate data from Dev environment to PROD
* ETL process on the data ware housing and OLAP data.
* Create the look-up tables for the data processing
* Experience with creating different reports and cubes using Cognos.
* Pivot tables and create various analysis report using MS-Excel
* Using statistical R-packages and R-programming for Factor, quantitative Analysis and k-means clustering.
* Using python script API for R- programming analysis
* Tableau reports and dashboards created and distributed in pdf format to the administration team.
* Worked on Database migration BI integration and cloud conversion.
* Big data Architectural analysis and Hadoop eco system migration with vm.
* Cost analysis and data analysis to decide the data center and platform configurations

IBM Cognos, Tableau, R-Programming Data manager, DataMart, Sql Server 2012, MS-Excel, Service Now.

Student, HR, payroll & Blackboard module

**Deloitte, Madison, WI Sep 15 – Sep 16**

**Web Analyst**

**Responsibilities:**

Worked in CARES and ACCESS projects different modules with Architecture & Maintenance Function Area (FA6) teams for Govt.Wisconsin, Department of Children and Families (DCF) & Department of Workforce Development (DWD).

* Worked as a Analyst in back end operations with Object Oriented Programming.
* Job scheduling, pl/sql module development using advanced pl/sql concepts collections, dynamic sql
* Write code to interface with external java applications.
* Performance tuning and sql query tuning. Write and tune the pl/sql code effectively to maximize performance
* Involved in System Analysis and Design methodology as well as Object Oriented Design and development using OOAD methodology to capture and model business requirements.
* Responsible for Module development, web services(RESTful, Restless), Frameworks(MVC), Java design patterns(Factory, singleton, etc.), JVM & Memory management and tuning
* New page Design, object mapping and migrate to html5 tag changes
* Incident Management through JIRA and JAMA
* Development and Unit test using IBM WebSphere
* SVN version control management for code repository.
* Experience with different development and maintenance projects
* Full stack custom design for the new implementation.
* Mapping the backend data from DB2 and Oracle 11g to the j2ee applications
* Cares, Access module integration and implementation based on the law updates
* Packages, functions, procedures, triggers written in oracle 11g sql, pl/sql.
* Job schedule, migration and update.
* Efficient J2EE stack designing analysis and implementation for different applications.
* XML scripts written for different purposes.
* Actively work as the Database Administrator and handle the database architecture using Erwin and kimball

Java EE ,JSP, Servlets, JSF, Spring DI/IOC, Hibernate, XML, HTML, JS, CSS, DB2, Web services, Rational Software Architect, Web sphere Application Server, UNIX, Junit, Log4J, SVN, Linux / Windows, oracle 11g, JIRA,JAMA and Wiki, Kimball, Erwin

**UCSC Extension – Santa Clara**  **Mar 15 – Aug 15**

**Big data Research Analyst**

**Responsibilities:**

Worked as a Hadoop Developer / Administrator with big data stack design architecture research team to build a strong data repository by developing, installing and configuring Hadoop ecosystem components that moved data from individual servers to HDFS.

* Developed MapReduce programs to parse the raw data, populate staging tables and store the refined data in partitioned tables in the Enterprise Data Warehouse (EDW).
* Create Hive queries that helped market analysts spot emerging trends by comparing fresh data with EDW reference tables and historical metrics.
* Enabled speedy reviews and first mover advantages by using Oozie to automate data loading into the Hadoop Distributed File System and Pig to pre-process the data.
* Provided design recommendations to improve review processes and resolved technical problems.
* Managed and reviewed Hadoop log files.
* Tested raw data and executed performance scripts
* Shared responsibility for administration of Hadoop, hive and pig.
* Installed and configured MapReduce, Hive and the HDFS
* Implemented CDH3 Hadoop cluster on CentOS. Assisted with performance tuning and monitoring.
* Created HBase tables to load large sets of structured, semi-structured and unstructured data coming from UNIX, NoSQL and a variety of portfolios.
* Created reports for the BI team using sqoop to export data into HDFS and Hive
* Developed multiple MapReduce jobs for data cleansing and preprocessing.
* Assisted with data capacity planning and node forecasting.
* Collaborated with the infrastructure, network, database, application and BI teams to ensure data quality and availability.
* Administration for Pig, Hive and Hbase installing updates, patched and upgrades.
* Installed, Designed, build and compare Hortonworks and MapR for the better performance and research purpose.

Data Mining, Data Analysis and Data Profiling using Hadoop, Amazon Web Services-AWS, Hive, Map Reduce, Pig Scripts, Oozie, Sql, HDFS, Python, java, CDH3, GitHub and excel, Hortonworks, MapR, Apache Falcon, Sqoop, Google Analytics.

**TLS IT - Stone Bridge, System Technology Sep 12 - Feb 14**

**Data base Administrator**

Worked as a Pl/sql developer/ Database Administrator (DBA) in TLS IT Solution network team for different applications.

**Responsibilities:**

* Coordinated with the front end design team to provide them with the necessary stored procedures and packages and the necessary insight into the data.
* Worked on SQL\*Loader to load data from flat files obtained from various facilities day to day.
* Created and modified UNIX shell scripts to loading the cleansed data into the base tables.
* Developed pl/sql triggers, stored procedures, functions and packages for moving the data from staging area to data mart.
* Created scripts to create new tables, views, sequences and queries for new enhancement in the application using TOAD.
* Performed sql and pl/sql tuning using various tools (EXPLAIN PLAN, SQL\*TRACE, TKPROF and AUTOTRACE etc.).
* Used bulk collections for better performance and easy retrieval of data by reducing context switching between sql and pl/sql engines.
* Extensively involved in using hints to direct the optimizer to choose an optimum query execution plan.
* Used Pragma autonomous transaction to avoid mutating problem in database trigger.
* Error handled using Exception handling and used advanced features and concepts (Dynamic Sql, collections, Bulk binding)
* ETL process through Informatica Power Center
* Worked with various sections of the tools such as debugger, target load plan and incremental aggregation to process data in Informatica
* Migrate the components such as sources and targets to another region using the Designer and Repository Manager screens.
* Used pushdown optimization and partitioning like tips to enhance code performance
* Partitioned the fact tables and materialized views to enhance the performance.
* Involved in Logical & Physical database layout design.

Informatica Power Center, Java, sql loader, DataMart, TOAD, Sql Developer, Oracle, UNIX.

**Client: Mepline Technical Services Nov 05 – Aug 12**

**Software Developer**

Intelligence Information System is deployed for use by the MEP Quantity Surveyors to prepare estimation sheet. Auto CAD drawing measurement, area calculations, commission percentage, unit calculations and point calculations, length of wire and pipe, type of wire and pipe and its cost, man power, time limit to finish the work. All calculations done by internally. In the form through pull down menus Quantity engineers select their options for the calculations and enter the needed details in the text box and get the report automatically and they can see the preview before printing. It easy to get the estimation sheet and help to finish their work without risk.

**Responsibilities:**

* Tracking the transactions and create the Analytics reports using Google analytical tool in the web based system.
* Involved in full life cycle of Intelligence Information System application development.
* Build model of this IIS system modules HLD & LLD.
* Transferred business narrative into a graphical representation (flow diagram) of business information needs and rules.
* Confirm and refine the model with the analysts and experts.
* Achieved to store the AutoCAD diagrams and its storage locations in the database.
* Creating and adding new fields to existing database.
* Participated in meetings with development team & support team.
* Installed and configured environment for both development and production.
* Created packages for easy handling.
* Handling Deployment Activity.
* Maintained Release Plan.
* Prepared Performance analysis report for both Oracle databases in development and production.

**Environment:** Oracle PL/SQL, J2EE, Hibernate, Google Analytics, Windows.

**Client: The Commerce Benefit Group Mar 04 – Oct 05**

**Role: Java Developer-Data Analyst**

Medical Accounting Administration system support business to handle various accounting activities, like different Ledger, Balance sheet preparation etc. It handles customer’s medical claiming forms, their accounting summary and detailed expenses. Valid expenses and their claiming amount pass to the approval section and keep it for their security.

**Project Description**

**Responsibilities :**

* Requirement gathering, analysis and customer interaction.
* Tableau reports and dashboards created and distributed in pdf format.
* Design the database object.
* Module development, Involved to create views for the Ford of Mexico region.
* Inline views join conditions given based on the requirement and conditions.
* Packages, Procedures, Triggers, functions developed for this application.
* Prepared Application Design Document.
* Involved in data modeling- constraints, relationship decided based on the business logic.
* Enhancements.
* Data Warehouse (DW) data integrated from different sources in different format (PDF, TIFF,J PEG, web crawl and RDBMS data MYSQL, oracle, SQL server etc).
* Small change document preparation.
* Client interaction- interact with business unit, assigned schedule and organized the meeting.
* Prepared unit test cases.
* Upload the developed database codes (Views and database objects) to the database under VIP schema in DEV environment and migrate to QA and Prod environment.
* Give privileges and create synonym to these objects.
* Packages, stand alone procedures and functions, triggers compiled and executed using DBCR-ADS tool.

**Environment:** Java 1.6 Hibernate web services framework, Oracle PL/SQL, SQL developer, Tableau 8.2, Oracle 11g, Access, NOSQL connector, Mongo DB, Linux – putty.

**Projects:**

**E—commerce Online retail business**

**Responsibilities :**

* Develop the web site designing
* Dynamic page creation and connect with database
* Design the database objects and tables.
* Secure data and coding to avoid sql injection
* Module development for payment process, signup, login, add or remove items from cart and

**Environment : PHP, HTML, Css, Java Script, Mysql**

**Web designing**

**Responsibilities :**

* Design and Develop the web site for small business
* Ten pages created and connect with database
* Design the database objects and tables.
* Create different forms and secure password
* Module development for send and receive email, footer, header and slider views

**Environment : XHTML, Java Script, Mysql**

**Trade market analysis (Big data Analysis)**

**Responsibilities :**

* Design and Develop the big data stack
* Analysis the financial trade market data
* Regression methods to build the model
* Graphs, charts for visualization
* Cleansing and mapping data.

**Environment : R-Programming, Statistical Analysis**

**Big data Analysis on Banking**

**Responsibilities :**

* Credit card analysis using mapR
* Stack design and development
* Virtual Machine setup and build the system
* Distrubuted system allocation
* Cleansing and mapping data.

**Environment : Map-R, hadoop**

**Marketing analysis (Big data Analysis)**

**Responsibilities :**

* Cleansing, mapping and grouping the data
* Analysis the business market data
* Design the dashboard & KPIs
* Generate pdf file to share to the Marketing team
* Cleansing and mapping data.

**Environment : Tableau, Excel, Microsoft Sql Server.**