**VENKATA SATYANARAYANA BILLA**

**Senior Spark Developer-Architect**

Mobile: **9866282929**

Email: venkatasbilla@gmail.com

# PROFESSIONAL SUMMARY:

11.5 years of experienced specializing in **Datawarehouse,Analytics ,Analysis ,Machinelearning ,Streaming Services** KAFKA with web based applications. Enjoys building relationships with stakeholders, understanding business issues, gathering and analyzing requirements, inventing creative solutions, documenting functional and technical design specifications, and working with Development, Testing, Production Support and Quality Assurance teams.

* 11.5 years of overall experience in IT industry, which includes hands on experience in Big data eco-system,Datascience related technologies like **Map Reduce, Hive,SPARK HBase, Pig, SQOOP, OOZIE, flume, Zookeeper and HDFS**,KAFKA,MLIB,NOSQL ,Mongodb,Machinelearning ,Deeplearning.
* 5.5 years of comprehensive experience as a Big Data & Analytics tools
* Very Strong Object-oriented concepts with complete software development life cycle experience – Requirements gathering, Conceptual Design, Analysis, Detail design, Development, System and User Acceptance Testing.
* Experience in working with MapReduce programs using Apache Hadoop for working with Big Data.
* Hands on experience in big data ingestion tools like **Flume and SQOOP**.
* Strong knowledge of Pig and Hive's functions, extending Hive and Pig core functionality by writing **UDFs**.
* Proficiency in different Hadoop distributions Cloudera and Hortonworks Distributions.
* Experience in analyzing data using **HiveQL, Pig Latin, Impala** and custom Map Reduce programs in Java.
* Hands on NoSQL database experience with Apache **HBase & Mongo DB**.
* Experience in adding and removing the nodes on Hadoop Cluster.
* Knowledge in job workflow scheduling and coordinating tools **like Oozie and Zookeeper**.
* Experience with various scripting languages like Linux/Unix shell scripts, Python 2.7 and Scala.
* Good Knowledge on **Apache Spark & Scala**.
* Good experience on **EMS, RV, JDK 5, JDBC, Servlets, JSP, Struts 2.1, Spring 2, AJAX, DOJO, JSON Hibernate 3, XML Processing (JAXB), JMS, and TIBCO**
* Strong front-end UI development skills using scripting languages like **JSP, HTML, JavaScript, JQuery and CSS.**
* Good experience on application server **WebSphere 6.0 and JBoss 4.x**.
* Experience to build tool like **Ant**.
* Good Knowledge on life cycle of **ETL** (Extraction, Transformation and Loading) using **Informatica Power Center** (Repository Manager, Mapping Designer, Workflow Manager, Workflow monitor).
* Developed number of complex Talend mappings and custom transformations to implement the business logic and to load the data incrementally.
* Extensively created mappings in **Talend** using **tMap, tJoin, tReplicate, tParallelize, tJava, tJavarow, tDie, tAggregateRow, tWarn, tLogCatcher,tMysqlScd and tFilter etc**.
* Utilized tStatsCatcher, tDie, tLogRow to create a generic joblet to store processing stats.
* Experienced in versioning, importing and exporting Talend jobs. Set up Triggers for Talend jobs in Job conductor.
* Good database experience using SQL Server, Stored Procedures, Cursors, Constraints and Triggers.
* Highly motivated, adaptive and quick learner.
* Domain Knowledge: Credit Cards domain experience with expertise in Non-Monetary & Monetary process, Payments, Fees, Collections, Statements functionalities, Banking, SMS Alerts & Email gateway solutions, IVR, Cards Origination, GL Accounting and middleware integration.
* Strong Analytical and Problem Solving skills, Multi-Tasking abilities, with proven experience in utilizing people and process knowledge to assist enterprises in making critical decision.

**Education & Certifications:**

* Bachelors in Computer Science Engineering, JNTU
* Achieved Cloudera Certified Developer for Apache Hadoop (CCDH).
* Good knowledge in IOT & Blockchain

**Technical Skills:**

|  |  |
| --- | --- |
| Hadoop/Big Data/Datascience Technologies | HDFS, Map Reduce, HBase, Hive, Pig, Impala, Sqoop, Flume, Oozie, Spark, SparkQL and Zookeeper.Machinelearning,Kafka,streaming ,Nosql,mongodb,Hbase,kafka, Tableau,Cloudera |
| Languages | Core Java, XML,HTML and HiveQL |
| J2EE Technologies | Servlets, JSP, JMS, JSTL, AJAX, DOJO, JSON and Blaze DS. |
| Frameworks | Spring 2, Struts 2 and Hibernate 3 |
| XML Processing | JAXB |
| Reporting Tools | BIRT 2.2. |
| Application & Web Services | WebSphere 6.0, JBoss 4.X and Tomcat 5. |
| Scripting Languages: | Java Script, Angular JS, Pig Latin, Python 2.7 and Scala. |
| Database(SQL/No SQL) | Oracle 9i, SQL Server 2005,MySQL ,DB2,HBase and Mongo DB 2.2 |
| IDE | Eclipse and Edit plus |
| Tools | SVN , Apache Ant, Junit and Star UML |
| ETL Tool | Talend Open Studio 5.5 |
| EAI Tools | TIBCO 5.6 |
| Bug tracking/ Ticketing: | Mercury Quality Center and Service Now |
| Operating System | Windows 98/2000 and Linux /Unix |

Professional Experience:

PRESENT PRODUCT: Telecom Billing (NGB Application): MTN

Role: Senior Hadoop Spark Developer-Architect

April 2015-TillDate

The purpose this Project NGB (Next Generation Billing System) to generate the INVOICES for millions of Subscribers in every Billing Cycle of Month. Already this Billing System implemented in ORCALE but it is consuming more time to generate the Invoices and This NGB product completely implementing with New Technologies BIGDATA, SPARK, SCALA, Distributed File system to process the data in MEMORY of Cluster to generate the 1 million invoices with in 10mins.

**Responsibilities:**

Articulate the architecture vision, conceptualize and experiment with alternative architectural approaches (Scala Application Architecture vs Java Application Architecture).

Designing Architecture’s SCALA Application and JAVA Application Architectures with DISTRIBUTED FILE SYSTEM and without Distributed File system and NOSQL storages (HBASE, MONGODB, and CASSANDRA).

Create models, component and interface specification documents, validate the architecture against requirements and assumptions.

Prepare documents and explain the architecture to sponsors and stakeholders.

Articulate the architecture vision, conceptualize and experiment with alternative architectural approaches.

Create models and component and interface specification documents.

Validate the architecture against requirements and assumptions.

Prepare documents and explain the architecture to sponsors and stakeholders.

Lead technical discussions related to new functionality, products and framework and ensure the code developed is reusable and configurable.

Lead defect prevention activity by aiming to maintain zero defects in E2E for assigned deliveries without any major escalation.

Set up and streamline process for Code review and Unit testing.

Perform and ensure code review is as per performance guidelines agreed with client for various releases (peer review). Drive automation and improvement initiatives across project (example customer account and product configuration automation).

Be actively involved in estimation process and reviews for the project.

**Environment:** Spark, Scala, Nodejs, AngularJs, Kafka, Hadoop, Java, Python, JavaScript, Zookeeper, Cluster, Redis, Spray Framework, Play Framework, Rpm, Akka, Docker, Jenkins, Git, Linux, Intelliji, Mysql, Cassandra,Mongodb, Docker.

**Client:MTN- APPLICATION**

**Role: Bigdata Architect**

The purpose of the project is to store terabytes of log information generated by the ecommerce website and extract meaning information out of it. The solution is based on the open source BigData s/w Hadoop . The data will be stored in Hadoop file system and processed using Map/Reduce jobs. Which intern includes getting the raw html data from the websites, Process the html to obtain product and pricing information, Extract various

reports out of the product pricing information and Export the information for further processing.

This project is mainly for the replatforming of the current existing system which is running on WebHarvest a third party JAR and in MySQL DB to a new cloud solution technology called Hadoop which can able to process large data sets (i.e. Tera bytes and Peta bytes of data) in order to meet the client requirements with the increasing competition from his retailers.

**Responsibilities:**

* Worked on Hortonworks (HDP 2.X) platform to perform Hadoop operations.
* Installed and configured Hive, Pig, Sqoop and Flume on the Hadoop cluster.
* Developed Simple to complex Map/reduce Jobs using Hive and Pig.
* The logs and semi structured content that are stored on HDFS were preprocessed using PIG and the processed data is imported into Hive warehouse which enabled business analysts to write Hive queries.
* Involved in loading data from UNIX/LINUX file system to HDFS.
* Handled importing of data from various data sources, performed transformations using Hive, MapReduce, and loaded data into HDFS.
* Continuous monitoring and managing the Hadoop cluster through Cloudera Manager.
* Installed and configured Hadoop MapReduce, HDFS, Developed multiple Map Reduce jobs in java for data cleaning and processing.
* Importing and exporting data into HDFS, Pig, Hive and HBase using SQOOP.
* Injected the data from logs and relational databases using Flume and SQOOP.
* Involved in gathering the requirements, designing, development and testing.
* Worked on loading and transformation of large sets of structured, semi structured data into Hadoop system.
* Worked on Hue interface for querying the data.
* Created Hive tables to store the processed results in a tabular format.
* Developed Hive Scripts for implementing dynamic Partitions.
* Developed Pig scripts for data analysis and extended its functionality by developing custom UDF's.
* Extensive knowledge on PIG scripts using bags and tuples.
* Experience in managing and reviewing Hadoop log files.
* Developed workflow in middleware to automate the tasks of loading the data into HDFS and pre-processing with Pig.
* Exported analyzed data to relational databases using SQOOP for visualization to generate reports for the BI team.

**Environment**: Hadoop (HDP 2.X), UNIX,SPARK Eclipse, TIBCO, HDFS, Java, MapReduce, Apache Pig, Hive, Flume, Python 2.7, SQOOP and MySQL,Kafka ,Aws ,cloudera

**Client: Asda Inc.**

**Responsibilities:**

* Responsible for building scalable distributed data solutions using Hadoop
* Created **HBase** tables to store variable data formats of PII data coming from different portfolios
* Implemented AWS Solutions using **EC2, S3, RDS**
* Good Knowledge in developing Java API to connect with **AWS S3 services** for storing and retrieving data stored in S3 buckets
* Developed API for using **AWS Lambda** to manage the servers and run the code in the AWS.
* Exported the analyzed data to the relational databases using **Sqoop** for visualization and to generate reports for the BI team
* Analyzed large amounts of data sets to determine optimal way to aggregate and report on it

         Developed Hive SQL in Hive Tables, configurations, set up and manage Hive over HDFS

         Developed Spark SQL and used in SQL Queries.

* Written **Hive queries** for data analysis to meet the business requirements
* Importing and exporting data into **HDFS** and **Hive** using **Sqoop.**
* Experienced in defining job flows.
* Got good experience with **NOSQL**database **HBase and Cassandra**
* Involved in creating **Hive** tables loading with data and writing hive queries which will run internally in map reduce way
* Designed and implemented **Map reduce**-based large-scale parallel relation-learning system
* Setup and benchmarked **Hadoop/HBase** clusters for internal use
* Implemented Hive Impala best practices to enhance query performance

**Environment/Tools:** Hadoop 2.0, Hive, Impala, Java, Sqoop, HBASE, Cdh5.2, sqoop, hbase,scala ,spark1.6, kafka aws ec2 ,tableau

**KOEING SOLUTIONS**

**June 2014 – Mar 2015**

**Bangalore**

**Client: W.W. Grainger,**

**Senior Bigdata Developer**

Grainger helps Customers save time and money by providing them the right products to keep their facilities up and running. Grainger works with more than 3500 suppliers to provide customers with access to more than 1 million products from variety of categories. Grainger’s diverse customer base comprised of facilities and maintenance professionals from businesses and institutions of all sizes

**Responsibilities**Solid Understanding of Hadoop **HDFS,** **Map-Reduce** and other Eco-System Projects

* Installation and Configuration of Hadoop Cluster
* Working with **Cloudera** Support Team to Fine Tune Cluster
* Working Closely with **SA** Team to make sure all hardware and software is properly setup for Optimum usage of resources
* Developed a custom **File System plugin for** Hadoop so it can access files on **Hitachi Data Platform**
* Plugin allows Hadoop MapReduce programs, **HBase**, **Pig** and **Hive** to work unmodified and access files directly
* The plugin also provided data locality for Hadoop across host nodes and virtual machines
* Wrote data ingesters and map reduce programs
* Developed **map Reduce jobs** to analyze data and provide heuristics reports
* Good experience in writing **data ingesters** and complex **MapReduce** jobs in **java** for data cleaning and preprocessing and fine tuning them as per data sets
* Extensive data validation using **HIVE** and also written **Hive UDFs**
* Involved in creating **Hive tables** loading with data and writing hive queries which will run internally in map reduce way
* lots of scripting (python and shell) to provision and spin up virtualized hadoop clusters
* Adding, Decommissioning and rebalancing nodes
* Created **POC** to store Server Log data into **Cassandra** to identify System Alert Metrics
* **Rack Aware** Configuration
* Configuring Client Machines
* Configuring, Monitoring and Management Tools
* **HDFS Support** and Maintenance
* Cluster HA Setup
* Applying Patches and Perform Version Upgrades
* Incident Management, Problem Management and Change Management
* Performance **Management and Reporting**
* Recover from Name Node failures
* Schedule **Map Reduce Jobs –FIFO and FAIR share**
* Installation and Configuration of other Open Source Software like Pig, Hive, HBASE, Flume and Sqoop
* Integration with RDBMS using swoop and **JDBC** Connectors
* Working with Dev Team to tune Job Knowledge of **Writing Hive** Jobs

**Environment:** AWS, EC2, S3, Lambda,Java, Apache HDFS Map Reduce, Pig Hive HBase Flume Sqoop, Cassandra

**SS**

**SYNVERSE TECHNOLOGIES,Hyderabad**

**April 2013- May 2014**

**Client: All State Insurance**

**SQL/Hadoop Developer**

The Allstate Corporation is the largest publicly held personal lines property and casualty insurer in America. Allstate was founded in 1931 and became a publicly traded company in 1993. Consumers access Allstate insurance products auto, home, life and retirement and services through Allstate agencies, independent, agencies. And Allstate exclusive financial representatives the company markets private passenger auto and homeowners insurance under the Allstate and Encompass and Esurance brand names. Esurance, an Allstate company, provides Esurance brand auto insurance products directly to consumers online, through call centers and through select agents, including Answer Financial

**Responsibilities**

* Developer Hadoop ecosystem : **Hadoop,** **MapReduce,** **Hbase,** **Sqoop,** **Amazon Elastic Map Reduce (EMR)**
* Developed a scalable, cost effective, and fault tolerant data ware house system on **Amazon EC2 Cloud.**
* Developed **MapReduce/EMR** jobs to analyze the data and provide heuristics and reports. The heuristics were used for improving campaign targeting and efficiency
* Response time for web services built on typical LAMP (php) stack was too slow developed a high performant / high volume / highly scalable platform for bidding in real-time understand from the client the extraction process and decide on the load strategy i.e. whether they want historical data or the current view
* Written complex **HSQL’s** to generate data required in the final reports and pass these **HSQL’s** to the Ruby programs to convert these **HSQL’**s to map Reduce programs
* Importing, exporting data into **HDFS** and **HIVE** using **Sqoop**
* Responsible for loading unstructured data into Hadoop file system (HDFS)
* Created and scheduled jobs for maintenance
* Configured **Database Mail**
* Monitored **File Growth**
* Maintained **Operators, Categories, Alerts, Notifications, Jobs and Schedules**
* Maintained **database response times**, proactively generated **performance reports**
* Automated most of the **DBA Tasks** and **Monitoring stats**
* Developed complex **stored procedures**, **views,** **clustered/non-clustered indexes**, **triggers** **(DDL, DML, LOGON**) and user defined functions
* Created a mirrored database using **Database Mirroring** with **High Performance Mode**
* Created **database snapshots** and stored procedures to load data from the snapshot database to the report database
* **Restore Development** and Staging databases from production as per the requirement
* Involved in resolving **Dead lock issues** and P**erformance issues**
* **Query Optimization** and Performance Tuning for **long running queries** and created new indexes on tables for faster **I/O**

**Environment:**  MS SQL Server 2005/2000, Windows 2000/2003 Server, DTS, Web Logic, Redhat Enterprise MS Access, XML, Hadoop, MapReduce, Hbase, Sqoop, Amazon Elastic Map Reduce CDH, Cassandra, NOSQL, Teradata.

**AJEL LTD, Hyderabad**

**Mar2011 to Feb2013**

**AEGON Insurance**

**SQL Developer**

AEGON is an international life insurance, pensions and asset Management Company with businesses in over 20 markets in the Americas, Europe and Asia. AEGON employs around 25,000 people and serves nearly 47 million customers worldwide.

**Responsibilities:**

* Installing, configuring Linux based systems
* Installed, Configured and Maintained Supporting open source Linux operating systems (CENTOS, Debian, Fedora)
* Monitoring the health and stability of Linux and Windows System environments
* Diagnosed and resolved problems associated with DNS, DHCP, VPN, NFS, and Apache
* Scripting expertise including BASH, PHP, PERL, Java script and UNIX Shell
* **Maintained** and **Monitored Replication** by managing the **profile parameters**
* Implemented **Log Shipping** and **Database Mirroring**
* Used **BCP Utility** and **Bulk Insert** for bulk operations on data
* Automated and enhanced daily administrative tasks including **disk space management** **Backup** and **recovery**
* Used **DTS** and **SSIS** to **Import** and **Export** various forms of data
* **Performance Tuning**, **capacity planning**, **Server Partitioning** and **Database security** Configuration are done on regular basis to maintain the consistency
* Created **alerts** and **notifications** to notify **system errors**
* Used **SQL Server Profiler** for troubleshooting, monitoring and **optimization** of SQL Server
* Worked with developers in creation of **Stored Procedures, triggers** and **User Defined Functions** to handle the complex business rules data and **audit analysis**
* Provided **24X7** on **call Support**
* **Generated reports** daily, weekly and monthly reports

**SQL Developer** Chirra Electronics Pvt Ltd is a company which provides customized solutions for DC-DC converters and SMPS. The project involved in integrating the “partner” (includes customers, design planners etc) information. The web system has reporting and a maintenance functions. Reporting access can be granted to most employees. Maintenance access is only granted to Customer & Sales Services employees. Most of the data in Partners comes from Pioneer Sales Reps, who enter information about their customers in the Field Information System (FIS), which sends updates to Partners.

**Responsibilities**

* To set up **SQL Server** configuration settings.
* Export or Import data from other data sources like flat files using **Import/Export** of **DTS.**
* Back up, package and distribute databases more efficiently by using **Red gate**
* Automate common tasks and use functionality in applications by using **Red gate**
* Rebuilding the indexes at regular intervals for better performance
* Designed and implemented comprehensive Backup plan and disaster recovery strategies
* Involved in trouble shooting and fine-tuning of databases for its performance and concurrency.
* Expertise and Interest include Administration, Database Design, Performance Analysis, and Production Support for Large (**VLDB)** and Complex Databases
* Monitored and modified Performance using execution plans and Index tuning.
* Manage the clustered environment.
* Using log shipping for synchronization of database.
* Implementation of **SQL Logins,** Roles and Authentication Modes as a part of Security Policies for various categories of User Support.
* Monitoring **SQL server performance** using profiler to find performance and deadlocks.
* Maintaining the database consistency with **DBCC** at regular intervals

**RELIANCE GLOBAL SERVICES PVT LTD, Hyderabad**

**Role: Software Engineer**

**Sep 2007 to Jan 2011**

**Client:**

**Enterprise Knowledge Management (EKM), INDIA**

**Role: Java Developer**

Enterprise Knowledge Manager is a powerful platform which will brings to enterprises a new found way to manage all their portfolios and realize productivity gains. Backed by a full text search engine that allows Enterprises to locate policies, standards and other documents associated with their artifacts in a secure manner, ensures that not only can artifacts and their meta-data be tracked but even email exchanges or documents exchanged in the course of portfolio management are all available in every context. Cost estimation for the new and existing projects. Provides reporting to track the enterprise repository.

**Responsibilities:**

* Involved in design and development phases of Software Development Life Cycle (SDLC)
* Developed all parts of modules Artifacts, Reporting, Costing and Admin Modules.
* Designing and developing of the User Interface for the application using JSP and DOJO (for validation and AJAX requests).
* Developed the runtime for the business reports using BIRT runtime tool.
* Followed agile methodology and SCRUM meetings to track, optimize and tailored features to customer needs.
* Developed the middle layer development and using spring framework
* Worked on developing the ORM layer using Spring framework’s ORM module
* involved in making compatible for WAS
* Prepared the document for unit test scripts and testing results.
* Worked on defect resolutions raised at QA and Production level issues.

**Environment:** Core Java, Spring, Log4j, JSP, CSS, JDBC, Oracle, SQL Server 2005, DB2, JMS, JAXB, DOJO, JBoss, Web sphere, Eclipse , XML , ANT and CVS.

**Client:**

**First Citizens Bank, INDIA**

**Role: Java Developer Oct 2007 – Nov 2008**

First Citizens Bank is one of the largest banking institutions in the world. Bank offers various financial and banking services to its customers. Worked on application named Access portal. It is a part of online banking that allows a customer to view quick summary of transactions and account details. It also shows mutual funds associated with account.

**Responsibilities:**

* Developed JavaScript behavior code for user interaction.
* Created database program in SQL server to manipulate data accumulated by internet transactions.
* Wrote Servlets class to generate dynamic HTML pages.
* Developed Servlets and back-end Java classes using Web Sphere application server.
* Developed an API to write XML documents from a database.
* Performed usability testing for the application using JUnit Test.
* Maintenance of a Java GUI application using JFC/Swing.
* Created complex SQL and used JDBC connectivity to access the database.
* Involved in the design and coding of the data capture templates, presentation and component templates.
* Part of the team that designed, customized and implemented metadata search and database synchronization.
* Used Oracle as Database and used Toad for queries execution and also Involved in writing SQL scripts, PL SQL code for procedures and functions

**Environment:** Java, Web Sphere 3.5, EJB, Servlets, JavaScript, JDBC, SQL, JUnit, Eclipse IDE and Apache Tomcat 6