# Kalyani Maganti Contact: +1-732-476-8952

Big Data Consultant [Email-Id:magkalyani@gmail.com](mailto:magkalyani@gmail.com)

*Sun Certified Java Developer*

*Hortonworks Certified Hadoop Associate*

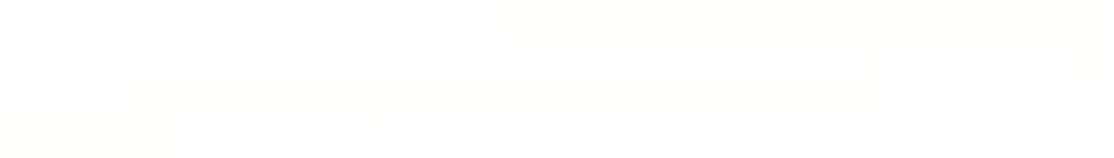


**PROFESSIONAL SUMMARY**

Results-driven IT consultant with 7+ years of business and technology experience implementing and overseeing multi-million dollar applications.

Expertise includes:

* **Intuitive Understanding** of the **Value of Data** (incl. Scala, Hadoop, Spark, Spark streaming, Spark Mllibs, Spark GraphX, and NoSQL Technologies) ***Where***, ***How*** and ***Why*** to Apply **Technology** (incl. New and Emerging Technologies) to **Solve** Business’s **Problems** and identify **Opportunities**.



* Offer comprehensive knowledge of applications including BPM, Complex Event Processing (Tibco Business Events), and **BIG DATA** applications.



**SPECIALTIES**

* Functional Programming concepts – creating succinct, clean and extensible codes without regressive defects
* Unstructured Large Scale Data Processing, transformation of data and processing it to meaningful entity structure.
* Recommendation Algorithm on Large Scale Datasets: Developing algorithms with Spark
* High Velocity Data Ingestion and ETL processing from data sources such as Twitter, Stock feeds or any other social mediaplatform.



**CERTIFICATIONS**

* Sun Certified Java Developer
* AWS Developer Associate
* Hortonworks Certified Hadoop Associate



**TECHNICAL SKILLS**

**Operating System** IBM AIX, Sun Solaris, HP-UX, Linux, Windows NT/2000

**Big Data Eco System Spark, Spark Streaming, Kafka, ElasticSearch, HBase,**

**Cassandra,** HDFS, MapReduce, Hive, Pig, Flume, Sqoop

**Languages** Scala, C, C++, Java, Python,PLSQL, SQL, **Hive, PIG Functional Programming** Scala, Monoids, Monads

**Database** Oracle, SQL Server, Informix, Sybase, Cassandra, HBase, MongoDB

**Cloud Servers/Devops** Saltstack, Vagrant, Docker Microservices, Ansible, Openstack,

AWS

**Tools** Maven, SBT, Ant scripts, Python scripts, GIT, SVN, Packer



**EDUCATION**

Bachelors of Technology, GITAM UNIVERSITY, INDIA



**PROFESSIONAL EXPERIENCE**

**JC Penney –** Plano, TX **June 2012 - till date**

## Sr Big Data Consultant

* Installed and configured Hortonworks Distribution Platform (HDP 2.4) on Amazon EC2 instances with 100 nodes
* Experienced on loading and transforming of large sets of structured, semi structured and unstructured data from HBase through Sqoop and placed in HDFS for further processing.
* Hands on experience on configuring Capacity scheduler.
* Configured YARN Queue Manager to accept multiple applications by setting User limit factor.
* Used Sqoop jobs for importing order data as snapshots to hdfs datalake.
* Used Python 2.7 scripting for injesting incremental data into hdfs.
* Developed automated scripts to import data from S3 to Datameer in a Datameer compressed format.
* Created FTP jobs (JSCAPE) to import data from COREMETRICS detailed files into datalake (hdfs/S3) through edge node.
* Developed MapReduce programs and Hive queries to analyze sales pattern and customer satisfaction index over the data present in various relational database tables.
* Implemented search on HDFS using operational metadata stored in ElasticSearch
* Analysed the data using the hive queries to find the top selling products in particular region during festive seasons inorder to increase its production in the future,and this analysed is has imported to tableau for visualization graphically.
* Provided ad-hoc queries and data metrics to the Business Users using Hive, Pig.
* Designed workflows and coordinators in Jenkins to automate and parallelize import jobs (Sqoop, Spark, JScape, etc) on Apache Hadoop environment by Hortonworks (HDP 2.4).
* Developed a fully automated continuous integration system using Git, Jenkins, MySQL and custom tools developed in Python and Bash.
* Experience in understanding the security requirements for Hadoop and integrating with

Kerberos authentication infrastructure- KDC server setup, creating realm /domain, managing

* Real streaming the data using Spark with Kafka.
* Experience in configuring Zookeeper to provide high availability and Cluster services coordination.
* Debug and proposed solutions to the power users and business users for the data extractions and analytics using the Datameer and HDFS files.
* Utilised market basket analysis to discover and understand customer purchasing behavior.
* Created aggregated view of data by integrating Apache Spark with ElasticSearch and indexing on elasticsearch for near realtime analytics and aggregation queries
* Extensively worked with ElasticSearch issuing ad-hocqueries.
* Expertise in using tableau for graphical visualization of the end result.

**SYNERGY SHARP –** San Ramon, CA **Aug 2010-May2012**

## Sr Big Data Consultant

* Provided architectural guidance for overall product development.
  + Created Yarn clusters on AWS with Hadoop, Spark, HBase.
  + Responsible for building scalable distributed data solutions using Hadoop.
  + Developed simple to complex MapReduce jobs using Hive and Pig.
  + Extensive experience setting up Hadoop clusters on Amazon EC2 with HDFS/YARN across 20 nodes.
  + Exported the analyzed data to the relational databases using Sqoop for visualization and to generate reports to BI team.

. • Managed datasets using MySQL, queried MYSQL database queries from

python using Python-MySQL connector to retrieve information from MySQLdb package.

* + Used hive to analyse the partitioned and bucketed data and compute various metrics for reporting.
  + Created components like hive UDFs for missing functionality in HIVE for analytics.
  + Experience in reviewing and managing the Hadoop logs.
  + Design and implement near real time data integration between kafka and Hadoop.
  + Mentor in Big Data stack Spark eco system to vast array of audience.
  + Define framework and best practices for projects using different technologies. Perform periodic code reviews to ensure code is following the defined best practices.

**AT&T –** San Ramon, CA **May 2008- Jun 2010**

## Software Engineer

* + - Responsible for architecture of back office applications.
    - Developed a POC for BOBPM using Spark and Hadoop intended to replace Tibco business events.
    - Provided trainings to back office teams in Cloud stack, Spark, and Hadoop.
    - Worked on upgrade from Tibco BE 3.x to BE 5.x.
    - Work closely with Production Support for Troubleshooting issues.
    - Determined various TIBCO product issues, worked with TIBCO Support and Engineering to determine a fix and getting a hotfix to resolve the issue.
    - Created Ontology’s consisting Concepts, Database Concepts, Scorecards and Events.
    - Responsible for writing Rule Functions, Virtual Rule Functions and Rule sets to handle events.
    - Created State modelers to model the lifecycles of instances of Concepts.
    - Used Decision Manager extensively to deploy business rules which are subject to change dynamically.
    - Implemented overall application based locking mechanism to avoid deadlocks/thread blocking in a multi-engine environment.
    - Created coding standards for all the Business Events applications and closely worked with Design Assurance teams to enforce them in back office teams.
    - Improved performance by determining critical points in the process flow where objects needs to be stored in local cache vs. distributed cache vs. backing store, thus reducing overall retrieval time.
    - Created BW instances to full fill the BPI requirements.
    - Designed, build, and tested adapter and web services integration using Tibco BW
    - Created queues in TIBCO EMS as per business requirements.
    - Involved in deployment of multiple engines in load balanced and fault tolerant mode across a cluster for high availability and performance
    - Defining and ensuring proper communication within team.