Job Title: Big Data Architect
Location: Chennai

Position Summary:

The Big Data Architect is an accomplished leader in Analytics, Data Engineering, Architecture, governance and Framework Design and someone who guides teams to success The Big Data Architect approaches business problems with creative thinking, and is focused on using the best forefront of next generation technology to address complex, value added real-world problems

Primary Function:

- Support presales for Big data opportunities
- Scope and execute big data projects for customers
- Support IP development for the practice

Primary Duties & Responsibilities
Delivery:

- Define client needs and oversee project milestones to ensure expectations, timelines, and budgets are met
- Define data platform architecture and design, have hands on capability to review code and make required changes
- Responsible for the overall quality of project deliverables and the successful implementation of defined solution for the customer
- Identify and qualify follow-on opportunities and engage senior leadership
- Establish procedures and recommend changes to policies that have a positive impact on the organization(s) and/or implementation team

Client Development:

- Occasionally interact with senior level management at client site or within the company, which involves negotiating or influencing others on matters of significance
- Build long-term, superior client relationships and proactively manage client expectations, and ensure that change control is used when scope boundaries are exceeded

Professional Development:

- Maintain a strong network and promote the organization at various meetings, forums, panels, publications, and conferences Begin to establish thought leadership in the industry
- Maintain technical certifications and attend training sessions to refine technical skills
- Responsible for oversight and apprentice training of junior resources as assigned in the field

Requirements Basic Qualifications

- Bachelor's degree in Computer Science, Information Technology, Engineering, or a related field
- At least eight (8) years of Information Technology work experience, to include all of the following
- Ability to deliver end to end Advanced analytics strategy, ranging from architecting Open source based Big data-analytic tools and building integrated solutions that meet emerging enterprise

needs Hands-on experience with various forms of data design, such as OLTP, OLAP, ODS, EDW (3rd normal), BI data marts (Star and Snow Flake schemas), DSS, with knowledge of various BI tools and data management tools

- At least six (6) years of experience in Structured, Semi-Structured and Un-structured data with progressive responsibilities in Data Warehousing and Business intelligence, data architecting, and analytical modeling techniques
- At least four (4) years of experience in Big Data technologies in Hadoop ecosystem – such as Hive, HDFS, MapReduce, Yarn, Kafka, Pig, Oozie, HBase, Sqoop, Nifi, and/or Ranger
  Other Position Requirements –
- Experienced in customization and optimization of anyone of the big data distributions: MapR, Cloudera or Hortonworks (HDP and HDF)
- Expertise in distributed, Columnar, MPP, NoSQL & Document Storage DBs in depth and scale
  Expertise in architecting, designing and developing big data frameworks for real time and batch analytics and ability to do capacity planning for large enterprise with petabyte scale
- Proven coding experience in anyone of the following programming language Scala, Python or R and Advanced knowledge in Scala/Python frameworks
- Proven knowledge in Spark 22 framework
- Demonstrated ability to configure and optimize and fine-tuning clusters for various Big data moving parts like Kafka, Nifi, Spark and HDFS
- Knowledge in Big Data multi data center distribution (active/active or active/passive) configuration and DiscCP
- Interpersonal skills to be able to work with Data Scientists and covert legacy machine learning models to Spark ML batch jobs using PySpark

Preferred Qualifications:

- Knowledge in HDFS, Cassandra or MongoDB
- Knowledge of any one of the following cloud big data offerings: Azure, HDInsight, or AWS EMR
- Master’s degree in Computer Science or Data Analytics
- Integration of Big data eco system with advanced ML products like Ayasdi or Linquamatics or data visualization products
- Knowledge of integrating PySpark with native anaconda libraries deep learning libraries like keras or tensor flow