GANESH G

Bangalore | (+91) 9885451224 | gganesh@outlook.in
in: linkedin.com/in/ganesh-g-94a484163 | **S**: gganesh_2

CCA cloudera
SPARK & HADOOP DEVELOPER

Profile

Certified Spark & Hadoop developer with 3 years of experience in dealing large volume of data, comprising loading from multiple sources to HDFS, effective storing, complex processing using various Hadoop frameworks and provide processed data to different environments.

Skills

- Apache Spark - Scala
- Apache Impala
- Apache Map Reduce
- Apache Oozie
- Apache Sqoop

- SQL & Hive QL
- Apache Flume
- Apache Pig
- Shell / Bash
- Java / Scala

Certification

CCA Spark and Hadoop Developer (CCA175):   23 Aug 2017 – 23 Aug 2020
License: 100-021-092
Certificate authority: Cloudera Inc.

Professional Summary

- In-depth knowledge of Hadoop architecture and various components such as HDFS, Resource Manager, Application Master, Node Manager, Name Node, Data Node and Map-Reduce concepts.
- Extensive knowledge in reviewing Hadoop log files, analysing data and developing code using hive QL, Pig Latin, Impala and Hadoop Map-Reduce applications, Spark applications in Scala.
- Very good exposure in designing both time driven and data driven automated workflow using Oozie and monitoring tools like Oozie and Zookeeper.
- Good exposure to monitoring, maintaining / managing Amazon EC2 Cloudera Distribution of Hadoop clusters using Cloudera Manager.
- Excellent technical, communication, analytical, trouble-shooting capabilities, problem solving skills and able to get on well with people including cross-cultural backgrounds.
- Good knowledge of version Controlling Systems like Apache Subversion and Git (Bitbucket / GitHub), and Jenkins for better control over building projects.
- Good working knowledge of IDEs like Eclipse, IntelliJ and used Maven, SBT build tools to achieve more functionality for build process.
- Very good understanding of Partitions, Bucketing concepts in Hive, used different file formats and compression codecs and designed both Managed and External tables in Hive and Impala to optimize and get most performance by managing storage issues.

Project

Digital Ad-Ratings (DAR):   Sep 2015 – Sep 2017
Nielsen – Digital Ad Ratings provides a new and highly accurate method for identifying and measuring the audiences exposed to online ad campaigns and enables measurement of any ad campaign, large or small, national or local, on any website. This approach leverages actual user demographics from leading online data providers to attach demos to larger portions of campaign impressions.

Work History

System Engineer: Sep 2017 - Current

Tata Consultancy Services   Siruseri, Chennai, TN.
Contact: Mozhiarasi J – +91 89391 47038
- Worked with admin team to automate cluster creation using AWS and Red hat ansible tower and Tibco. Involved in migration of existing project to Cloud AWS environment. (50% overall cost cutting).
- Received appreciation from customer for successful completion of moving code to Spark platform and cloud environment.
- Developed and deployed new jobs, made code changes to existing jobs that are required while migrating project to cloud environment.
- Involved in code migration from Impala to Spark-SQL and Spark core in scala which process raw data and store processed data back to Amazon S3.
- Stored data in different file formats with different compression codecs in HDFS for effective storing and which can improve performance and reduce I/O usage while reading from HDFS.

Assistant System Engineer: Sep 2015 - Sep 2017

Tata Consultancy Services   Siruseri, Chennai, TN.
Contact: Rajkumar Vairusamy – +91 98650 91165 / +1 (908) 947-8312
- Developed Map-Reduce programs to parse the data from multiple compressed / non-compressed file formats, populates staging tables and store the refined data in partitioned Hive / Impala tables.
- Solved performance issues in hive, impala with understanding of joins, grouping / aggregation and run time implementation of hive and impala (30% reduction in total time).
- Worked in complete Software Development Life Cycle (analysis, design, development, testing, implementation and support) adhering Agile Methodologies.
- Developed user defined functions (UDF) in Java when necessary to use in Pig and Hive queries.
- Involved in importing data from SQL Server, Oracle, csv and text from local/external file system to HDFS using different techniques on regular basis into Hive tables, which are partitioned, vice versa.

Education

Bachelor of Technology: **73.28% (Computer Science)**   2015
N. B. K. R Institute of Science and Technology   Vidyanagar, Nellore

Intermediate: **88.4%**   2011
Sri Chaitanya Sai Bharathi junior college   Vijayawada

SSC: **91.16%**   2009
A. P. Residential School   Lepakshi, Anantapuram