Gaurav Mishra

Bangalore, Karnataka - Email me on Indeed: indeed.com/r/Gaurav-Mishra/527ff276667c4594

WORK EXPERIENCE

Hadoop Admin with Multiplatform (Cloudera,Hortonworks, Apache hadoop)

IBM - Bangalore, Karnataka - July 2015 to Present

nstalled and configured HDFS,MapReduce, Hive ; implemented CDH , HDP Hadoop Cluster on CentOS.
Responsible to manage data coming from different sources and involved in HDFS maintainance and loading of various type of data.
Worked with delivery team to setup new Hadoop users. This job includes setting up linux users, Kerberos principal and testing HDFS,Hive,Pig & MapReduce access for the users.
Experience in adminstering large Hadoop clusters which includes activities like deploy Hadoop cluster, add and remove nodes ,keep track of jobs, configure name-node high-availabilty,schedule and configure and take backups.
Performed data pipeline using Flume, Sqoop and Java Mapreduce to ingest customer behavorial data and financial histories into HDFS for analysis.Worked on Sqoop jobs with incremental load to populate Hive external tables
Configured Fair Schedulers to provide Service Level Agreement for multiple users.
Supported Map reduce program those running on clusters.
Engaged in interaction with client and business users for new requirements and support.
Impact analysis, planning and implementation of change requests on the Hadoop environment.
Generate reports Weekly/Monthly and provide the business for Analysis .

Currently being assigned as a Hadoop admin in WPP which involves handling huge data on different Opcos. (Kantar,GroupM, JWT etc). Hadoop admin includes working on Apache hadoop as well as Cloudera & HotonWorks distribution platform

Key skills

Expertise
Details
OS Platform
Linux,
Languages
C, C++, Linux, CoreJava, SQL,Python , Bash Script ,
Hadoop Versions
Apache Hadoop 1.2 & 2.6 ,
Skills
HDFS, Mapreduce, YARN, Sqoop,Flume, ZooKeeper, Oozie Hbase (NoSQL) , Hue, Cloudera Manager, Impala, NIFI,Ambari
Distributed plaform
Cloudera (CDH 4&5)
HortonWorks (HW 2.2 & 2.3)

Project #
Project Name

Employer
Client
Duration
Team Size
Description:
Wire Plastic & Products Plc.
IBM
Wire Plastic & Products (WPP), UKI
July, 2015 – till date.
12
WPP Plc, (Wire and Plastic products) is a British multinational advertising and public relations company. They are mainly into marketing and advertising of different organization. It has taken over around 108 Opco. , out of which IBM supports for nine major Opcos.. There are two lower environment in each Opcos. There is an on-call system where one must support for 15 days in two months interval as the primary contact for Hadoop team and handle production issues in Off peak hours.The purpose of the project is to perform the analysis on the Effectiveness and validity of controls and to store terabytes of log information generated by the source providers as part of the analysis and extract meaningful information out of it. The solution is based on the open source Big Data software Hadoop. The data will be stored in Hadoop file system and processed using Map Reduce jobs, which intern includes getting the raw data, process the data to obtain controls and redesign/ change history information, extract various reports out of the controls history and export the information for further processing. We installed in Opco environment of 500 nodes which started with
90 nodes and with addition of nodes was done in every phase(90days) looking on stability of Opcos. Environment. For 9 Opcos. It was divided among team to handle 3 Opcos for each member .The platforms supported are HortonWorks, Cloudera &Apache Hadoop. We manage processing large sets of structured, semi-structured and unstructured data and supporting systems. I in As an admin the responsibilities includes Hadoop Cluster environment i.e adding and removing cluster nodes, cluster capacity planning, performance tuning, cluster Monitoring, Troubleshooting.

Project #
Project Name
Employer
Client
Duration
Role
Team Size
Description
:
:
:
:
:
:
:
:
2
Common Frameworks(CFX)
Tata Consultancy Services
American Express Company, USA
Feb 2014 – Jun, 2015.

Hadoop Administrator
5

American Express Company, also known as Amex, is an American multinational financial services corporation .The company is best known for its credit card, charge card and traveler cheque's businesses. The project involves hadoop support to amex platform. Hadoop environment was set of scartch of 20 nodes and we ended up in 170 nodes with phase to phase testing and moving it into production . The platforms that were involved where Apache Hadoop & Cloudera distributed platform. The hadoop environments were setup from root level which includes installation of OS and setting up hadoop environment , adding users , creating group providing necessary access (User& access management),setting up clusters(Cluster Management) .Beside hadoop support, was involved in migration in the last phase of project which included upgrading of hdfs.

Keywords : Bigdata technologies, HDFS, MapReduce, Oozie, Sqoop, Flume, Hive, Hbase, YARN, Cloudera & Hotonworks Disributed platform. HUE,
Scripting : Bash, Perl ,

Senior Software Engineer
TCS ION - TATA CONSULTANCY SERVICES   -   Bangalore, Karnataka - October 2011 to July 2015

I joined as a fresher and worked on Midlleware technologies.

2014: Got a chance to move into Big data technologies as a Admin . In a elaborated way worked on HDFS, MapReduce, Sqoop, Flume, Hbase, Hive , Python Scripting, Cloudera

EDUCATION

B.Tech in Mechanical Engineering
KIIT - Bhubaneshwar, Orissa
2007 to 2011

SKILLS

Bigdata technologies, HDFS, MapReduce, Oozie, Sqoop, Flume, Hive, Hbase, YARN, Cloudera & Hotonworks Disributed platform. HUE, Scripting : Bash, Perl , (2 years)

CERTIFICATIONS

SCJP Certified
SCJP certifified.