

Run Lead - Big Data Platform Run Lead - Big Data Platform Run Lead - Intel Salt Lake City, UT

Over 5+ years of comprehensive IT experience in BigData and BigData Analytics, Hadoop, HDFS, MapReduce, YARN, Hadoop Ecosystem and Shell Scripting. -5+ years of development experience using Java, J2EE, JSP and Servlets. -Highly capable for processing large sets of Structured, Semi-structured and Unstructured datasets and supporting BigData applications. -Hands on experience with Hadoop Ecosystem components like Map Reduce (Processing), HDFS (Storage), YARN, Sqoop, Pig, Hive, HBase, Oozie, ZooKeeper and Spark for data storage and analysis. -Expertise in transferring data between a Hadoop ecosystem and structured data storage in a RDBMS such as MY SQL, Oracle, Teradata and DB2 using Sqoop. -Experience in NoSQL databases like Mongo DB, HBase and Cassandra. -Have excellent knowledge on Python Collections and Multi-Threading. -Skilled experience in Python with proven expertise in using new tools and technical developments -Experience in Apache Spark cluster and streams processing using Spark Streaming -Worked on several python packages like numpy, scipy, pytables etc. -Expertise in moving large amounts of log, streaming event data and Transactional data using Flume. -Experience in developing MapReduce jobs in Java for data cleaning and preprocessing. -Expertise in writing Pig Latin, HiveScripts and extended their functionality using User Defined Functions (UDF's). -Expertise in handling structured arrangement of data within certain limits (Data Layout's) using Partitions and Bucketing in Hive. -Expertise in preparing interactive Data Visualization's using Tableau Software from different sources. -Hands on experience in developing workflows that execute MapReduce, Sqoop, Pig, Hive and Shell scripts using Oozie. -Experience working with Cloudera Hue Interface and Impala. -Hands on experience developing Solr Indexes using MapReduceIndexer Tool. -Expertise in Object-oriented analysis and design (OOAD) like UML and use of various design patterns. -Experience in Java, JSP, Servlets, EJB, Web Logic, Web Sphere, Hibernate, SpringJBoss, JDBC, RMI, Java Script, Ajax, JQuery, XML and HTML. -Fluent with the core Java concepts like I/O, Multi-threading, Exceptions, RegEx, Data Structures and Serialization. -Performed unit testing using Junit Testing Framework and Log4J to monitor the error logs. -Good Knowledge of Python and Python Web Framework Django. -Experienced with Python

frameworks like Webapp2 and, Flask. -Experience in process improvement, normalization/de-normalization, data extraction, cleansing and manipulation. -Converting requirement specification, Source system understanding into Conceptual, Logical and physical Data Model, Data flow (DFD). -Expertise in working with transactional databases like Oracle, SQL server, My SQL, and Db2. -Expertise in developing SQL queries, Stored Procedures and excellent development experience with Agile Methodology. -Ability to adapt to evolving technology, strong sense of responsibility and accomplishment. -Excellent leadership, interpersonal, problem solving and time management skills. -Excellent communication skills both Written (documentation) and Verbal (presentation). Authorized to work in the US for any employer

Work Experience

Run Lead - Big Data Platform Intel - Austin, TX June 2015 to Present

Manage onshore and off-shore Hadoop teams consisting of Developers, Architects, Data Engineers and Administrators. Report to stakeholders on bi-weekly basis on the overall health of the Enterprise Big Data platform. Oversee Synchrony core 5 Hadoop environments significantly ingesting large datasets sourced from real-time feeds, variety user interactions with mobile apps and internal transactions. Single point of contact for all platform operations ("Run") issues from immediate response, coordination, escalation, root cause analysis and resolution. Ensuring SLA's for availability, performance, security, maintenance - upgrades, installation and user administration across all Data Lake environments. Administer and maintain Pivotal/Hortonworks Hadoop, Greenplum and GemFire clusters across all environments. Installation/configuration of the Hadoop eco-system tools and continuous enhancement and expansion of the enterprise data lake. Collaborate with development teams to assist in code promotion across environments and deployments in production, including CMDB CI creation and updates. Proactively monitor cluster health and perform performance tuning activities. Perform capacity planning and expansion activities working across Infrastructure and other enterprise service teams. Perform cluster maintenance with patching/upgrades/migration, user provisioning, automation of routine tasks, re-processing of failed jobs, configure and maintain security policies. Ensure the Enterprise Data Lake initiative is continually providing unprecedented customer 360 capabilities and associated services at all time. Certified HDP Hadoop Administrator HealthScape

Advisors LLC - Chicago, IL February 2013 to May 2015 Cluster planning and engineering of POC and Production Clusters Strong experience on Hadoop distributions Hortonworks & Cloudera Administer, troubleshoot and debug cluster problems on RHEL/Ubuntu Troubleshoot and debugs cluster problems on RHEL based Linux infrastructure. Kerberos Administrator/Kerberize cluster and ensure sound domain name to-realm mapping Periodically Check HDFS health Configure YARN Capacity Scheduler based on infrastructure needs/YARN tuning Perform upgrades, patches and fixes using effective roll-out method Ensure HDFS is Balanced and performing optimally at all times Commission/Decommission Hadoop cluster nodes Review namenode WebUI for information concerning Datanode volume failures Back-up existing Hadoop databases such as oozie, hive metastore, HDFS Metadata backup etc Active Directory/LDAPS integration and management Purge older log files Build cluster according to workload pattern/cluster type Configure Cluster services for High Availability Ensure volume and hdfs encryption(Data at rest encryption) Experienced in AWS configuration optimization for Hadoop Backup Procedures and Disaster Recovery ACL's and audits to meet compliance specifications using Ranger Open tickets and troubleshoot cluster problems with support Experienced in AWS Storage methodologies Expertise in cloudbreak setup/configuration and blueprint creation. Engineer data pipeline management using Falcon Defining/executing feeds, processes, data pipelines, jobs mirroring between production and testing cluster using Falcon ETL offload using Atlas Securing hadoop services (Webhdfs/hive/yarn) using Knox for REST API calls. Trains and leads teams in understanding and implementing hadoop solutions in organization. Monitor cluster/organization servers for intrusions/detection, log file reviews and threats using SIEM solution Alien vault. Documents and reports findings to compliance and audit teams. Perform Active directory/windows server administration Big Data Analyst Lorven Technologies August 2010 to January 2013 Worked with most Hadoop distributions Hortonworks Developed data governance process and controls and ensured compliance with enterprise data architecture principles and standards for the various systems and components. Analyzing, profiling data for quality and reconciling data issues. Build, test and deploy Hadoop solutions using most Hadoop ecosystem components. Administer Hadoop

cluster, monitor performance with Ganglia. Design robust Hadoop solutions for complex business problems. Utilize new and latest Open Source tools for addressing business challenges. Work with multiple customer teams and support teams to execute Hadoop engagements. Integrated Hadoop into traditional ETL, accelerating the extraction, transformation, and loading of massive structured and unstructured data Python developer GAD GROUP TECHNOLOGY, INC - Chicago, IL June 2009 to July 2010 Build command-line app with Restful API layers that cater for end users who are mostly kids. Pull data out of HTML and XML file with Scrapy/Beautiful soup. Interacting with web-app using Flask. Performed Data analysis using Python Pandas. Processing Data records and returning computed results using Mongo DB Aggregation framework. Parse aggregated data into Apache Solr and graph database Orient DB. IT Analyst /Programmer Logos Infotech Inc - St. Louis, MO April 2007 to May 2009 Worked both independently and in a team-oriented collaborative environment. Worked with Microsoft SQL server. Documented and provided status of project and technical information related to the application/software supported (Web2Py & .NET). Supported remote users at their home office, hotel, or customer site utilizing remote tools and troubleshooting over phone using VPN. Knowledge of DSL/Cable Modem Routers, Windows Server 2012 and 2008, CISCO Switches and Routers and VOIP (i.e. Cisco Call Manager). Education Bachelor's Skills APACHE HADOOP HDFS (2 years), APACHE HADOOP OOZIE (2 years), ETL (4 years), EXTRACT, TRANSFORM, AND LOAD (4 years), Hadoop (7 years) Additional Information Technical Skills: Operating Systems: Linux, windows, Mac OS, Unix Cloud Techs: AWS, EC2, Elasticsearch, Cloudwatch, EMR. Hadoop Eco System: HDFS, MapReduce, Pig, Hive, SparkSQL, HBase, Apache Crunch, Solr, Sqoop, Spark Streaming, Spark, Oozie, Zookeeper, Hue, AVRO, Cassandra, MongoDB. Database Servers: HBase, MongoDB, DynamoDB, Cassandra, DB2, Teradata, MYSQL, Oracle, MS SQL Server 2005/2008/2012 ETL Design Tools: Teradata Load Utilities (BTEQ, Fast, Multi load), SSIS, Informatica. Report DesignTool: SSRS, Tableau. Application Servers: JBoss 7.x, WebSphere 6.x, WebLogic 11g, JBoss 5.0 Programming Languages: Java, Shell Scripts, Scala, Python and R. Web Technologies: Servlets, JSP, JSTL, JDBC IDEs: Eclipse, Netbeans, RAD, Jdeveloper, TOAD, SQL Developer Frameworks: Spring,

Struts, Hibernate 4.x/3.x, Hadoop, Impala.

Name: Michael Chaney

Email: phillipbrown@example.com

Phone: 283.452.8875x29969