

## **PUNEET MISHRA (Green Card)**

**Looking For Remote Work Only**

**LEAD HADOOP ADMIN**

**75017, Irving, TX**

**Email:** pmttech6633@gmail.com

- ✓ Over 10+ years of experience including 4 years of experience with Hadoop Ecosystem in installation and administrated of all **UNIX/LINUX** servers and configuration of different Hadoop eco-system components in the existing cluster project.
- ✓ Extensive experience in installing, configuring and administrating **Hadoop cluster** for major Hadoop distributions like **CDH5 and HDP**.
- ✓ Experienced in setting up automated monitoring and escalation infrastructure for Hadoop Cluster using Ganglia and Nagios.
- ✓ Experience with complete Software Design Lifecycle including design, development, testing and implementation of moderate to advanced complex systems.
- ✓ Hands on experience in installation, configuration, supporting and managing Hadoop Clusters **using Apache**, Horton works, **Cloudera and MapReduce**.
- ✓ Experience in importing and exporting the data using Sqoop from HDFS to Relational Database systems/mainframe and vice-versa.
- ✓ Hands on experience on configuring a Hadoop cluster in a professional environment and on Amazon Web Services (**AWS**) using an **EC2** instance.
- ✓ Excellent knowledge of in **NOSQL databases** like **HBase, Cassandra**. Experience in monitoring and troubleshooting issues with Linux memory, CPU, OS, storage and network.
- ✓ Experience in creating custom Lucene/Solr Query components.
- ✓ Expertise with the tools in Hadoop Ecosystem including **Pig, Hive, HDFS, Map Reduce, Sqoop, Solr, Spark, Kafka, Flume, Yarn, Oozie, and Zookeeper**.
- ✓ Used **NoSQL database with Cassandra, HBase, Hive, MongoDB and Monod**.
- ✓ Experience in developing and scheduling **ETL** workflows in Hadoop using **Oozie**. Also have substantial experience writing MapReduce jobs in **Java, Pig, Flume, Zookeeper and Hive and Storm**.
- ✓ Strong experience in System Administration, Installation, Upgrading, Patches, Migration, Configuration, Troubleshooting, Security, Backup, Disaster Recovery, Performance Monitoring and Fine-tuning on Linux (RHEL) systems.
- ✓ Extensively involved in Test Plan re-design, Test Case re-Creation, Test Automation and Test Execution of web and client server applications as per change requests.
- ✓ Experience in working on internet related technologies such as **HTML**, JavaScript, VBScript and XML.
- ✓ Having Strong Experience in **LINUX/UNIX** Administration, expertise in Red Hat Enterprise Linux 4, 5 and 6, familiar with Solaris 9 & 10 and IBM AIX 6.
- ✓ Created POC to store Server Log data into Cassandra to identify System Alert Metrics.

### **Technical Skills**

### Big Data Technologies:

HDFS, Hive, Map Reduce, Cassandra, Pig, Hcatalog, Phoenix, Falcon, Scoop, Flume, Zookeeper, Mahout, Oozie, Avro, HBase, MapReduce, HDFS, Storm, CDH 5.3, CDH 5.4

Tools Quality center v11.0, ALM, TOAD, JIRA, HP QTP, HP UFT, Selenium, Test NG, JUnit

Programming Languages Shell Scripting, Puppet, Scripting, Python, Bash, CSH, Ruby, PHP

QA Methodologies Waterfall, Agile, (TM) V-model.

Front End Technologies HTML, XHTML, CSS, XML, JavaScript, AJAX, Servlets, JSP

Java Frameworks MVC, Apache Struts 2.0, Spring and Hibernate

Defect Management Jira, Quality Centre.

Domain Knowledge GSM, WAP, GPRS, CDMA and UMTS (3G) Web Services: SOAP (JAX-WS), WSDL, SOA, Restful (JAX-RS), JMS

Application Servers Apache Tomcat, Web Logic Server, Web Sphere, Jboss

Databases Oracle 11g, MySQL, MS SQL Server, IBM DB2 NoSQL Databases HBase, MongoDB Cassandra Data Stax Enterprise 4.6.1

Cassandra RDBMS Oracle 9i, Oracle 10g, MS Access, MS SQL Server, IBM DB2, PL/SQL Operating Systems Linux, UNIX, MAC, Windows

## **Lead Hadoop Admin**

**Nov 2015 to present**

**Intuit - Irving, TX**

### **Roles and Responsibilities:**

- ✓ Worked on Distributed/Cloud Computing (Map Reduce/ Hadoop, Hive, Yarn, Hue, Pig, HBase, Sqoop, Flume, Spark, AVRO, Zookeeper, Tableau, Bedrock generic ODBC/JDBC, etc.) Horton works (HDP 2.2.4.2), for 4 clusters ranges from POC to PROD contains nearly 100 nodes.
- ✓ Provided 24x7 operation support for large scale Hadoop and MongoDB clusters across production, UAT and development environments.
- ✓ Performed a Major upgrade in production environment from HDP 1.3 to HDP 2.2. As an admin followed standard Back up policies to make sure the high availability of cluster.
- ✓ Implemented Flume, Spark, Spark Stream framework for real time data processing. Developed analytical components using Scala, Spark and Spark Stream. Implemented Proofs of Concept on Hadoop and Spark stack and different big data analytic tools, using Spark SQL as an alternative to Impala
- ✓ Involved in Analyzing system failures, identifying root causes, and recommended course of actions.
- ✓ Experienced in administrating Big Data clusters with HDFS, Kafka, Zookeeper, Hive, Yarn, Hue, Oozie, etc.
- ✓ Imported logs from web servers with Flume to ingest the data into HDFS.
- ✓ Involved in implementing security on Horton works Hadoop Clusters using Kerberos by working along with operations team to move non secured cluster to secured cluster.

- ✓ Responsible for upgrading Horton works Hadoop HDP2.2.0 and Map reduce 2.0 with YARN in Multi Clustered Node environment. Handled importing of data from various data sources, performed transformations using Hive, Map Reduce, Spark and loaded data into HDFS.
- ✓ Used MySQL databases by Cluster Configuring and maintenance of HDFS, YARN Resource Manager, MapReduce, Hive, HBASE, Kafka, or Spark;
- ✓ String understanding and experience of ODBC/JDBC with various clients like Tableau, Microstrategy, and server components; and
- ✓ Monitoring and tune cluster component performance.
- ✓ Migrated services from a managed hosting environment to AWS including: service design, network layout, data migration, automation, monitoring, deployments and cutover, documentation, overall plan, cost analysis, and timeline.
- ✓ Performance tuning of Hadoop clusters and Hadoop MapReduce routines.
- ✓ Working on this project spanning from Architecting, Installation, Configuration and Management of Hadoop Clusters.
- ✓ Managing Amazon Web Services (AWS) infrastructure with automation and configuration management tools such as Chef, Ansible, Puppet, or custom-built designing cloud-hosted solutions, specific AWS product suite experience.
- ✓ Development projects Extensively on Hive, Spark, Pig, Sqoop and Gem fire XD throughout the development Lifecycle until the projects went into Production. Created reporting views in Impala using Sentry Policy files.
- ✓ Responsible for Handler configuration and handler ESB mappings. Also Involved in Integrating Hive with Mulesoft ESB to land data into applications running on Sales force and vice versa.
- ✓ Implemented dual data center set up for all Cassandra cluster. Performed many complex system analysis in order to improve ETL performance, identified high critical batch jobs to prioritize
- ✓ Implemented Spark solution to enable real time reports from Cassandra data. Was also actively involved in designing column families for various Cassandra Clusters.

**ENVIRONMENT:** RHEL, Ubuntu, Cloudera Manager, Cloudera Search, CDH4, HDFS, HBase, Hive, Pig, ZooKeeper, Monitoring Cluster with automated scripts, Map Reduce2 (YARN), PostgreSQL, MySQL, HDFS, Kafka, Zookeeper, Hive, Yarn, Hue, Oozie, QAS and Ganglia.

**Lead Hadoop Administrator**  
**AT&T Services - San Ramon, CA**

**June 2014 to Sep. 2015**

**Description:** AT&T Enterprise Business provides networking, mobility, cloud, security, hosting, application management, unified communications, VPN, VoIP, and conferencing solutions for medium to large businesses. An extensive portfolio of integrated solutions that will help customers adapt to the shifting demands for business. AT&T wants to implement a fully integrated, low-cost software system with the scale and flexibility to support their growing data needs. The system will ingest over 50 million records per day from workflow database, transform it, store it in high volume store, and

export it too different downstream in order to process the data for generating reports to business users on time.

**Roles and Responsibilities:**

- ✓ Hands on Installation and configuration of Hortonworks Data Platform HDP 2.3.4
- ✓ Worked on installing production cluster, commissioning & decommissioning of Data Nodes, Name Node recovery, capacity planning, and slots configuration
- ✓ Worked on Hadoop Administration, responsibilities include software installation, configuration, software upgrades, backup and recovery, cluster setup, cluster performance and monitoring on daily basis, maintaining cluster up and run on healthy.
- ✓ Implemented the security requirements for Hadoop and integrate with Kerberos authentication and authorization infrastructure.
- ✓ Designed, developed and implemented connectivity products that allow efficient exchange of data between the core database engine and the Hadoop ecosystem.
- ✓ Involved in defining job flows using Oozie for scheduling jobs to manage apache Hadoop jobs.
- ✓ Implemented Name Node High Availability on the Hadoop cluster to overcome single point of failure.
- ✓ Worked on YARN capacity scheduler by creating queues to allocate resource guarantee to specific groups.
- ✓ Worked on importing and exporting data from Oracle database into HDFS and HIVE using Sqoop.
- ✓ Monitored and analyzed of the Map Reduce job executions on cluster at task level.
- ✓ Extensively involved in Cluster Capacity planning, Hardware planning, Performance Tuning of the Hadoop Cluster.
- ✓ Wrote automation scripts and setting up crontab jobs to maintain cluster stability and healthy.
- ✓ Installed Ambari on an already existing Hadoop cluster.
- ✓ Implemented Rack Awareness for data locality optimization.
- ✓ Optimized and tuned the Hadoop environments to meet performance requirements.
- ✓ Hand-On experience with AWS cloud with EC2, S3.
- ✓ Collaborating with offshore team.
- ✓ Ability to document existing processes and recommend improvements.
- ✓ Shares knowledge and assists another team member as needed.
- ✓ Assist with maintenance and troubleshooting of scheduled processes.
- ✓ Participated in development of system test plans and acceptance criteria.
- ✓ Collaborate with offshore developers in order to monitor ETL jobs and troubleshoot steps.

**Environment:** Hortonworks HDP2.3x, Ambari, Oozie 4.2, Sqoop1.4.6, Hive 1.2, Mapreduce2, Ambari, blueprints, Oracle Sql Developer, Teradata, SVN, SFTP, SSH, Eclipse, Jdk 1.7, Maven.

**Sr.Hadoop Admin**  
**2014**

**UMBBank Los Angeles CA**

**Aug. 2013 to May**

### **Roles and Responsibilities:**

- ✓ Installed/Configured/Maintained Apache Hadoop clusters for application development and Hadooptools like Hive, Pig, HBase, Zookeeper and Sqoop.
- ✓ Installed and configured Hadoop, MapReduce, HDFS (Hadoop Distributed File System), developed multiple MapReduce jobs in java for data cleaning.
- ✓ Developed data pipeline using Flume, Sqoop, Pig and Java map reduce to ingest customer behavioral data and financial histories into HDFS for analysis.
- ✓ Involved in collecting and aggregating large amounts of log data using Apache Flume and staging data in HDFS for further analysis.
- ✓ Experience in methodologies such as Agile, Scrum and Test driven development
- ✓ Worked on installing cluster, commissioning & decommissioning of DataNodes, NameNode recovery, capacity planning, Cassandra and slots configuration.
- ✓ Responsible for developing data pipeline using HDInsight, flume, Sqoop and pig to extract the data from weblogs and store in HDFS.
- ✓ Involved in migration of ETL processes from Oracle to Hive to test the easy data manipulation.
- ✓ Worked on importing and exporting data from Oracle and DB2 into HDFS and HIVE using Sqoop.
- ✓ Installed, configured, and administered a small Hadoop clusters consisting of 10 nodes. Monitored cluster for performance and, networking and data integrity issues.
- ✓ Responsible for troubleshooting issues in the execution of MapReduce jobs by inspecting and reviewing log files.
- ✓ Formulated procedures for installation of Hadoop patches, updates and version upgrades.
- ✓ Working with data delivery teams to setup new Hadoop users. This job includes setting up Linux users, setting up Kerberos principals and testing HDFS, Hive.
- ✓ Load log data into HDFS using Flume, Kafka and performing ETL integrations.
- ✓ Managed 350+ Nodes HDP 2.2.4 cluster with 4 petabytes of data using Ambari 2.0 and Linux Cent OS 6.5.
- ✓ Created 25+ Linux Bash scripts for users, groups, data distribution, capacity planning, and system monitoring.
- ✓ Install OS and administrated Hadoop stack with CDH5 (with YARN) Cloudera Distribution including configuration management, monitoring, debugging, and performance tuning.
- ✓ Upgraded the Hadoop cluster from CDH4.7 to CDH5.2.
- ✓ Supported MapReduce Programs and distributed applications running on the Hadoop cluster.
- ✓ Scripting Hadoop package installation and configuration to support fully-automated deployments.
- ✓ Worked with systems engineering team to plan and deploy new Hadoop environments and expand existing Hadoop clusters.

**Environment:** Hive, Pig, HBase, Zookeeper and Sqoop, ETL, Ambari 2.0, Linux Cent OS, HBase, MongoDB, Cassandra, Ganglia and Cloudera Manager.

**Hadoop Administrator**

**Sep. 2010 to Oct. 2012**

## **App Labs - Hyderabad, Andhra Pradesh**

**Description:** Nationwide Mutual Insurance Company & Affiliated Companies is a group of large U.S. insurance and financial services companies. It offers a full range of insurance and financial services across the country, including car, motorcycle, boat, homeowners, pet, farm, life and commercial insurance.

### **Roles and Responsibilities:**

- ✓ Responsible for Cluster configuration maintenance and troubleshooting and tuning the cluster.
- ✓ Good experience on cluster audit findings and tuning configuration parameters.
- ✓ Implemented Kerberos security in all environments.
- ✓ Solid Understanding of Hadoop HDFS, Map-Reduce and other Eco-System Projects.
- ✓ Install and configure Phoenix on HDP 2.1. Create views over HBase table and used SQL queries to retrieve alerts and Meta data.
- ✓ Involved in creating Hive tables loading with data and writing hive queries which will run internally in map reduce way.
- ✓ Implemented Capacity Scheduler to share the resources of the cluster for the map reduce jobs given by the users.
- ✓ Experience in configuring the cluster using FIFO or FAIR share.
- ✓ Demonstrate and understanding of concepts, best practices and functions to implement a Big Data solution in a corporate environment.
- ✓ Inputs to development regarding the efficient utilization of resources like memory and CPU utilization based on the running statistics of Map and Reduce tasks.
- ✓ Worked on High Availability for Name Node using Cloudera Manager to avoid single point of failure.
- ✓ Develop Session tasks and Workflows using Power Center Workflow Manager.
- ✓ Manage and review data backups and log files.
- ✓ Used Ganglia to monitor the cluster around the clock.
- ✓ Set up automated processes to analyze the System and Hadoop log files for predefined errors and send alerts to appropriate groups.
- ✓ Commissioning and Decommissioning Nodes from time to time.
- ✓ Set up and manage HA name node and Name node federation using Apache 2.0 to avoid single point of failures in large clusters
- ✓ Work with network and Linux system engineers to define optimum network configurations, server hardware and operating system.
- ✓ Working with Cloudera Support Team to Fine Tune Cluster.
- ✓ Production support responsibilities include cluster maintenance.

**Environment:** RHEL 4.x/5/6, Solaris 9, 10&11, HP-UX, Centos, SUSE 10, 11, VERITAS Volume Manager 3.x/ 4.x, VERITAS Storage Foundation 5, RedHat Cluster, VERITAS Cluster Server 4.1, Tripwire, NFS, DNS, SAN/NAS, puppet, chef, Splunk.

**Linux Admin**

**May 2006 to July 2010**

**SYNTEL - Pune, Maharashtra**

**Roles and Responsibilities:**

- ✓ Provisioning, building and support of Linux servers both Physical and Virtual using VMware for Production, QA and Developers environment.
- ✓ Installed, configured and Administrated of all UNIX/LINUX servers, includes the design and selection of relevant hardware to support the installation/upgrades of Red Hat Cent OS, Ubuntu operating systems.
- ✓ Network traffic control, IPsec, Quos, VLAN, Proxy, Radius integration on Cisco Hardware via Red Hat Linux Software.
- ✓ Responsible for managing the Chef Client nodes and upload the cookbooks to chef-server from Workstation.
- ✓ Used Agile/scrum Environment and used Jenkins, GitHub for Continuous Integration and Deployment.
- ✓ Responsible for configuring real time backup of web servers. Log file was managed for troubleshooting and probable errors.
- ✓ Responsible for reviewing all open tickets, resolve and close any existing tickets.
- ✓ Document solutions for any issues that have not been discovered previously.
- ✓ Worked with File System includes UNIX file System and Network file system. Planning, scheduling and implementation of O/s. Worked with File System includes UNIX file System and Network file system. Planning, scheduling and implementation of O/s. patches on both Solaris & Linux. Diligently teaming with the infrastructure, network, database, application and business intelligence teams to guarantee high data quality and availability.
- ✓ Highly experienced in optimizing performance of Web Sphere Application server using Workload Management (WLM).
- ✓ Patch management of servers and maintaining server's environment in Development/QA/Staging /Production.
- ✓ Performing Linux systems administration on production and development servers (RedHat Linux, Cent OS and other UNIX utilities).
- ✓ Installing Patches and packages on Unix/Linux Servers.
- ✓ Installation, Configuration, upgradation and administration of Sun Solaris, RedHat Linux.
- ✓ Installation and Configuration of VMware vSphere client, Virtual Server creation and resource allocation.
- ✓ Performance Tuning, Client/Server Connectivity and Database Consistency Checks using different Utilities.
- ✓ Shell scripting for Linux/Unix Systems Administration and related tasks. Point of Contact for Vendor escalation
- ✓

- ✓ Environment: Linux/Centos 4, 5, 6, Logical Volume Manager, VMware ESX, Apache and Tomcat Web Server HPSM, HPSA. Installed Cent OS using Pre-Execution environment boot and Kick-start method on multiple servers.
- ✓ Monitoring System Metrics and logs for any problems.
- ✓ Running Cron-tab to back up Data.
- ✓ Applied Operating System updates, patches and configuration changes.
- ✓ Maintaining the MySQL server and Authentication to required users for Databases.
- ✓ Appropriately documented various Administrative & technical issues

**Environment:** Red hat Linux/Centos 4, 5, 6, Logical Volume Manager, VMware ESX 5.1/5.5, Apache and Tomcat Web Server, Oracle 11,12, Oracle Rac 12c, HPSM, HPSA.



# Puneet Mishra

- Irving, TX, USA

## Contact Information

- pmttech6633@gmail.com (Preferred)
- 2154887750 (Preferred)

## Work History

### Total Work Experience: 13 years

- **Lead Hadoop Admin | Intuit**  
Nov 01, 2015 - No End Date | Irving TX United States
- **Lead Hadoop Administrator | AT&T Services**  
Jun 01, 2014 - Sep 01, 2015 | San Ramon CA United States
- **Sr.Hadoop Admin | UMBBank**  
Aug 01, 2013 - May 01, 2014 | Los Angeles CA United States
- **Hadoop Administrator | Insurance Company**  
Sep 01, 2010 - Oct 01, 2012
- **Linux Admin | SYNTEL**  
May 01, 2006 - Jul 01, 2010

## Skills

- **software** | 10yrs | 2017
- **database** | 8yrs | 2017
- **infrastructure** | 8yrs | 2017
- **performance tuning** | 8yrs | 2017
- **configuration** | 8yrs | 2017
- **installation** | 8yrs | 2017

- **monitoring** | 8yrs | 2017
- **production** | 8yrs | 2017
- **upgrades** | 8yrs | 2017
- **integration** | 7yrs | 2017
- **mysql** | 6yrs | 2017
- **ubuntu** | 6yrs | 2017
- **apache hadoop** | 6yrs | 2017
- **chef** | 6yrs | 2017
- **network** | 6yrs | 2017
- **real-time** | 6yrs | 2017
- **web server** | 6yrs | 2017
- **oracle** | 6yrs | 2015
- **troubleshooting** | 6yrs | 2015
- **backup** | 5yrs | 2015
- **apache**
- **hadoop**

## Work Preferences

- Likely to Switch: Most Likely
- Willing to Relocate: No
- Work Authorization:
  - US
- Work Documents:
  - Green Card
- Desired Hourly Rate: 80+ (USD)
- Desired Salary: 160000+ (USD)
- Security Clearance: No
- Third Party: Yes
- Employment Type:
  - Contract - Corp-to-Corp
  - Contract - W2
  - Contract to Hire - Independent
  - Contract to Hire - W2
  - Part-time
  - Contract - Independent
  - Contract to Hire - Corp-to-Corp

## Profile Sources

- Dice:

<https://www.dice.com/employer/talent/profile/cb18ea5b7ae5fa8d58042c429e20b5bfe241276c>