Bhupal

Summary

Over 9 years of diversified experience in Software Design & Development. Experience as Hadoop developer solving business use cases for several clients. Experience in the field of software with expertise in backend applications

Education

Master of Business Administration, Wilkes University, Wilkes-Barre, PA

Bachelor of Information Technology, Punjab Technical University

Skills:

|  |  |  |
| --- | --- | --- |
| Scripting Language | : | Java, VBScript, .NET |
| Hadoop Ecosystem | : | HDFS, Map Reduce, Pig, Hive, Sqoop, Flume, Zookeeper, HBase |
| ETL Tools | : | Informatica Power Center 6.1/ 7.1/ 9.1 |
| Operating Systems | : | MS-DOS, Windows95/98/NT/XP/7, Linux, Unix |
| Web Technologies | : | JSP, JDBC, CSS |
| Databases | : | Oracle, My SQL |
| Application /Web Server | : | Apache Tomcat 4.0, Web Logic, TFS |
| Functional Testing Tools | : | Quick Test Pro, Selenium, Load Runner, Quality Center, HPALM, JIRA |

Professional Experience:

**Samsung, Ridgefield, NJ Jan’16 – Till Date**

**Role: Hadoop Developer/Administrator**

**Environment: Windows 7/Linux, Hadoop 2.0, Yarn, SharePoint 2014, Hadoop, S3, Hive, Pig, Map Reduce, Impala, Sqoop, Zookeeper, Kafka, HBASE, Putty, MySQL, Cloudera, Agile**

Roles & Responsibility:

* Setting up multi cluster environment running on CDH 5.4.
* Worked with highly unstructured and semi structured data of 2048 TB in size.
* Setting up both prod and dev nodes. 62 total nodes. HDFS Storage has 52 nodes including 12 disk /node having 3.7 TB memory.
* Involved in data loading from external sources with Impala queries to target tables
* Setting Monitoring, alerts and events on the CDH IN batch/shell scripting
* Involved in setting and monitoring all services running on cluster like HBASE, Flume, impala, hive, pig, kafka.
* Involved in importing data from RDBMS with Sqoop jobs. Troubleshoot if any of the jobs fails to finish.
* Involved in giving access to Cloudera manager, Hue, oozie to internal and external users by creating SR request
* Granting and revoking users privileges to the both dev and prod clusters
* Everyday running health tests on the services running on Cloudera
* Download data from FTP with shell scripts to the local clusters for data loading to target tables involving portioning and bucketing techniques.
* Setting up jobs in Crontab as well as creating oozie workflows, coordinations.
* Monitoring all crontab jobs as well as oozie jobs, Debug the issues when any of the jobs fails to complete.
* Sending daily status of the services running on the cluster to the scrum master.
* Involved in creating documentation regarding all jobs running on cluster

**Stonegate Mortgage, Indianapolis, IN Mar’15 – Till Date**

**Role: Big Data/Hadoop** Developer

**Environment: Windows 7/Linux, SharePoint 2014, hadoop 2.0, Hive, Pig, Map Reduce, Sqoop, Zookeeper, TFS, VS 2015, Putty, MySQL, Cloudera, Agile, Teradata**

Roles & Responsibility:

* Worked on a live **60 nodes Hadoop cluster** running**CDH5.2**
* Worked with highly **unstructured and semi structured data of 90 TB** in size (270 GB with replication factor of 3)
* Extracted the data from Teradata/RDMS into HDFS using **Sqoop** (Version 1.4.6)
* Created and worked **Sqoop**(version 1.4.6)**jobs with incremental load** to populate Hive External tables.
* Extensive experience in writing **Pig**(version 0.15)**scripts to transform raw data** from several big data sources into forming baseline big data.
* Developed **Hive** (version 1.2.1) scripts for end user / analyst requirements to perform ad hoc analysis
* Very good understanding of **Partitions, Bucketing**concepts in Hive and designed both**Managed and External** tables in Hive to optimize performance
* **Solved performance issues** in Hive and Pig scripts with understanding of Joins, Group and aggregation and how does it translate to MapReduce jobs.
* Strong Knowledge on Multi Clustered environment and setting up Cloudera Hadoop Eco-System. Experience in installation, configuration and management of Hadoop Clusters. .
* Provided design recommendations and thought leadership to sponsors/stakeholders that improved review processes and resolved technical problems.
* Worked on managing big data/ hadoop logs
* Worked in Linux based environment
* Worked on BI reporting tool Talend for generating reports.
* Integrate Talend with Hadoop for processing big data jobs.
* Good knowledge of Solr, Kafka, Spark.
* Shared responsibility for administration of Hadoop, Hive and Pig.
* Installed and configured Storm, Solr, Flume, Sqoop, Pig, Hive, HBase on Hadoop clusters.
* Provide detailed reporting of work as required by project status reports

**California ISO, Folsom, CA Feb’14- Jan’ 2015**

**Role: Big Data/ Hadoop** Developer

**Environment: Windows 7/Linux/Unix, SharePoint 2013, Hadoop 2.0, Eclipse, Hive, Pig, Map Reduce, Sqoop, Hbase, Zookeeper, HPALM, Putty, Oracle/Teradata, Cloudera, Agile**

**Roles & Responsibility:**

* Design and Implementation of ETL process in Big Data/ Hadoop Eco-systems.
* Hands on experience on Cloudera and migrating big data from Oracle with Sqoop (Version 1.4.3).
* Very good experience with both MapReduce 1 (Job Tracker/Task Tracker) and MapReduce 2 (YARN)
* Worked and import data from Teradata/oracle with SQOOP (version 1.4.3)
* Implementation of de-duplication process to avoid duplicates in daily load.
* Developed several advanced Map Reduce/ Python programs in JAVA as part of functional requirements for Big Data.
* Developed Hive (Version 1.1.1) scripts as part of functional requirements as well as hadoop security with Kerberos.
* Worked with the admin team in designing and upgrading CDH 3 to CDH 4
* Developed **UDFs** in Java as and when necessary to use in PIG and HIVE queries
* Experience in using **Sequence files, RCFile, AVRO file formats**
* Developed **Oozie** workflow for scheduling and orchestrating the ETL process
* Process and analyze ETL/ big data jobs in Talend
* Successfully tested and generate reports in Talend.
* Successfully integrated Hive tables with MySQL database.
* Working in UNIX based environment for data operations
* Experience working on NoSQL databases including Hbase.
* Experience in deployment of code changes using team city build.
* Involved in handling code fixes during production release.
* Managed Hadoop clusters include adding and removing cluster nodes for maintenance and capacity needs.
* Provide detailed reporting of work as required by project status reports.

**Ann Taylor, NYC Feb’ 13 - Jan’14**

**Role: Big Data/ Hadoop** Developer

**Environment: Windows 7, SharePoint 2013, Hadoop 1.0, Eclipse, Pig, Hive, Flume, Sqoop, HBase, Putty, HPALM, WinSCP, Agile, MySQL**

**Roles & Responsibility:**

* Worked on Solr to search/analyze real time big data.
* Used Sqoop (version 1.4.3) to import data into HDFS and Hive from MySQL/ Oracle
* Responsible for building scalable distributed data solutions using Hadoop
* Extensive experience in writing Pig scripts to transform raw data from several data sources in to forming baseline data.
* Developed several advanced Map Reduce programs to process data files received.
* Developed Pig Scripts, Pig UDFs and Hive Scripts, Hive UDFs to load big data files into Hadoop.
* Created workflows and scheduled jobs in Apache oozie.
* Continuous monitoring and managing the Hadoop cluster through Cloudera Manager.
* Extracted feeds form social media sites such as Twitter using Flume and Solr.
* Developed Hive scripts for end user / analyst requirements for adhoc analysis
* Involved in loading data from UNIX file system to HDFS.
* Involved in gathering business requirements and prepared detailed specifications that follow project guidelines required to develop written programs.
* Usage of Sqoop (Version 1.4.3) to import data into HDFS from MySQL database and vice-versa.
* Bulk loaded data into HBase using NOSQL.
* Experience in storing and retrieval of documents in Apache Tomcat
* Very good experience in monitoring and managing the Hadoop cluster using **Cloudera Manager**.
* Good working knowledge of **Cassandra**
* Knowledge transfers sessions on the developed applications to colleagues.

General Motors, Warren, MI **April’ 11 – Dec ‘12**

Role: ETL/ Informatica Analyst

Environment: Windows XP/NT, Informatics 6.2, Unix, Oracle, SQL, PL/SQL

Roles & Responsibility:

* Participated in documenting the existing operational systems.
* Involved in the requirements gathering for the warehouse. Presented the requirements and a design document to the client.
* Created ETL jobs to load data from staging area into data warehouse.
* Analyzed the requirements and framed the business logic for the ETL process.
* Involved in the ETL design and its documentation
* Experience in manufacturing Unit and their activities like Planning, Purchase and Sale activities.
* Involved in creation of use cases for purchase and sale.
* Designed and developed complex aggregate, join, lookup transformation rules (business rules) to generate consolidated (fact/summary) data using Informatica Power center 6.0.
* Designed and developed mappings using Source qualifier, Aggregator, Joiner, Lookup, Sequence generator, stored procedure, Expression, Filter and Rank transformations
* Development of pre-session, post-session routines and batch execution routines using Informatica Server to run Informatica sessions
* Evaluated the level of granularity
* Evaluated slowly changing dimension tables and its impact to the overall Data Warehouse including changes to Source-Target mapping, transformation process, database, etc.
* Collect and link metadata from diverse sources, including relational databases Oracle, XML and flat files.
* Created, optimized, reviewed, and executed Teradata SQL test queries to validate transformation rules used in source to target mappings/source views, and to verify data in target tables
* Extensive experience with PL/SQL in designing, developing functions, procedures, triggers and packages.
* Developed Informatica mappings, re-usable Sessions and Mapplets for data load to data warehouse.
* Designed and developed Informatica mappings and workflows; Identify and Remove Bottlenecks in order to improve the performance of mappings and workflows and used Debugger to test the mappings and fix the bugs

MacMillan Higher Education, NYC **Aug’09- Feb’11**

Role: ETL/ Informatica Analyst

Environment: Informatica 6.1, PL/SQL, MS Access, Oracle, Windows, Unix

Roles & Responsibilities:

* Extensively worked with the data modelers to implement logical and physical data modeling to create an enterprise level data warehousing.
* Created and Modified T-SQL stored procedures for data retrieval from MS SQL SERVER database.
* Automated mappings to run using UNIX shell scripts, which included Pre and Post-session jobs and extracted data from Transaction System into Staging Area.
* Extensively used Informatica Power Center 6.1/6.2 to extract data from various sources and load in to staging database.
* Extensively worked with Informatica Tools - Source Analyzer, Warehouse Designer, Transformation Developer, Mapplet Designer, Mapping Designer, Repository manager, Workflow Manager, Workflow Monitor, Repository server and Informatica server to load data from flat files, legacy data.
* Created mappings using the transformations like Source qualifier, Aggregator, Expression, Lookup, Router, Filter, Rank, Sequence Generator, Update Strategy, Joiner and stored procedure transformations.
* Designed the mappings between sources (external files and databases) to operational staging targets.
* Involved in data cleansing, mapping transformations and loading activities.
* Developed Informatica mappings and also tuned them for Optimum performance, Dependencies and Batch Design.
* Involved in the process design documentation of the Data Warehouse Dimensional Upgrades. Extensively used Informatica for loading the historical data from various tables for different departments.

ICICI Lombard Insurance, India **Aug’06 to June’09**

Role: ETL Informatica Analyst

Project: Enterprise Web Application Full life cycle Implementation

Environment: ETL, Oracle8i,Windows 95/NT, Oracle, Tomcat, UNIX, XML

Roles &Responsibilities:

* Extensively worked on Informatica tools such as Designer (Source Analyzer, Warehouse Designer, Mapping Designer, Transformations), Workflow Manager, and Workflow Monitor
* Developing mapping using needed transformations in tool according to technical specification
* Resolving any performance issue with transformations in Informatica and mapping with the help of technical specifications
* Created and Run the UNIX scripts for all Pre-Post session ETL jobs.
* Participated in Review of Test Plan, Test Cases and Test Scripts prepared by system integration testing team.
* Developed number of Complex Informatica Mappings and Reusable Transformations
* Extensively used various transformations like Aggregator, Expression, connected and unconnected Lookup's and update strategy transformations to load data into target.   
  • Experience in debugging and performance tuning of targets, sources, mappings and sessions.