

Contact

www.linkedin.com/in/venkata-mallepudi-b71b2493 (LinkedIn)

Top Skills

Hadoop

Hive

SQL

Languages

English

Telugu

Hindi

Certifications

Developing Data Products

Hadoop Fundamentals

Hortonworks Certified Apache
Hadoop 2.0 Developer

Introduction to Pig

Machine Learning Foundations: A
Case Study Approach

Publications

An integrated platform for collecting
mobile phone data and learning
demographic features

Venkata Mallepudi

Data Analyst / Data Scientist

Aliso Viejo, California

Summary

- Certified Cloudera and Hortonworks Hadoop Developer.
- Developed, released and presently maintaining an android mobile application using Android Studio, SQLite, Mysql, and Google Firebase and performing analytics on the data collected from it using R.
- 4+ years of experience in analyzing data using Apache Hive.
- 3+ years of experience on Cloudera platform for data management and analytics.
- 1+ years of experience working on Tableau for Visualization.
- Good knowledge on data management on different storage platforms like S3, Google storage.
- Perform descriptive and exploratory analysis using R and Tableau.
- Experience developing predictive models in R.
- Experience using Sqoop and Flume to load data into HDFS.
- Strong understanding of Hadoop architecture and its components: HDFS and MapReduce, Yarn, Spark.
- Knowledge on Pig, Hive, HBase, Sqoop, Flume, HCatalog, Oozie and Zookeeper.
- Experience installing CDH and HDP stacks.
- Knowledge on setting, configuring, monitoring, and administering Hadoop and other sub-projects like Pig, Hive and Hbase.
- Knowledge on troubleshooting and supporting Linux Operating Systems.
- Worked on Iterative Graph Processing system – Apache Giraph that runs on Hadoop framework and uses map-phase of MapReduce programming model.
- Good knowledge on automating processes using PHP.
- Around 3 years of experience in SQL Server 2005/2008 and Oracle 10g.
- Extensive working experience in database design and development, creating complex database queries, writing Stored Procedures and Functions.
- Worked on Oracle Transportation Management (OTM) as a Techno-Functional Expert.

- Good understanding of OOP concepts.
- Experience in all phases of System Development Life cycle (SDLC).
- Ability to analyze and perform root cause analysis for complex problems.

Experience

M2Catalyst, LLC

Data Analyst/Data Engineer/Data Scientist

July 2015 - Present (4 years 9 months)

Aliso Viejo, California

- Develop predictive models to determine the location of the cell-towers using crowd-sourced data.
- Perform descriptive and exploratory analysis on the data to identify different scenarios and frame proper questions to develop predictive models for each of those different scenarios.
- Perform statistical inference and hypothesis testing and draw conclusions with strong data analysis.
- Work on massive amounts of data and performing large-scale data analysis to derive useful business insights to customer about market behavior.
- Design and maintain analytical database structures
- Perform exploratory and statistical analysis over issues to identify the cause.
- Perform data management tasks like Storage, compression, backup and recovery.
- Perform data analysis using Apache Hive, R, Microsoft Excel and Python.
- Perform data visualization using R, Microsoft Excel, Tableau.
- Perform feature-engineering to generate more accurate results with regression and other machine learning models.
- Automating the execution process of Apache Hive queries using PHP.
- Cluster tuning, monitoring and maintenance.
- Data exports to clients in different file formats like CSV, TSV and JSON.
- Data exports to different storage services like Amazon S3 buckets, Synology, Google Cloud.
- Experience building dashboards in Tableau with multiple external data sources.
- Communicate and present the data to both internal and external customers conveying the complete story around it.

- Validate implementation across clients, through data analysis, which support client's business objectives
- Experience using Command line interface to run php scripts, and hive queries.
- Familiarity with Issue-Tracking Tools.
- Experience with git.

IGATE

Technical Lead

February 2015 - July 2015 (6 months)

New Jersey

- Lead a team of 4 testers in understanding the requirements and identifying the possible use-case scenarios to test.
- Prepare weekly schedules for each tester and assign test cases that needs to be executed.
- Prepare test cases and execute them on time.
- Update the respective developer/development-team about the failed test cases and work closely with them to get the issues resolved.
- Generate weekly reports on the status of the test cases and circulate to the whole team involved.
- Train new testers on writing test-cases and updating the status of each test-case after execution.

Louisiana State University

1 year 10 months

System Administrator (Graduate Assistant)

August 2012 - May 2014 (1 year 10 months)

- Responsible for maintaining SQL server data bases.
- Designing SSIS packages for migration of data from different sources like Excel , Access to SQL Sever using SSIS/2008.
- Responsible for identifying, troubleshooting and resolving hardware, software and network related issues on systems of end users.
- Configure and install Windows/Macintosh operating systems on new computers.
- Track and fix trouble tickets/requests, monitor progress and communicate with manager about unresolved problems.
- Work with vendors to resolve hardware related issues. Analyze bugs in system and application software.
- Optimize computer usage by recommending enhancements to system capabilities and performance.

Masters Project: Implementing Contrail's Bubbles phase in Apache Giraph

May 2013 - March 2014 (11 months)

I worked on developing a genome assembler. Genome assembler assembles the data provided by the DNA sequencers. The data is a fixed length strings of nucleotides which are the building blocks of DNA structure. The size of the dataset varies from a few Megabytes to several Terabytes depending upon the species.

- Installed, configured and troubleshooted Apache Hadoop and Apache Giraph on cluster.
- Worked on an error-correcting phase called "Bubbles Phase".
- Implemented bubbles-phase of DNA assembler using map-reduce programming model in JAVA.
- Implemented the phase in both Old-MapReduce API and New-MapReduce API.
- Implemented bubbles phase on Apache Giraph which has only map-phase and no reduce-phase.
- Monitored the performance of each datanode while running the job.
- Identified and solved cluster and Hadoop related issues while working on the cluster.
- Tested the code with 2 Million short-reads as input (2.5GB – dataset).
- Tried handling different types of datasets like Sequence files and RCFiles.

Front End: Hadoop Framework: JAVA.

Mobile Application for LSU College of Engineering

October 2013 - December 2013 (3 months)

A mobile application for the sophomores of LSU is created to easily access the details of available faculty and classrooms. A chat room for students to discuss and clarify doubts is also provided.

- Created a login page to allow only sophomores to access the application.
- Created dynamic pages for the students to post their queries and discuss them.
- Handled database design and created the tables.
- Created SQL queries and stored procedures in SQL Server 2010.

Front End: Xamarin Studio, C#.

Back End: SQL Server 2010.

Web Application for Chemical Engineering Department in LSU

May 2013 - August 2013 (4 months)

A web application for the Chemical Engineering department is created using HTML, CSS, JavaScript and JQuery. It provides all the information about the department and people should login for the special events and discussions.

- Created a web application using ASP.NET and handled the database using SQL Server 2010.
- Created login page, and registration page from the scratch.
- Embedded audio and video features within the webpage.
- Provided upload and download options of documents to specific users of the application.

Front End: ASP.NET, C#.

Back End: SQL Server 2010

Southern Marsh Company

October 2012 - December 2012 (3 months)

Worked on an online application called Shopify to create an online store to Southern Marsh Company especially for user sign up, Generation of coupon codes and updating of database during online sales.

- Developed a web application to store and retrieve information relating to the web pages requested by the user.
- Involved in the analysis, design, coding and testing of the application.
- Handled database design and maintenance for the online store.
- Created SQL queries and stored procedures in SQL Server 2010.

Front End: ASP .NET. C#.

Back End: SQL Server 2010.

Tech Mahindra (formerly Mahindra Satyam)

1 year 10 months

Developer

March 2012 - July 2012 (5 months)

E2 open is an enhancement project for generating reports in a regular basis to the customer and client.

- Worked extensively on SQL queries to generate data in reports.
- Customized reports as per the clients requirement.
- Developed reports using Fusion Transportation Management (module in Oracle Transportation Management).

Developer

October 2010 - March 2012 (1 year 6 months)

Worked as a Developer on Oracle Transportation Systems (OTM). This serves to provide DHL an interface to perform transportation with minimal cost and itinerary for its customers.

Configuring scenarios based on the geography, creating users and assigning roles are my primary tasks.

- Involved in the analysis, design, coding and testing of the application.
- Created SQL queries and stored procedures.
- Worked on bulk import/export to database from Microsoft Excel.
- Worked on ASP.NET to provide customized web pages as per the requirement.
- Handled Login Page and Tax Calculations screens using ASP.NET.
- Additional features like downloading details after Tax calculations on to a separate Excel file with just a single click have been provided.
- Provided XML mappings for the fields required in integration layer (Service Oriented Architecture)

Sanketika Vidya Parishad

Web Cache Optimization using Least Grade Page Replacement Algorithm

September 2009 - February 2010 (6 months)

- Developed a web application to store and retrieve information relating to the web pages requested.
- Involved in the analysis, design, coding and testing of the application.
- Web cache is maintained in the local proxy to store information about the server, date & time of the web pages requested.
- Developed backend database using Oracle 10g.
- Automated updating of the web page information in the cache is provided.
- Front End: JAVA
- Back End: Oracle 10g.

Sanketika Vidya Parishad

Complaint tracking System

August 2008 - December 2008 (5 months)

Employees can login and raise complaints and can check the respective status of the complaint regularly.

- Developed a window form application on Complaint Tracking System.
- Involved in the analysis, design, coding and testing of the application.
- Worked extensively in maintaining the database consistency.
- Worked on SQL queries in updating the statuses of the complaints.

Front End: JAVA.

Back End: SQL Server 2008.

Education

Louisiana State University

Master's degree, Computer Science · (2012 - 2014)

Andhra University

Bachelor's degree, Computer Science · (2006 - 2010)