

Data Scientist Data Scientist Data Scientist - BBVA Compass Birmingham, AL Over 8+ years of Experience in Data Architecture, Design, Development, and Testing of business application systems, Data Analysis and developing Conceptual, logical models and physical database design for Online Transactional processing (OLTP) and Online Analytical Processing (OLAP) systems. * Experienced in designing star schema, Snowflake schema for Data Warehouse, and ODS architecture. * Experienced in Data Modeling & Data Analysis experience using Dimensional Data Modeling and Relational Data Modeling, Star Schema/Snowflake Modeling, FACT & Dimensions tables, Physical & Logical Data Modeling. * Experienced in big data analysis and developing data models using Hive, PIG, and Map reduce, SQL with strong data architecting skills designing data-centric solutions. * Hands on experience with big data tools like Hadoop, Spark, Hive, Pig, Impala, PySpark, SparkSQL.

* Very good knowledge and experience on AWS, Redshift, S3, and EMR. * Proficient in Data Analysis, mapping source and target systems for data migration efforts and resolving issues relating to data migration. * Excellent development experience SQL, Procedural Language (PL) of databases like Oracle, Teradata, Netezza, and DB2. * Very good knowledge and working experience on big data tools like Hadoop, Azure Data Lake, AWS Redshift. * Experienced in Data Scrubbing/Cleansing, Data Quality, Data Mapping, Data Profiling, Data Validation in ETL * Experienced in creating and documenting Metadata for OLTP and OLAP when designing systems.

* Expertise in synthesizing Machine learning, Predictive Analytics and Big data technologies into integrated solutions. * Extensive experience in development of T-SQL, DTS, OLAP, PL/SQL, Stored Procedures, Triggers, Functions, Packages, performance tuning and optimization for business logic implementation. * Experience in using various packages in R and Python like ggplot2, caret, dplyr, Rweka, gmodels, RCurl, tm, C50, twitter, NLP, Reshape2, rjson, dplyr, pandas, NumPy, Seaborn, SciPy, matplotlib, Scikit-learn, BeautifulSoup, Rpy2. * Experienced using query tools like SQL Developer, PL/SQL Developer, and Teradata SQL Assistant. * Excellent knowledge of Machine Learning, Mathematical Modeling and Operations Research. Comfortable with R, Python, SAS and Weka, MATLAB, Relational databases. Deep understanding & exposure of Big Data Eco-system. * Expertise in designing complex Mappings and have expertise in performance tuning

and slowly-changing Dimension Tables and Fact tables * Extensively worked with Teradata utilities BTEQ, Fast export, and Multi-Load to export and load data to/from different source systems including flat files. * Hands on experience in implementing LDA, Naive Bayes and skilled in Random Forests, Decision Trees, Linear and Logistic Regression, SVM, Clustering, neural networks, Principle Component Analysis. * Expertise in extracting, transforming and loading data between homogeneous and heterogeneous systems like SQL Server, Oracle, DB2, MS Access, Excel, Flat File and etc. using SSIS packages. * Proficient in System Analysis, ER/Dimensional Data Modeling, Database design and implementing RDBMS specific features. * Experience in UNIX shell scripting, Perl scripting, and automation of ETL Processes. * Extensively used ETL to load data using Power Center / Power Exchange from source systems like Flat Files and Excel Files into staging tables and load the data into the target database Oracle. Analyzed the existing systems and made a Feasibility Study. * Excellent understanding and working experience of industry standard methodologies like System Development Life Cycle (SDLC), as per Rational Unified Process (RUP), Agile Methodologies. * Proficiency in SQL across a number of dialects (we commonly write MySQL, PostgreSQL, Redshift, SQL Server, and Oracle) * Experienced in developing Entity-Relationship diagrams and modeling Transactional Databases and Data Warehouse using tools like ERWIN, ER/Studio, and Power Designer and experienced with modeling using ERWIN in both forward and reverse engineering cases. Authorized to work in the US for any employer Work Experience Data Scientist BBVA Compass - Birmingham, AL July 2017 to Present Architect framework BBVA Compass Bancshares, Inc - Birmingham, AL 2007 to Present is a bank holding company headquartered in Birmingham, Alabama. It has been a subsidiary of the Spanish multinational Banco Bilbao Vizcaya Argentaria since 2007 and operates chiefly in Alabama, Arizona, California, Colorado, Florida, New Mexico, and Texas. Responsibilities: * Design, Develop and implement Comprehensive Data Warehouse Solution to extract, clean, transfer, load and manage quality/accuracy of data from various sources to EDW Enterprise Data Warehouse. * Architect framework for data warehouse solutions to bring data from source system to EDW and provide data mart solutions for Order/Sales operation, Salesforce activity, Inventory tracking, in-depth data mining

and analysis for market projection etc. * Utilized Apache Spark with Python to develop and execute Big Data Analytics and Machine learning applications, executed machine learning use cases under Spark ML and MLLib. * Handled importing data from various data sources, performed transformations using Hive, Map Reduce, and loaded data into HDFS. * Performed data cleaning and feature selection using MLLib package in PySpark and working with deep learning frameworks such as Caffe, Neon. * Tested Complex ETL Mappings and Sessions based on business user requirements and business rules to load data from source flat files and RDBMS tables to target tables. * Utilized Spark, Scala, Hadoop, HBase, Cassandra, MongoDB, Kafka, Spark Streaming, MLLib, Python, a broad variety of machine learning methods including classifications, regressions, dimensionally reduction etc. and Utilized the engine to increase user lifetime by 45% and triple user conversations for target categories. * Develop a high performance, scalable data architecture solution that incorporates a matrix of technology to relate architectural decision to business needs. * Conducting strategy and architecture sessions and deliver artifacts such as MDM strategy (Current state, Interim State, and Target state) and MDM Architecture (Conceptual, Logical and Physical) at the detail level. * Design and development of dimensional data model on Redshift to provide advanced selection analytics platform and developed Simple to complex Map Reduce Jobs using Hive and Pig. * Designed and developed NLP models for sentiment analysis. * Designed and provisioned the platform architecture to execute Hadoop and machine learning use cases under Cloud infrastructure, AWS, EMR, and S3. * Developed and configured on Informatica MDM hub supports the Master Data Management (MDM), Business Intelligence (BI) and Data Warehousing platforms to meet business needs. * Transforming staging area data into a STAR schema (hosted on Amazon Redshift) which was then used for developing embedded Tableau dashboards * Worked on machine learning on large size data using Spark and MapReduce. * Let the implementation of new statistical algorithms and operators on Hadoop and SQL platforms and utilized optimizations techniques, linear regressions, K-means clustering, NativeBayes and other approaches. * Developed Spark/Scala, Python for regular expression (regex) project in the Hadoop/Hive environment with Linux/Windows for big data resources. * Proficiency in SQL across a

number of dialects (we commonly write MySQL, PostgreSQL, Redshift, Teradata, and Oracle) * Responsible for full data loads from production to AWS Redshift staging environment and Worked on migrating of EDW to AWS using EMR and various other technologies. * Worked on Teradata SQL queries, Teradata Indexes, Utilities such as Mload, TPump, Fast load and Fast Export.

* Application of various machine learning algorithms and statistical modeling like decision trees, regression models, neural networks, SVM, clustering to identify Volume using Scikit-learn package in python, MATLAB. * Worked on data pre-processing and cleaning the data to perform feature engineering and performed data imputation techniques for the missing values in the dataset using Python. * Created Data Quality Scripts using SQL and Hive to validate successful data load and quality of the data. Created various types of data visualizations using Python and Tableau. * Worked with BTEQ to submit SQL statements, import and export data, and generate reports in Teradata. * Build and maintain scalable data pipelines using the Hadoop ecosystem and other open sources components like Hive and HBase. * Created Hive architecture used for real-time monitoring and HBase used for reporting and worked for map reduce and query optimization for Hadoop Hive and HBase architecture. * Involved in Teradata utilities (BTEQ, Fast Load, Fast Export, Multiload, and TPump) in both Windows and Mainframe platforms. * Built analytical data pipelines to port data in and out of Hadoop/HDFS from structured and unstructured sources and designed and implemented system architecture for Amazon EC2 based cloud-hosted solution for the client.

Environment: Erwin 9.6.4, Oracle 12c, Python, PySpark, Spark, Spark MLlib, Tableau, ODS, PL/SQL, OLAP, OLTP, AWS, Hadoop, Map Reduce, HDFS, Python, MDM, Teradata 15, Hadoop, Spark, Cassandra, SAP, MS Excel, Flat files, Tableau, Informatica, SSIS, SSRS, AWS EC2, AWS EMR, Elastic Search. Data Scientist BBVA Compass Bancshares, Inc - Chicago, IL August 2016 to June 2017 Description: At AIM Specialty Health (AIM), it is our mission to promote appropriate, safe, and affordable health care. As the specialty benefits management partner of choice for today's leading healthcare organizations, we help improve the quality of care and reduce costs for today's most complex tests and treatments.

Responsibilities: * Developed applications of Machine Learning, Statistical Analysis, and Data Visualizations with challenging data Processing

problems in sustainability and biomedical domain. * Worked on Natural Language Processing with NLTK module of python for application development for automated customer response. * Used predictive modeling with tools in SAS, SPSS, R, Python. * Responsible for design and development of advanced R/ Python programs to prepare to transform and harmonize data sets in preparation for modeling. * Identifying and executing process improvements, hands-on in various technologies such as Oracle, Informatica, and BusinessObjects. * Handled importing data from various data sources, performed transformations using Hive, Map Reduce, and loaded data into HDFS. * Interaction with Business Analyst, SMEs, and other Data Architects to understand Business needs and functionality for various project solutions. * Created SQL tables with referential integrity and developed queries using SQL, SQL*PLUS, and PL/SQL. * Involved with Data Analysis primarily Identifying Data Sets, Source Data, Source Meta Data, Data Definitions and Data Formats * Wrote simple and advanced SQL queries and scripts to create standard and Adhoc reports for senior managers. * Created PL/SQL packages and Database Triggers and developed user procedures and prepared user manuals for the new programs. * Prepare ETL Architect & design document which covers ETL Architect, SSIS design, Extraction, transformation, and loading of Duck Creek data into the dimensional model. * Design Logical & Physical Data Model using MSVisio 2003 data modeler tool. * Participated in Architect solution meetings & guidance in Dimensional Data Modeling design. * Applied linear regression, multiple regression, ordinary least square method, mean-variance, the theory of large numbers, logistic regression, dummy variable, residuals, Poisson distribution, Bayes, Naive Bayes, fitting function etc to data with help of Scikit, SciPy, NumPy and Pandas module of Python. * Applied clustering algorithms i.e. Hierarchical, K-means with help of Scikit and SciPy. * Developed visualizations and dashboards using ggplot, Tableau * Worked on development of data warehouse, Data Lake and ETL systems using relational and non-relational tools like SQL, No SQL. * Built and analyzed datasets using R, SAS, MATLAB, and Python (in decreasing order of usage). * Applied linear regression in Python and SAS to understand the relationship between different attributes of the dataset and causal relationship between them * Pipelined (ingest/clean/munge/transform) data for feature extraction toward downstream classification. * Used

ClouderaHadoopYARN to perform analytics on data in Hive. * Wrote Hive queries for data analysis to meet the business requirements. * Expertise in BusinessIntelligence and data visualization using R and Tableau. * Validated the Macro-Economic data (e.g. BlackRock, Moody's etc.) and predictive analysis of world markets using key indicators in Python and machine learning concepts like regression, Bootstrap Aggregation and Random Forest. * Worked in large-scale database environments like Hadoop and MapReduce, with working mechanism of Hadoop clusters, nodes and Hadoop Distributed File System (HDFS). * Interfaced with large-scaledatabase system through an ETL server for data extraction and preparation. * Identified patterns, data quality issues, and opportunities and leveraged insights by communicating opportunities with business partners.

Environment:Machine learning, AWS, MS Azure, Cassandra, Spark, HDFS, Hive, Pig, Linux, Python (Scikit-Learn/SciPy/NumPy/Pandas), R, SAS, SPSS, MySQL, Eclipse, PL/SQL, SQL connector, Tableau. Python Developer Walgreens - Chicago, IL March 2015 to July 2016 Description:The Walgreen Company is an American company that operates as the second-largest pharmacy store chain in the United States behind CVS Health. It specializes in filling prescriptions, health and wellness products, health information, and photo services. Responsibilities: * Designed and Developed UI for creating Dashboard application using AngularJS, D3, C3, HTML, CSS, Bootstrap, JavaScript and jQuery. * Developed and implemented Python scripts to automate retrieval, parsing and reporting of configuration parameters from Network Devices connected to customer networks. * Involved in user interface design and development using JSP, Servlet, HTML5, CSS3,andJavaScript. * Wrote and tested Python scripts to create new data files for Linux server configuration using a Python template tool. * Modified controlling databases using SQL generated via Python and Perl code, collected and analyzed data with Python programs using SQL queries from the database of data collected from the systems under tests. * Developed new user interface components for different modules using Kendo UI with various controls including Grid controls, and chart controls etc. * Involved in write application level code to interact with APIs, Web Services using AJAX, JSON and hence building type-ahead feature for zip code, city, and county lookup using jQuery, Ajax, and jQuery UI. * Worked on updating the existing clipboard to have the new

features as per the client requirements. * Skilled in using collections in Python for manipulating and looping through different user-defined objects. * Taken part in the entire lifecycle of the projects including Design, Development, and Deployment, Testing and Implementation and support. * Developed views and templates with Python and Django's view controller and templating language to create a user-friendly website interface. * Automated different workflows, which are initiated manually with Python scripts and Unix shell scripting. * Used Python unit and functional testing modules such as unit test, unittest2, mock, and custom frameworks in-line with Agile Software Development methodologies. * Wrote and executed various MYSQL database queries from Python using Python-MYSQL connector and MySQL dB package. * Created and maintained Technical documentation for launching Hadoop Clusters and for executing Hive queries and Pig Scripts. * Developed Sqoop scripts to handle change data capture for processing incremental records between new arrived and existing data in RDBMS tables. * Installed Hadoop, Map Reduce, HDFS, AWS and developed multiple MapReduce jobs in PIG and Hive for data cleaning and pre-processing. * Managed datasets using Panda data frames and MySQL, queried MYSQL database queries from python using Python-MYSQL connector and MySQL dB package to retrieve information. * Involved in the Web/Application development using Python, HTML5, CSS3, AJAX, JSON, and jQuery. * Developed and tested many features for a dashboard using Python, Java, Bootstrap, CSS, JavaScript, and jQuery. * Generated Python Django forms to record data of online users and used PyTest for writing test cases. * Implemented and modified various SQL queries and Functions, Cursors and Triggers as per the client requirements. * Prototype proposal for Issue Tracker website using Python/Django connecting MySQL as Database. * The developed overall layout that meets cross-platform compatibility using Bootstrap, media queries and Angular UI Bootstrap.

Environment: Python, HTML5, CSS3, AJAX, JSON, jQuery, MySQL, NumPy, SQL Alchemy, Matplotlib, Hadoop, Pig Scripts.

Python Developer Citi Bank - Irving, TX July 2013 to February 2015

Description: The project was to build an algorithm that accurately classifies credit card holders among multiple classes based on the historical data available on multiple variables. Further, the aim was to improve bank's efficiency by reducing default rate while offering new

products. Moreover, I was Involved in a project to identify the employees' access level, based on his/her current & historical tasks and duties..

Responsibilities:

- * Involved in the software development lifecycle (SDLC) of tracking the requirements, gathering, analysis, detailed design, development, system testing and user acceptance testing.
- * Developed entire frontend and backend modules using Python on Django Web Framework.
- * Involved in designing user interactive web pages as the front-end part of the web application using various web technologies like HTML, JavaScript, Angular JS, jQuery, AJAX and implemented CSS for better appearance and feel.
- * Knowledge of cross-browser and cross-platform development of HTML and CSS based websites in Windows like IE 6, IE 7, IE 8 and FF.
- * Interactive in providing change requests, trouble reports and requirements collection with the client.
- * Actively involved in developing the methods for Create, Read, Update and Delete (CRUD) in Active Record.
- * Working knowledge of various AWS technologies like SQS Queuing, SNS Notification, S3 storage, Redshift, Data Pipeline, EMR.
- * Developed a fully automated continuous integration system using Git, Jenkins, MySQL and custom tools developed in Python and Bash.
- * Implemented Multithreading module and complex networking operations like race route, SMTP mail server and web server Using Python.
- * Used NumPy for Numerical analysis for the Insurance premium.
- * Implemented and modified various SQL queries and Functions, Cursors and Triggers as per the client requirements.
- * Managed code versioning with GitHub, BitBucket, and deployment to staging and production servers.
- * Implemented MVC architecture in developing the web application with the help of Django framework.
- * Used Celery as task queue and Rabbit MQ, Redis as messaging broker to execute asynchronous tasks.
- * Designed and managed API system deployment using a fastHTTP server and Amazon AWS architecture.
- * Involved in code reviews using GitHub pull requests, reducing bugs, improving code quality, and increasing knowledge sharing
- * Install and configuring monitoring scripts for AWS EC2 instances.
- * Working under UNIX environment in the development of application using Python and familiar with all its commands.
- * Developed remote integration with third-party platforms by using RESTful web services.
- * Updated and maintained Jenkins for automatic building jobs and deployment.
- * Improved code reuse and performance by making

effective use of various design patterns and refactoring code base. * Updated and maintained Puppet RSpec unit/system test. * Worked on debugging and troubleshooting programming related issues. * Worked in the MySQL database on simple queries and writing Stored Procedures for normalization. * Deployment of the web application using the Linux server. Environment: Python 2.7, Django 1.4, HTML5, CSS, XML, MySQL, JavaScript, Backbone JS, JQuery, Mongo DB, MS SQL Server, JavaScript, Git, GitHub, AWS, Linux, Shell Scripting, AJAX, JAVA. Hadoop Developer SITEL India Pvt LTD - Hyderabad, Telangana August 2012 to June 2013 Description: Sitel Group combines comprehensive customer care capabilities with unparalleled digital, training and technology expertise to help build brand loyalty and improve customer satisfaction. We partner with our clients to effectively harness our industry's explosive digital transformation to ensure an innovative, end-to-end solution to managing and enhancing the customer experience.

Responsibilities: * Designed and developed multiple MapReduce jobs in Java for complex analysis. Importing and exporting the data using Sqoop from HDFS to Relational Database systems and vice-versa. * Integrated Apache Storm with Kafka to perform web analytics. Uploaded clickstream data from Kafka to HDFS, HBase, and Hive by integrating with Storm. * Configured Flume to transport web server logs into HDFS. Also used Kite logging module to upload web server logs into HDFS. * Developed UDF functions for Hive and wrote complex queries in Hive for data analysis * Performed Installation of Hadoop in fully and Pseudo Distributed Mode for POC in early stages of the project. * Analyze, develop, integrate, and then direct the operationalization of new data sources. * Generating Scala and Java classes from the respective APIs so that they can be incorporated into the overall application. * Responsible for working with different teams in building Hadoop Infrastructure * Gathered business requirements in meetings for successful implementation and POC and moving it to Production and implemented POC to migrate map reduce jobs into Spark RDD transformations using Scala. * Implemented different machine learning techniques in Scala using Scala machine learning library. * Developed Spark applications using Scala for easy Hadoop transitions. * Successfully loaded files to Hive and HDFS from Oracle, Netezza and SQL Server using SQOOP * Uses Talend Open Studio to load files into Hadoop HIVE tables and performed

ETL aggregations in Hadoop Hive. * Developed Simple to Quebec and Python MapReduce streaming jobs using Python language that is implemented using Hive and Pig. * Designing & Creating ETL Jobs through Talend to load huge volumes of data into Cassandra, Hadoop Ecosystem, and relational databases. * Worked on analyzing, writing Hadoop MapReduce jobs using Java API, Pig, and Hive. * Developed some machine learning algorithms using Mahout for data mining for the data stored in HDFS * Used Flume extensively in gathering and moving log data files from Application Servers to a central location in Hadoop Distributed File System (HDFS) * Worked with Oozie Workflow manager to schedule Hadoop jobs and high intensive jobs * Responsible for cluster maintenance, adding and removing cluster nodes, cluster monitoring, and troubleshooting, manage and review data backups, manage and review Hadoop log files. * Extensively used Hive/HQL or Hive queries to query data in Hive Tables and loaded data into HIVE tables. * Creating UDF functions in Pig & Hive and applying partitioning and bucketing techniques in Hive for performance improvement * Creating indexes and tuning the SQL queries in Hive and Involved in database connection by using Sqoop. * Involved in Hadoop Name node metadata backups and load balancing as a part of Cluster Maintenance and Monitoring * Used File System Check (FSCK) to check the health of files in HDFS and used Sqoop to import data from SQL server to Cassandra * Monitored Nightly jobs to export data out of HDFS to be stored offsite as part of HDFS backup * Used Pig for analysis of large datasets and brought data back to HBase by Pig * Developed Python Mapper and Reducer scripts and implemented them using Hadoop streaming. * Created schema and database objects in HIVE and developed Unix Scripts for data loading and automation * Involved in the training of big data ecosystem to end-users. Environment: Java1.5, J2EE, Hibernate, Spring, JUnit, WebLogic HTML, CSS, JavaScript, Jenkins, Node.js, jQuery, Linux, CICD, Spring Boot, Maven, Log4J and Junit, Eclipse, REST, SQL Navigator. Java Developer Napier Healthcare Pvt. Ltd - Hyderabad, Telangana October 2009 to July 2011 Description: Napier Healthcare was established in 1996 as a Healthcare IT Products and Services Company. With deep domain knowledge and singular focus, we have built a robust, feature-rich suite of solutions to deliver value to the stakeholders of the healthcare industry worldwide. Responsibilities: *

Development, Testing, Maintenance and product delivery. * Developed a scalable and maintainable application using J2EE Framework, Hibernate, MVC Model, Struts, and J2EE Design Patterns. * Prepared SOW (Statement of Work) by communicating with agencies and organized meetings about requirements. * Followed Java & J2EE design patterns and the coding guidelines to design and develop the application. * Extensively used JSTL tags and Struts tag libraries. Used Struts tiles as well in the presentation tier. * Developing the application using Struts and Spring based frameworks. * Actively involved in designing and implementing the application using various design patterns. * Coordinating with clients and closing production issues relating to software development.

* Identifying and Evaluate Technology Solutions, Problem Solving, and Troubleshooting. * Done with Server-side validations using Struts Validation framework. * Processed JSON response data by consuming RESTful web services and used an Angular filter for implementing search results. * Used Struts-config.xml file for defining Mapping Definitions and Action Forward Definitions * Developed the Action Classes which is in between view and model layers, Action Form Classes which is used to maintain session state of a web application, created JSPs (Java Server pages) using Struts tag libraries and configured in struts-config.XML, web.xml files. * This application is designed using MVC architecture to maintain easily. * Hibernate is used for database connectivity and designed HQL (Hibernate Query language) to create, modify and update the tables. * Created new Soap services using JAX-WS specifications. * Wrote JUnit test cases for testing.

Environment:Java, Struts, Hibernate, JSP, Servlets, SOAP UI, HTML, CSS, Java Script, JUnit, Apache Tomcat Server, EJB, NetBeans Education Bachelor's Skills PYTHON (10+ years), MYSQL (10+ years), ORACLE (10+ years), PL/SQL (10+ years), SQL (10+ years) Additional Information

TECHNICAL SKILLS Python Libraries Beautiful Soup, NumPy, SciPy, Matplotlib, python-twitter, Pandas data frame, urllib2 Data Analytics Tools/Programming Python (NumPy, SciPy, pandas, Gensim, Keras), R (Caret, Weka, ggplot), MATLAB, Microsoft SQL Server, Oracle PL/SQL, Python. Database Oracle11g, MySQL 5.x, and SQLServer Version Control SVN, Clear case, CVS Reporting Tools MS Office (Word/Excel/PowerPoint/ Visio/Outlook), Crystal Reports XI, SSRS, Cognos 7.0/6.0. BI Tools Microsoft Power BI, Tableau, SSIS, SSRS, SSAS, Business Intelligence

Development Studio (BIDS), Visual Studio, Crystal Reports, Informatica 6.1. Database Design Tools and Data Modeling MS Visio, ERWIN 4.5/4.0, Star Schema/Snowflake Schema modeling, Fact & Dimensions tables, physical & logical data modeling, Normalization and De-normalization techniques, Kimball & Inmon Methodologies IDE's PyCharm, Emacs, Eclipse, NetBeans, Sublime, SOAP UI Web/App. Servers WebSphere Application Server 8.0, Apache Tomcat, Web Logic 11g/12c, JBoss 4.x/5.x CloudTechnologies AWS, S3.

Name: Nicole Hudson

Email: thomaskristin@example.net

Phone: 658-459-3839x746