**Indeed.com – Data Scientist Job Vacancies**

Indeed.com-01

# Data Scientist

## Data Scientist - Sykes Enterprises, CT

Sunnyvale, CA

## Work Experience

**Data Scientist**

**Sykes Enterprises, CT**

December 2017 to Present

• Deep Learning: Predicted type of customer service request by implementing deep learning model, on call center telemetric data to reduce an employee's average request handling time when assisting customers   
• Recommender Systems: Built and deployed a recommender system on Microsoft Azure to reduce agent's average handle time of issue and increased the number of calls answered in a quarter by 5%   
• People Analytics: Planned strategic workforce scaling by implementing employee attrition (Microsoft Azure, SQL, and R) predictive model; suggested recommendations to reduce effect of job abandonment by 10%

**Data Science Consultant**

**Limra International Insurance**

July 2017 to December 2017

• Insurance Analytics: identified trends that affect length of service (LOS) of agents and estimated LOS using Linear regression; suggested business proposals that increased LOS   
• Predicting Insurance Agent Churn: Identified the factors impacting agent churn by building logistic regression model and provided insights to develop agent retention programs that decreased churn by 5%   
• Consumer Analytics: Built customer retention prescriptive models for subscription based products using advanced machine learning algorithms. This resulted in decreasing customer churn by approximately 8%

**Data Science Analyst**

**IBM**

January 2015 to December 2016

• Retail orders Analytics: Created classification model (Logistic Regression, SQL, and R) to analyze customer complaints on incorrect orders and performed Root Cause Analysis; reduced incorrect orders by 75%   
• HR Analytics: Predicted the absenteeism of employees by modelling employee behavior data with classification model, assisted the management to reduce the absenteeism   
• Software Engineering: Delivered Financial Solution to fix incorrect monthly bill cycle and correctly billed approximately $1 million towards client and won recognition from Subject Matter Expert   
• Dashboarding: Designed interactive dashboards in Tableau to track Key Performance Indicators (KPIs) that assist in meeting service level agreement; proposed business strategies leading to increasing of team performances   
• Reporting: Extracted data (SQL), prepared insightful reports and identified performance hindering metrics; minimized percent of aged incidents by 60% in 4 months

## Education

**Master in Business Analytics and Project Management**

**University of Connecticut School of Business**

May 2018

## Skills

SQL (2 years), DATA SCIENCE (2 years), CLUSTERING (Less than 1 year), DATA MANIPULATION (Less than 1 year), DATA MINING (Less than 1 year)

## Additional Information

SKILLS   
Data Science: Classification, Regression, Time Series Forecasting, Machine Learning, Data mining, Feature Engineering, Data Scraping, Data Manipulation and Visualization, Text mining, Clustering, Support Vector Machine, Random Forest, A/B Testing, Market Basket Analysis, Natural Language Processing   
Technical Skills: SQL, R, Python (NumPy, Pandas, Scikit), SAS, JMP, Tableau, MS Excel, Hive, Azure

# Indeed.com-02

# Chief Data Scientist

## Data Scientist

Seattle, WA

• 8 years of experience in statistical analyses, experiment design, economic research, and impact evaluation   
• Experience in building and leading data science team that assembles relevant data and creates value from data   
• Established track record of Health Economics publications (16 papers, some covered by major new outlets).

Authorized to work in the US for any employer

## Work Experience

**Chief Data Scientist**

**CareSkore**

Mountain View, CA

June 2016 to March 2018

• Lead the data science team to develop machine learning and predictive analytics models for population health management products   
• Combine clinical, socioeconomic, demographic, claim and behavioural data to better understand patients' risk profiles such as missing appointments, sepsis, readmission, mortality, likelihood to pay, among other models   
• Work closely with marketing and engineering team to define products' goals and guide implementations

**Senior Research Consultant**

**UnitedHealth Group - Health Services Research, Market Insights Group**

Mountain View, CA

September 2015 to June 2016

• Designed and conducted exploratory analyses, economic research studies, observational asnd impact evaluation studies using insurance claims and financial data   
• Collaborated with cross-functional stakeholders to understand their needs, formulate and complete suitable studies

**Senior Health Economist**

**American Cancer Society - Economic & Health Policy Program, Intramural Research**

Atlanta, GA

October 2013 to September 2015

• Led statistical analyses, data modelling and programming on obesity prevention and access to care initiatives   
• Authored 10 scientific papers and communicated results to policy analysts and journalists

**Research Assistant**

**University of Illinois at Chicago - Institute for Health Research and Policy**

Chicago, IL

June 2009 to October 2013

• Investigated relationships between dietary behaviours and nutrient intake in adults and children in the US   
• Worked on data management, data analysis, codebook writing and manuscript preparation for multiple projects

**Global Health Economics Graduate Intern**

**Amgen Inc - Global Health Economics Department**

Thousand Oaks, CA

June 2012 to August 2012

• Designed and executed an observational study using a large electronic medical record oncology database   
• Obtained training on cost effectiveness analysis and cost benefit analysis

**Lecturer**

**University of Illinois at Chicago - Department of Economics**

Chicago, IL

August 2006 to May 2009

Taught Microeconomics Theory and Business Application

## Education

**Ph.D. in Economics in Economics**

**University of Illinois at Chicago**

Chicago, IL

2013

**M.A. in Economics & Applied Econometrics in Economics & Applied Econometrics**

**University of Illinois at Chicago**

Chicago, IL

2008

## Skills

boosting (Less than 1 year), data science (3 years), Python (3 years), SAS (3 years)

## Additional Information

SKILLS   
• Programming: Python, PySpark, R, SQL, PostgreSQL, Stata, SAS   
• Data Science: Regression, classification, boosting, bagging.

# Indeed.com-03

# Senior Data Science Engineer

## Machine Learning Engineer/Data Scientist

Atlanta, GA

• Several years' experience in data science and machine learning.   
• Engineering team lead with over ten years of experience.

Authorized to work in the US for any employer

## Work Experience

**Senior Data Science Engineer**

**MailChimp**

Atlanta, GA

November 2015 to Present

• Built a production-grade audience-expansion system based on Elasticsearch and Python.   
• Prototyped a product-recommendation engine using Spark, Hive, and Scala.   
• Worked in Spark and sklearn on a redesign of the model used to predict email 'bounces'.   
• Led an eight-person team of data engineers and data scientists.   
• Worked across the company to translate business objectives into software.

**Data Systems Engineer**

**MailChimp**

Atlanta, GA

February 2015 to November 2015

Built a streaming-data platform that processes tens of billions of events per day. Written in Scala, with Akka Streams and Kafka.

**Software Architect**

**Travel Syndication Technologies**

Atlanta, GA

September 2012 to February 2015

• Worked on a five-developer team as technical lead. I was mainly designing and building ETL software in Scala.   
• Created robust and legible architectures.   
• Unified data from many different sources into coherent models; involved application of domain modeling skills and relational schema design.   
• Technologies: Scala, Java, MongoDB, MySQL; Hibernate, JPA, Anorm; Play, Akka.   
TECHNOLOGY OVERVIEW

## Education

**PhD in Applied Optical Physics**

**University of Nottingham**

September 2002 to September 2006

**MSc in Physical Chemistry**

**University of Nottingham**

September 1998 to June 2002

## Skills

Scala (3 years), Haskell (2 years), Machine Learning (2 years), Data Science (2 years), Spark (2 years), Python (3 years), Java (3 years)

## Links

[https://stackoverflow.com/users/383053/joffer](https://www.indeed.com/url?q=https%3A%2F%2Fstackoverflow.com%2Fusers%2F383053%2Fjoffer&h=73d14718)

## Certifications/Licenses

**Coursera Machine Learning**

Andrew Ng's course

# Indeed.com-04

# Data Science Engineer

## Data Science Engineer

San Francisco, CA

Data Science Engineer at tech non-profit HRDAG. Previously, a Computational Systems engineer at NERSC at the Lawrence Berkeley   
National Laboratory. Proficient in Python, R, Java, C/C++, shell scripting for machine learning applications. Experienced wit h Caffe, Torch,   
Theano.   
Machine Learning, Tensorflow.

Sponsorship required to work in the US

## Work Experience

**Data Science Engineer**

**Human Rights Data Analysis Group**

San Francisco, CA

June 2017 to 2018

Drove a data deduplication project to support the C olombia peace accord. Developer for data sciencefor peace project in El   
Salvador.   
• Developed code base in Python and R for data wrangling and aggregating a war dataset consisting of 6+ million records in inconsistent formats.   
• Named Entity Resolution using Hierarchical Agglomerative Clustering and Random Forests to match records that refer to the same victims that occurred in a military officer's career.

**High Performance Computer Systems Engineer**

**Lawrence Berkeley National Laboratory**

Berkeley, CA

June 2016 to June 2017

Berkeley, CA June 2016 - June 2017   
• Optimizing the allocations to high-speed memory on a GPU-like CPU called Intel Xeon-Phi processor, a multi-level memory system.   
• Achieved speedups of >70% on state-of-the-art benchmark codes by utilizing MCDRAM optimally.   
• Lead author of a conference paper titled "Optimize high-bandwidth memory usage on Knight's Landing using benchmarks and miniapps".

**Analyst**

**Deutsche Bank CIB Center**

Mumbai, Maharashtra

June 2014 to June 2015

• Developed a N atural Language Processing tool using NLTK and Scrapy to scrape news websites for counterparty news.   
• Bucket news based on keywords that indicate distress, and on sentiment of the news. This tool warned the team of bankruptcy of 2 Russian banks in 2014, and the Deutsche Bank was able to withdraw 5.5mn in deals.

**Operations Analyst**

June 2013 to June 2014

• Developed VBA tools to reconcile internal records of Service Level Agreements with partner bills.   
• Developed automated KPI dashboards for service uptime, call quality and data coverage in the telecom circle.

## Education

**M.S. in Industrial Engineering and Operations Research**

**University of California, Berkeley**

Berkeley, CA

August 2015 to May 2016

**B.Eng. in Computer Engineering**

**Delhi College of Engineering**

New Delhi, Delhi

August 2009 to June 2013

## Skills

PYTHON (2 years), DATA SCIENCE (1 year), MACHINE LEARNING (Less than 1 year), C++ (Less than 1 year), DATABASE (Less than 1 year)

## Links

[http://www.linkedin.com/agrimabahl](https://www.indeed.com/url?q=http%3A%2F%2Fwww.linkedin.com%2Fagrimabahl&h=7f0fc39d)

[https://github.com/agrimabahl](https://www.indeed.com/url?q=https%3A%2F%2Fgithub.com%2Fagrimabahl&h=3668dcca)

## Additional Information

TECHNICAL SKILLS   
GitHub: https://github.com/agrimabahl   
Languages: C++, Java, R, Python (Machine Learning, Data Science stack: Jupyter, Tensorflow, Keras, NumPy, SciPy,   
Pandas, scikit-learn, word2vec), Matlab   
Performance Tools: OpenMP, Intel TBB, Intel VTune, Linux shell scripting, CMake, gcc/icc compilation, git   
Database Tools: SQL Server, Oracle, MySQL, Hadoop

# Indeed.com-05

# Manager - Data Science

## Product Manager - Data Science

Tulalip, WA

• More than 2 decades of experience with in Information Systems Management including 6 years in Business Intelligence and Data Science / Analytics as a Product Manager, 10 years in Software Development Management and 7 years in implementing large scale Enterprise Systems (ERP)

Sponsorship required to work in the US

## Work Experience

**Manager - Data Science**

**Destination Canada**

Vancouver, Canada

January 2018 to Present

Canadian Crown Corporation wholly owned by the Government of Canada that uses data-driven marketing strategies to stimulate international demand and tourism export revenue for Canada   
   
• Led the Data Science initiative for the organization with annual budget of over CAD 125 million   
• Led the team to analyse data and implement various Machine Learning models for dimension reduction, classification, clustering and regression (SVM, Random Forest, RNN, HCA, t-SNA, PCA etc) to predict / classify tourist spending / visiting pattern   
• Led the data analytics team to provide insight on credit card spend, mobile / GPS data, border counts data from CBSA and Statistics Canada survey of foreign tourists to Canada and provided actionable market intelligence to Destination Canada’s key markets and tourism partners via BI dashboards   
• Led the initiative to streamline the data from various departments to align with data structure conducive with data science models

**Product Manager - BI and Data Analytics**

**ZE PowerGroup Inc**

Richmond, BC

2012 to December 2017

One of the biggest source of Enterprise Commodity Data providers in the world with more than 8000 data feeds and serving the fortune 500 companies in commodity trading, banks and utility companies around the globe.   
   
• Recommended, initiated and managed concept to launch of integration of third party products like R, MATLAB, Tableau, Power BI, Spotfire, Grass GIS and ESRI including product analysis, selection and managing the feature sets and their compatibility with the product suite   
• Managed the team of data scientists to provide customized analytics for clients using Parametric / Non Parametric modeling, Supervised / Unsupervised Learning, Linear / Logistic Regression, SVM, KNN, ANN, Random Forest, LDA, QDA etc.   
• Led the initiative to set up big data platform for the organization including Spark, HDFS, Cassandra, Sqoop and Kafka   
• Led the Business Analytics projects predictive analytics for clustering client based on data feed and predicting feed failure   
• Reported to Chief Product Officer and interacted on weekly basis to update the status of various projects and discuss the strategic direction of Product Management and Integration with third party products   
• Product manager for the tech team of 15 developers   
• Managed product vision, roadmap, feature briefs, wireframes, user stories and epics for the products and organized Competitive Analysis and Market Analysis for the feature / technology integration with the core product suite   
• Managed various stakeholders expectations by optimally balancing the backlog against product features by value, effort and criticality to meet the market demand

**Manager, Software Development and Enterprise Integration**

**Kal Tire**

Vernon, BC

2011 to 2012

Canada's largest independent tire dealer serving retail, commercial and mining industries with over   
230 retail stores in Canada and operations spanning 18 countries and 5 continents.   
• Led and mentored the Systems Development and Enterprise Integration team of 10 analysts,   
developers and subcontractors   
• Led the software development and integration of legacy ERP system with state of the art retail   
enterprise system (Tomax retail.net)   
• Planned and coordinated all systems development and integration activities   
• Provided technical direction and leadership   
• Designed roadmap for aligning various legacy systems with technology mandate of the organization   
• Defined and implemented the processes, toolkits and templates required for custom integration   
and system development including source code management, integration design, systems   
design, development, unit testing, integration testing and functional testing

**Systems Architect**

**Simon Fraser University**

Surrey, BC

2002 to 2011

One of the most respected universities in Canada and often ranked in top 3 comprehensive   
universities in Canada by prestigious Macleans magazine.   
• Managed several IT projects in various web technologies (.Net, J2EE, PHP) team sizes ranging   
between 4 and 10 members   
• Planned, developed and managed overall project schedule by creating project planning   
documentation, communicating project status, and facilitating interaction amongst stakeholder   
groups   
• Planned, developed and managed test strategy and test plan to conduct end to end testing   
covering all applications to ensure seamless integration of all applications   
• Successfully provided business process improvement solutions, managed several software   
development projects and delivered within the scheduled baseline plan over 90% of the time   
• Liaised with internal and external project stakeholders to understand and define project   
requirements, deliverables and estimated timeframes   
• Evaluated, sourced and negotiated several IT products based on the requirements from the stakeholders

**Enterprise IT Manager**

**Godrej InfoTech Limited**

Mumbai, Maharashtra

1997 to 2002

The first Indian ERP consulting company to get ISO certified and is the infotech division of Godrej   
Conglomerate, one of the leading business houses in India for over 100 years; CMM level 4 company.   
• Led a team of 30 functional and technical staff in ERP system Implementation and Customization business unit   
• Provided consultancy to various business units on ERP implementation projects through process   
mapping, gap analysis, pilot project, migration and live phases totalling over $10 millions   
• Analysed and studied several ERP/CRM systems like Baan, SAP and JD Edwards, Siebel, i2 and their alignment with client business processes and recommended improvements to existing   
business processes in line with best industry standards through detailed gap analysis   
• Designed and developed ERP module test manuals   
• Designed and developed data migration strategies, data templates and migration tools for ERP   
modules   
• Led several projects in integrating the legacy applications and migrating legacy application data   
to ERP in distribution (sales / purchase), finance, human resources and manufacturing by analyzing the legacy systems   
• Developed operation and working capital budget for the staff, operation and projects supervised   
and assisted in the departmental capital budget preparation   
• Gathered requirements from key / end users about the business intricacies and reported to the top management of the business divisions and presented to them about the project status, risks   
and scope change on periodic basis

**ERP Implementation Manager**

**Larsen & Toubro Limited**

Mumbai, Maharashtra

1992 to 1997

One of the largest and most respected companies in corporate India, specializing in technology,   
engineering, construction and manufacturing, with manufacturing units throughout India and internationally and offices all over the world.   
• Managed the mapping, piloting, migrating team for the full scale ERP implementation for the division   
• Selected as an ERP implementation manager for the division by demonstrating excellent   
analytical skills and sound understanding of key business dynamics and detailed understanding of the existing legacy systems   
• Researched large scale ERP systems (Baan / SAP) for the group and managed ERP   
implementation for the division   
• Managed strategic sourcing of procurement worth $10 million for turnkey EPC (Engineering,   
Procurement and Construction) projects   
• Led a collaborative sourcing initiative by involving suppliers, procurement and engineering   
divisions, which led to the savings of $2 million in purchasing costs   
• Developed and implemented various in-house applications to overcome the rigidity of the legacy   
mainframe system and saved tremendous subcontracting costs thereby providing a viable and potential business process improvements which were key to the success of the division

## Education

**Master of Business Administration in Business Administration**

**Simon Fraser University**

2008

**Master of Science in Information Technology**

**Simon Fraser University**

Vancouver, BC

2006

**Bachelor of Engineering in Mechanical Engineering**

**Calcutta University**

Vancouver, BC

1992

**Project Management Professional**

**Calcutta University**

Kolkata, West Bengal

## Skills

Machine Learning, Data Science, Artificial Intelligence (5 years)

## Additional Information

Software Skill Set   
   
Analytics Applications: Anaconda, Jupyter, R, Python (Scikit, Pandas, Numpy, Matplotlib etc.), Google Analytics, MATLAB, Advanced MS Excel (including VBA)   
Analytics: Parametric / Non Parametric, Supervised / Unsupervised Learning, Linear / Logistic Regression, SVM, KNN, ANN, Random Forest, LDA, QDA, Hierarchical Clustering, PCA, Ensemble Methods, Neural Networks, Deep Learning   
Business Intelligence: Tableau, Power BI, Qlik, TIBCO Spotfire    
Big Data: Hadoop, Spark, Hive, Cassandra, Sqoop, Kafka   
Java: J2EE, JSF, ORM (Hibernate), JMS, JAX-WS, JMS, Log4J, Maven   
Other Web Technologies: .Net / C#, OData, SOAP, REST, PHP, Groovy, Grails   
Database: MySQL, MSSQL, Oracle   
ERP / CRM: SAP, Baan   
GIS: ESRI, Grass   
Project Management: MS Project, JIRA, Confluence   
Operating Systems: Windows, Solaris, Linux

# Indeed.com-06

# Data Scientist

## Data Scientist

Washington, DC

Data Science, Machine Learning, Project Management

Authorized to work in the US for any employer

## Work Experience

**Data Scientist**

**Booz Allen Hamilton**

Arlington, VA

July 2016 to Present

Performed cloud based analytics using AWS. Natural Language and Machine Learning were used daily. Made great use of word embedding models. Research was conducted daily for general datascience. Programming was done in python. Worked with Elastic Search and other NoSQL databases (e.g. MongoDB) on a daily basis. Worked with visualizations using open source tools like D3.js, Seaborn, matplotlib, ggplot2, and Kibana. Gained knowledge of the big data stack while working with Data Engineers. Transitioned Architecture from RDBMS to Big Data Stack. Set up visualization platform. Built Chatbots. Conducted deep learning research using Tensorflow. Conducted Supervised Learning for classification. Performed Distribution Fitting. Statistical Data Analysis was performed on a daily basis.

**Data Science Mentor**

**Springboard**

March 2016 to Present

Responsibilities   
I teach my colleagues how to use R, Probability and Statistics, Data Wrangling, and different topics in AI according to their goals for the course. Topics include Machine Learning (includes Anomaly Detection) and Natural Language.

**Freelance Data Scientist**

**Numer.ai and Kaggle.com**

March 2015 to Present

On Kaggle.com I have worked with over 20 different data projects. Numerai.ai is one dataset concerning finance. On both websites I focus on statistical/data modeling, machine learning, feature engineering, and data wrangling. I have placed in the top 10% on a few projects on Kaggle.com. I use both websites to test new techniques in data modeling. I use the Kaggle forum to stay current on new techniques, technologies, and algorithms.

**Data Analyst**

**CSR, Incorporated**

August 2011 to December 2015

Responded to ad-hoc data requests and client queries. Creates and edits code for SPSS as well as formulas for Excel for reports. Creates reports using data submitted by Federal grantees. Conducts data verification. Trains new employees. Provided suggestions for political reports. Worked primarily with Criminal Justice data. Used Exploratory Data Analysis and Anomoly Detection to spot outliers. Automated workflow using R Statistical Package. Performed text analytics.

## Education

**B.S. in Criminal Justice**

**Virginia Commonwealth University**

2010

**Master's in Computer Science (Machine Learning)**

**Georgia Tech**

## Skills

Machine Learning (2 years), Data Analysis (5 years), Project Management (5 years), Data Science (3 years), Python (3 years), R (3 years), SPSS (4 years), Excel (5 years), Linux (3 years), Research (4 years)

## Links

<https://github.com/MelvinDunn>

[https://www.kaggle.com/melvindunn](https://www.indeed.com/url?q=https%3A%2F%2Fwww.kaggle.com%2Fmelvindunn&h=134398ae)

[https://www.linkedin.com/in/melvin-dunn-73b88026%EE%80%90](https://www.indeed.com/url?q=https%3A%2F%2Fwww.linkedin.com%2Fin%2Fmelvin-dunn-73b88026%25EE%2580%2590&h=0010fbe0)

## Certifications/Licenses

**Secret Clearance**

## Additional Information

Computer Skills   
SPSS, SQL, Ruby, Python, Linux (Ubuntu), Microsoft Office 2010: Excel, Word, Access, PowerPoint, Machine Learning, Neural Networks, Random Forests, Deep Learning, Regression, Supervised Learning, Unsupervised Learning, Additional Information Technology Experience   
Proficient in Microsoft, Structured Query Language (SQL), Microsoft Office, R Programming Language, Python Programming Language, Linux (various distros), Data Regression, classification, probability and statistical algorithms. Machine supervised/reinforcement/unsupervised learning. DataWrangling, Web Scraping, Dealing with Unstructured Data. Docker, Amazon Web Services (AWS), ElasticSearch, Kibana, MongoDB, CUDA, OpenCV (Python), Deep Learning, Convolutional Neural Networks, Knowledge of C++, Computer Vision, Computational Photography. Knowledge of Hortonworks and HBase. Socrata, Tensorflow, Keras. CEH and CISSP Associate Certifications.

# Indeed.com-07

# Data Science Software Developer

## Data Science Software Developer - KULWANT

Washington, DC

## Work Experience

**Data Science Software Developer**

**KULWANT**

December 2016 to Present

**Senior Software Engineer**

June 2013 to September 2013

• Developed components for fraud components and unit tested Java code   
• Utilized JIRA and Crucible for defect resolution and code reviews

**Senior Software Engineer**

**LEOBURNETTINTERACTIVE**

June 2008 to April 2013

• Developed web applications for Phillip Morris   
• Primary backend developer on enterprise project for Devry creating RESTful Web Services using Spring,   
Hibernate, JPA, MySQL, Maven, and Subversion   
• Developed and implemented new system to generate links for promotional emails utilizing Grails framework

**Software Developer**

**U B S**

July 2006 to January 2008

• Designed, developed, and deployed web-based application centered around Struts framework enabling users to change passwords on company website for Altria   
• Solely responsible for maintenance and application of patches of Java Application allowing authentication to the   
Internet for Devry   
• Developed payment processing RESTful service for Coke

**Business Consultant**

**STERLINGCOMMERCE**

August 2005 to July 2006

• Determined the functional business requirements, installed and configured Warehouse Management software   
• Coordinated with development resources offshore to resolve defects and provide enhancements in a timely and efficient manner

**Software Engineer**

**I B M**

October 2004 to July 2005

Developed JSP pages using Action Forms to populate data and validating data in Action classes in using   
WebSphere Application Developer in Struts   
• Documented business process with Sequence Diagrams using Microsoft Visio

**Programmer**

**FIDELITONE**

May 2002 to March 2004

Led all phases of software development for B2B application (IBM Partner Agreement Manager) including   
establishing requirements, designing, testing and implementing processes to enable business-to-business   
transactions primarily by XML and FTP

**Technical Consultant**

**ACCENTURE**

April 2000 to March 2001

Worked on Chase's CRM application Customer Assist as a junior Java developer.

## Education

**Masters of Science in Journalism**

**NORTHWESTERN UNIVERSITY**

August 2015

**Bachelor of Science in Management Information Systems**

**UNIVERSITY OF ILLINOIS**

Chicago, IL

December 1999

## Skills

DATA SCIENCE (1 year), DATA ANALYTICS (Less than 1 year), DATA VISUALIZATION (Less than 1 year), HP-UX (Less than 1 year), JSON (Less than 1 year)

## Links

<http://data.rollcall.com/congressionalmadness>

## Additional Information

SKILLS/TOOLS   
Tableau   
SQL   
GitHub   
ADDITIONAL COURSEWORK   
( atGeneralAssemblyandMakerSquare )   
   
Completed courses: Data Science Data Analytics UX/UI Design Web Development   
   
• Developed expertise in predictive modeling, supervised and unsupervised machine learning, natural language   
processing, data visualization and web scraping.   
• Learned Python libraries sklearn, pandas, pyspark, BeautifulSoup, scrapy, seaborn, matplotlib, requests, json   
   
Projects:   
• Developed model to forecast final NBA game scores using linear regression and random forests. Data was   
obtained by scraping NBA's website.   
• Created predictions for Michelin Guide's restaurant using Support Vector Machine and logistic regression   
• Data mined Indeed to predict salaries of data scientist using natural language processing

# Indeed.com-08

# Data Science Researcher

## Data Scientist

Denver, CO

Seeking a full-time data science position that can utilize my modeling skills, statistical knowledge,   
and creativity to discover valuable business insights contained in large stores of data.

Authorized to work in the US for any employer

## Work Experience

**Data Science Researcher**

**Solar Durability and Lifetime Extension Center**

Cleveland, OH

October 2016 to May 2018

• Designed data pipelines to automate machine learning extraction of crack features in solar panel backsheets   
• Implemented cycle counting and clustering algorithms on weather data time series to quantify climates   
• Utilized parallel computing clusters and a custom SLURM job architecture to decrease run time   
• Collaborated nationwide with 5 other institutions on a Department of Energy SunShot grant   
• Managed multiple undergraduate students working on data acquisition

**Mechanical Engineering Intern**

**Janicki Industries**

Hamilton, WA

May 2016 to July 2016

• Analyzed data from customers' CAD models and compared to past projects to determine part pricing   
• Researched competitors' and developed a strategic plan to pursue a multi-million dollar military contract   
• Designed an autoclave cart to minimize processing time for a series of 24 large, complex composite parts

**Mechanical Engineering Intern**

**Cascade Designs**

Seattle, WA

May 2015 to August 2015

• Developed thermodynamic models and ran calculations in Matlab for an emerging portable stove project   
• Conducted rigorously standardized experiments to validate theoretical models   
• Communicated observations from thermal imaging data to management in technical reports   
• Studied field failures of snow shovels to improve manufacturing processes and product robustness

**Product Design Intern**

**Tarptent Ultralight Shelters**

Nevada City, CA

May 2014 to August 2014

• Led the design of key improvements to the Bowfin, MoTrail, and Rainshadow3 models   
• Developed a patent-pending trekking pole support system to reduce weight and increase head room   
• Modeled designs in CAD and sewed prototypes for field testing and design validation   
• Integrated carbon fiber support structures into tent fly to decrease setup time and overall weight

## Education

**MS in Mechanical Engineering**

**Case Western Reserve University**

Cleveland, OH

August 2018

**BS in Mechanical Engineering**

**Case Western Reserve University**

Cleveland, OH

May 2017

## Skills

MACHINE LEARNING (2 years), DATA VISUALIZATION (2 years), Python (1 year), R (2 years), DataScience (2 years), Analytics (2 years), Git (2 years), Business Intelligence (3 years), Business Management (3 years)

## Awards

**Gustav Kuerti Award**

May 2017

"To the senior in mechanical and aerospace engineering who has demonstrated the highest level of scholarship"

**Florence Keuerleber Prize**

May 2017

"To an undergraduate student who has seriously pursued and excelled in the study of a modern language while majoring in another area"

**Max Kade Award**

April 2015

For excellence in German

**Best Chapter Officer**

April 2015

Case Western Reserve University Beta Theta Pi

## Publications

**A Non-Destructive Method for Crack Quantification in Photovoltaic Backsheets Under Accelerated and Real-world Exposures**

[https://www.sciencedirect.com/science/article/pii/S0141391018301551](https://www.indeed.com/url?q=https%3A%2F%2Fwww.sciencedirect.com%2Fscience%2Farticle%2Fpii%2FS0141391018301551&h=55c3714c)

July 2018

Volume 153 of Polymer Degradation and Stability

## Additional Information

RELEVANT SKILLS   
• Programming Languages: R, Python, Matlab, C++ and their integrated development environments   
• Machine Learning and Statistical Methods: cross validation, bootstrap, random forests, neural networks   
• Business Acumen: 3 years owning online business, responsible for market and profit and loss analyses   
• Data Visualization: artistic talent for communicating data insights through concise, informative figures   
• Agile Software Development: Git repositories, function documentation and packaging, Bash terminal   
• Databases: familiarity with SQL, Excel, and the Hadoop ecosystem

# Indeed.com-09

# Sr. Data Analysis Manager supporting Data Science

## Sr. Data Analysis Manager supporting Data Science - Capital One Bank

Olney, MD

Authorized to work in the US for any employer

## Work Experience

**Sr. Data Analysis Manager supporting Data Science**

**Capital One Bank**

McLean, VA

2009 to January 2018

McLean, VA 2009 - 2018   
Sr. Data Analysis Manager supporting Data Science   
Responsible for model operations management to consult in the design of, implement, execute, maintain, and monitor all batch risk models and related deposit policies.   
• Managed team of 4 to 6 Analyst to conduct whole range of model life cycle activities for Consumer and Small Business Deposit and Fraud Risk models and related policies including Funds Availability,   
Courtesy Overdraft Limit management, Remote Deposit limit management, and Deposit Fraud.   
• Code and maintain models to operate efficiently on the optimal target platform and automatically   
deliver results to operational platforms for decisioning.   
• Provide first line support for all model and related policy operational issues including Sev One Ops.   
• Monitor planned and surprise data environment changes and evaluate operational impacts.   
• Consult on next generation machine learning platform design.   
• Consulted, advised, prepared and transmitted analytical data as subject matter expert in loan and deposit, application and servicing data from Chevy Chase Bank to Capital One TeraData Bank   
Analytical Data Warehouse.

**Assistant VP/Vice President Data Based Solutions**

**Chevy Chase Bank**

Bethesda, MD

1997 to 2009

Bethesda, MD 1997 - 2009   
Assistant VP/Vice President Data Based Solutions   
Data Based Solutions was the internal consultant to Card, Lending and Retail Branch divisions of Chevy   
Chase Bank. Our mission was to bring statistical models and methods to create a competitive advantage by extending the residual value of data to maximize overall profitability.   
• Manage department of 3-5 SAS analysts to develop and implement predictive scoring models.   
• Specify modeling techniques including linear or logistic regression, CHAID, or neural network.   
• Coded Neural Network model in SAS for batch execution.   
• Oversee development of predictive models by external vendors.   
• Determine and specify sampling techniques for development and validation samples.   
• Formulates and communicates recommendations for business process improvements.   
• Recalibrate and rebuild models when necessary or as additional data sources or implementation   
platforms become available.   
240-419-7404 Dennis E. Miller

**Knowledge Based Marketing**

**Wired and Wireless Telecommunications, Cable Entertainment, Electric**

Chapel Hill, NC

1995 to 1997

Statistical Analyst/Senior Statistical Analyst   
Customer Management Services was a full service database marketing and intelligence start up launched through University of North Carolina servicing multiple industries.   
• Constructed models for Wired and Wireless Telecommunications, Cable Entertainment, Electric   
Power Utilities, Financial Services and Membership Societies for both Consumer and Business-to-   
Business Applications, Attrition, Cross-Sell/Up-sell, Response, New Home Mover, Fraud and   
Profitability.   
• Develop predictive models using various technologies and software including: Neural Network,   
Regression, and CHAID.   
• Made client presentation and recommendations on model systems.   
• Assess and interpret model fit and effectiveness, and monitor population and characteristic stability and performance of implemented models. Recalibrate and rebuild models as necessary.

**Senior Electrical Engineer**

**Halliburton NUS Environmental Corp**

Raleigh, NC

1989 to 1995

Commercial Architectural/Engineering firm servicing the US Commercial Nuclear industry.   
• Manage analysis of critical AC/DC electrical power systems in commercial nuclear power facilities.   
• Conduct detailed electrical design engineering of plant modifications.   
• Developed database applications supporting computational analysis.   
• Radioactive Waste treatment facility engineering design support.

## Education

**Master of Business Administration in Technology Management**

**University of Phoenix**

**Bachelors of Science in Electrical Engineering Technology**

**Messiah College/Temple University**

## Skills

SAS Programming (10+ years), Data Science (10+ years), Management (10+ years), Microsoft Office (10+ years), Data Mining (10+ years), Data Warehousing (10+ years), Banking (10+ years), Risk Management (10+ years), Business Process Improvement (10+ years)

## Links

<http://www.linkedin.com/in/dennisemiller>

## Patents

**System for and method for determining overdraft protection (#20140304131 A1)**

[https://patents.google.com/patent/US20140304131A1/en](https://www.indeed.com/url?q=https%3A%2F%2Fpatents.google.com%2Fpatent%2FUS20140304131A1%2Fen&h=0ac47e84)

2014

A system and method in accordance with example embodiments may calculate a velocity of money associated with a particular account and a particular customer. Additionally, a system and method in accordance with example embodiments may calculate an overdraft risk score based on the velocity of money, where the overdraft risk score may be used to determine an overdraft protection limit for an account.

# Indeed.com-10

# FACEBOOK - Lead Data Science Engineer

## FACEBOOK - Lead Data Science Engineer - IDC Analytics and Data Science

San Jose, CA

\* Delivered a Product using Data Science, Data Lake and Machine Learning for 'Digital Marketing'. The   
product covers Customer 360, churn, recommendation engine, credit and risk rating, digital marketing   
and campaign management. Implementation of this product resulted in significant increase in sales,   
revenue and customer satisfaction.   
\* Strong technical knowledge in building Data Lakes and Advanced Analytics on premise and in the   
cloud (AWS and Azure) environments using technologies such as Hadoop, Spark, Scala, Python   
(NumPy, SciPy and others), MLib, NLP (NLTK and OpenNLP), MATLAB, R, Hive, SQL, Kafka,   
HBase, Impala, MQTT, Cassandra, MapReduce, Java/J2EE, Pig, MongoDB, NoSQL, S3, EC2, EMR,   
Redshift, Kinesis etc.   
\* Delivered Fraudulent Claims detection system by Integrating EDW, BI, Data Lake and Data Science   
for a major P&C. This system helped business in saving time, analyzing risk and improved rating   
mechanism.   
\* Derived Enterprise Data Warehouse and Business Intelligence roadmaps, transformed into programs   
and projects, executed and delivered scalable, best performing EDW-BI platforms and solutions so   
that Business can make quick and accurate decisions bases on the DATA.   
\* Designed, Developed and Implemented DATA DRIVEN SOLUTIONS and EDW-BI, ETL, ELT,   
Advanced Analytics and Data Visualization Solutions using SAS, SPSS, Tableau, D3, QlikView,   
MSBI, Power BI, Talend, Informatica, Datastage etc.   
\* Highly experienced in processing and integrating structured and unstructured data in batch and real   
time. Developed scalable and best performing technical solutions to transform unstructured data   
(social networking, click stream, twitter and others).   
   
1  Kamal Basha M S   
 \* Always open to learn new tools and technologies, passionate about providing best value for business   
by delivering data driven solutions and products.

## Work Experience

**FACEBOOK - Lead Data Science Engineer**

**IDC Analytics and Data Science**

August 2017 to Present

Design, development and implementation of Advanced Analytics   
Platform to support several customer and vendor centric use cases.   
   
As a Lead Data Science Engineer performing the following activities:   
   
• Developing use cases by working with Business Executives.   
• Identifying data source and developing data pipeline using HIVE, SQL, Python, Java, Scala,   
MySQL, Oracle, PostgreSQL, Presto and Raptor.   
• Test and Training models in Data Lake, sharing scored results with Business Executives.   
• Improving the accuracy of the models through feature engineering and variables selection.   
• Implemented models such as Linear regression, Logistic Regression, Naïve Bayesian classifiers   
and Decision trees.   
• Developing Data visualization layers using Tableau and open source technologies to present   
scored results.   
• Data Analysis and Data mapping, design, development and implementations end-to-end technical   
solutions.

**Data Scientist and Architect**

**Eureka**

September 2016 to July 2017

Eureka - End to End Digital Marketing Campaign Management System: Design, development and implementation of EUREKA product which contains BIG Data and Data Science tools and technologies.   
Eureka supports digital marketing, customer 360, churn and recommendation engine.   
   
As a Data Scientist and Architect performing the following activities:   
   
• Data Analysis and Data mapping - Identified 32+ data sources such as CDR, Exadata, e- Commerce, ClickStream, Triangulation and others in relation with Use Cases customer 360,   
churn, recommendation engine.   
• Defined roadmap for future state BIG Data Advanced Analytics Platform based on Internal and External use cases.   
   
2• Extracted the value of Real-World Data such as Triangulation and Clickstream to provide insight   
into customer demographics and behavior to find viable pathway towards growth in the competitive telecom industry.   
• Involved in Data Science models such as K-means clustering, Association rules, Linear   
regression, Logistic Regression, Naïve Bayesian classifiers, Decision trees, SVM and others to accurately predict and project Customer 360, Churn and Next Best Offers.   
• Developed Real-Time, scalable Recommendation engines for Next Best Offer.   
• Created Customer 360-degree view with variety of classifications. Developed web based user   
interface for Marketing and Sales team with upstream and downstream APIs to trigger channel   
marketing with external vendors.   
• Developed Data Ingestion pipeline for Data Lake using Spark, Scala, Sqoop, Python (NumPy,   
SciPy and others), MATLAB, Hive, Impala, HBase, Cassandra, Kafka, Flume, Oozie and other   
Cloudera stack tools and technologies.   
• Developed web based UI for executive management using Java, Python, MySQL, Hive, Node.js,   
AngularJS, LAMP, D3, Tableau and others.

**Data Scientist and Architect**

**Prelude Systems**

August 2012 to August 2016

Fraudulent Claims Detection System: 21st Century Insurance Purchased by Zurich and merging   
systems of Farmers Auto Insurance with 21st Century systems.   
   
Data Science and BIG Data: There is an increased focus on Data Science and BIG data to use it for business advantage with focus on sales, marketing, search engine data and fraudulent claims.   
   
Contribution: As a Solutions Architect and Junior Data Scientist:   
• Delivered Data Science Solutions using Predictive Modeling, Machine Learning, Text Mining,   
Probability and Mathematical models on high volume data.   
• Implemented Advanced Analytics models such as K-means clustering, NLP, Associate rules,   
Linear regression, Logistic regression, Naïve Bayesian classifiers, Decision tress, Time Series   
Analysis and Text Analytics.   
• Involved in POCs related to BIG data solutions and advanced analytics using Hadoop eco system   
tools such as Spark, Scala, MapReduce, Pig, Hive, HBase, Impala, Cassandra along with other   
technologies such as Python, D3, Tableau, QlikView, R, SAS etc.   
• Designed 'Data Solution' that converts BIG Data in to meaningful structured data to meet the Current and Future needs of Business.   
• Provided Data Integration Solution to bring BIG data related ELT process and Structured   
(RDBMS and others) Data related ETL process in to integrated platform.   
• Created a Roadmap to develop and implement Data Solutions using new age technologies.   
• Analysis of existing EDW-BI systems of Farmers and 21st Century Insurance.   
• Analysis of BIG data and other source systems including claims, consumer, policy, search logs   
and third party data.   
• Analysis of Structured Data and related ETL process.   
• Design and Development of Data Lakes and Sandbox for Advanced Analytics.   
• Involved in procuring KPIs and Business Needs with objectives of merging two EDW-BI systems.   
• Participated in procuring KPIs and Business Needs keeping BIG Data and Structured data as key   
aspects.

**EDW-BI Architect**

**UNIVERSITY OF CALIFORNIA**

San Diego, CA

October 2007 to July 2012

Enterprise IT Architecture: UCSD ranked 7th in U.S. research universities. UCSD campus enrollment is   
about 29000 students and annual revenues are $2.4 billion. Student systems, Department specific   
systems, EDW-BI are highly dependent on Data specific solutions.   
   
Contribution: As an EDW-BI Architect:   
   
3• Understanding different departments Business needs, analysis of current IT environment.   
• Review of OLTP and OLAP systems including student systems and department systems.   
• Analysis of Data dependencies, redundancies, quality and correctness.   
• Creating the current state IT environment report and providing road map for future state IT   
environment to align with business objectives and vision.   
• Defined Business Intelligence roadmap and directed teams to implement 'Real time Data   
Warehousing and Decision Support Systems'. Implemented EDW-BI solutions.   
• Responsible for data needs of about 45000 users community on daily basis that includes   
students, financial departments, employees, engineering, R & D, medical and all other   
departments within UCSD campus.   
• Responsible for Data Governance, Data Management, Regulatory Reporting and BI reporting   
and analytical systems. Designed and Developed Metadata Management System.

**EDW-BI Architect**

**ICW GROUP**

March 2007 to September 2007

Enterprise Data Warehouse - BI Solutions - Predictive Analytics: Enterprise Data Warehouse (EDW)   
design, architecture and infrastructure to support BI solutions for Accounting, Actuary, Claims, Finance and Auto departments. Also, proving the design for 'Enterprise Data Warehouse' and ODS to support   
'Predictive Analytics' initiatives.   
   
Contribution: As an 'EDW-BI Architect':   
• Prepared 'EDW Assessment' report and presented to senior management along with issues,   
impacts and resolution.   
• Defined a new strategy to implement 'Enterprise Data Warehouse (EDW) and Business   
Intelligence (BI)' initiative that will save $2 millions for the organization.   
• Prepared a Program Plan to implement EDW & BI solutions to support Accounting, Finance,   
Actuary, Claims and Auto departments.   
• Defined resource needs and budget to implement EDW & BI Roadmap.   
• Developed a strategy to implement 'Predictive Analytics' initiative.   
• Provided guidelines to build 'Production Support Team' and 'User Support Team'.   
• Implementing Basic Infrastructure and Architecture to support future EDW & BI projects.

**EDW-BI Architect**

**21ST CENTURY INSURANCE**

June 2006 to February 2007

Enterprise Data Warehouse (EDW) and Business Intelligence (BI): The EDW-BI department supports   
multiple business areas such as Finance, Actuary, Claims and Marketing along with Regulatory   
Reporting and Data Governance. EDW also maintains SOX standards, SDLC and PMLC deliverables.   
   
Contribution: As an EDW-BI Architect:   
• Analysis of existing Data Warehouse and creating a current state of Data Warehouse Report.   
• Involved in procuring KPIs and Business Needs to design Enterprise Data Warehouse and Business Intelligence System.   
• Designed EDW-BI road map based on current state DW and future business needs that includes   
ETL, Data Logic, BI Reports, Analytics and Real Time Data Warehouse.   
• Responsible in transforming DW in to EDW-BI with Subject specific Data Marts such as Claims,   
Policy, Customer, Premium, Agents and Billing.   
• Involved in Logical and Physical Data Models, OLAP based Star Schemas using Ralph Kimball   
methodologies.   
• Involved in design, development and Implementation of EDW-BI complete system using IBM   
Data Stage, Microstrategy and other tools and technologies.   
• Participated in User Acceptance Testing and Delivery of BI Reports and Analytical components.   
 

**EDW-BI Architect**

**STATE FARM INSURANCE**

September 2001 to May 2006

New Initiative to build EDW & BI systems to support Decision Making Process. These systems process   
internal and external data to build highly analytical systems with subject lines such as Claims, Premium,   
Agents, Sales and Other Aspects.   
   
Contribution: As an EDW-BI Architect:   
• Responsible for Source Systems and Source Data Analysis including Policy Processing Systems,   
Claims Systems, Billing Systems, ERP applications and other systems.   
• Involved in procuring KPIs and Business Needs to design Enterprise Data Warehouse.   
• Designed Subject specific Data Marts such as Claims, Policy, Customer, Premium, Agents and Billing.   
• Involved in Logical and Physical Data Models building process using Ralph Kimball   
methodologies.   
• Responsible for Data Mapping, Data Quality and Data Accuracy. Implemented several automated   
processes in and around data.   
• Designed ETL process to build EDW using new age technologies such as Informatica.   
• Involved in design, development and Implementation of BI reports and dashboards using Cognos   
BI tool.   
   
Electronic Bill Presentment and Payment System (EBPPS) - As a Solutions Architect designed,   
developed and implemented EBPP system using JAVA, J2EE and Web Services.

**Subject Matter Expert**

**FARMERS INSURANCE GROUP**

August 1999 to August 2001

E-Billing: This project helped customers to choose electronic payment methods such as pay by credit   
card and electronic transfer of funds from their bank.   
   
Contribution: Led the teams in feasibility study, project planning, execution, implementation and user   
training. The project budget $2.75 million, implemented within budget utilizing 15 resources.   
   
• Performed Technical Lead and Subject Matter expert roles as and when needed along with project management responsibilities.   
• Responsible for Quality procedures, production implementation, Onsite-Offshore   
coordination, phased roll out, user training and knowledge transfer.

**Senior System Analyst**

**WELLS FARGO**

October 1997 to July 1999

SSA   
Multi Branch Banking: Wachovia merged with 'Core States' bank. The Accounting application of Wachovia needs to process the accounting data of 'Core States' bank using existing Wachovia systems.   
Enhanced the existing system to perform transactions between multiple branches.   
   
Contribution: Involved in Data mapping, Analysis of existing system, Development of programs to cleanse the data and to migrate the data and Led the team of developers to build one-time datamigration   
programs.

**Senior Systems Analyst**

**K-MART**

November 1996 to September 1997

Merchandise Hierarchy System: Enhancement of existing Merchandise Hierarchy System at K-Mart to introduce new level of grouping known as 'Category Cluster'. The reports at category cluster level provide   
valuable information which results in better planning and monitoring the performance.   
   
5   
   
Contribution: Independently involved in the Analysis of existing system and Modification of impacted   
programs, Unit testing, Integration & Functional testing and Production Implementation.

**Software Engineer**

**Tech Mahindra**

June 1993 to October 1996

Stores Accountancy System: Involved in the development of Stores Accountancy System, Employee   
Information System, Payroll System and Accounts Receivable system for leading Manufacturing   
Company.   
   
Contribution: As a Systems Analyst responsible for Analysis, Design, Development, Modifications,   
Unit, Integration and Functional Testing.

## Education

**MBA in MIS**

**JNTU**

India

November 2013 to February 2016

**Bachelor of Engineering in Engineering**

**Osmania University**

India

August 1989 to May 1993

## Skills

BI (10+ years), Business Intelligence (10+ years), Data Science (4 years), EDW (10+ years), Java (6 years), Python (10+ years)

## Certifications/Licenses

**DELL EMC Data Science Associate**

July 2016 to Present

**AWS Solutions Architect**

February 2017 to Present

**IBM Business Intelligence Designer**

April 2006 to Present

## Additional Information

TECHNICAL SKILLS   
   
\* Data Science, BIG Data, BI and EDW: Spark, Scala, Python, R, Hadoop, Hive, HBase, Kafka,   
Impala, Cassandra, MongoDB, Storm, NiFi, Flume, Cloudera Stack and HortonWorks Stack, NumPy,   
SciPy, MLib, NLP (NLTK and OpenNLP), MATLAB Tableau, Power BI, Cognos, Business Objects,   
MicroStrategy, Informatica, DataStage, Talend, SAS, Google Analytics and others.   
\* Database: Oracle, DB2, SQL server, Teradata, Mongo DB, MySQL, PostgreSQL and others.   
\* Web Technologies: Python, Java, J2EE, CSS, XML, PHP, HTML, LAMP and others.   
\* ERP: Oracle Applications, PeopleSoft, SAP HANA and BW.   
\* Other Tools and Technologies: Jenkins, Jira, Confluence, GitHub, Erwin, ER Studio, Autosys,   
SourceForge, Control-M, Zena, Primavera, TeamTrack and others.   
\* Operating Systems: UNIX, Linux, Microsoft Windows series, MS-DOS and others.

# Indeed.com-11

# Data Scientist Manager / Senior Machine Learning Engineer

## Senior Data Scientist / Machine Learning Engineer

New York, NY

https://www.linkedin.com/in/jdmarold   
http://www.jakemarold.com   
   
Natural Language Processing (NLP) techniques: statistical parsing, part-of-speech tagging, metadata extraction, NER - named entity recognition, text extraction from multiple document types, web crawling and scraping). Topic modeling using LDA, LSI, Doc2Vec, Word2Vec, and other embedding techniques.    
   
Supervised and unsupervised machine learning techniques: recurrent, convolutional, and feed-forward neural networks, traditional machine learning algorithms including logistic regression, SVM, decision trees / random forest, etc. Graph and network analysis.    
   
Automatic Speech Recognition (ASR) techniques: analysis of mono and stereo 8kHz and 16kHz audio, audio transcription, acoustic model adaptation, language model adaptation, speaker diarization   
   
Frontend techniques: data exploration and visualization dashboards using angular framework, bootstrap, HTML5, typescript, javascript, etc.   
   
Tensorflow, Theano, Keras, H2O.ai, SciKit-Learn, SpaCy, Gensim, Pandas, NumPy, SciPy, SymPy, BeautifulSoup, MultiProcessing, Flask, NetworkX, NLTK, MRJob, RegEx, pySpark, Hadoop HDFS, git Seaborn, Matplotlib, Bootstrap, ElasticSearch, AngularJS, Angular2-5.

Authorized to work in the US for any employer

## Work Experience

**Data Scientist Manager / Senior Machine Learning Engineer**

**J.P. Morgan Chase & Co.**

New York, NY

November 2015 to Present

Lead data scientist developer. Responsible for strategic machine learning and technology decision making. Involved at multiple levels of application development lifecycle - ideation, strategy, oversight, prototype, data movement, data science, and some frontend development. Relationship management with select technology partners.    
   
Improved performances across multiple lines of business by applying NLP and machine learning techniques (experience with multiple domains of textual information - contributions have helped to generate over \$10 million in funding for various projects)   
   
Original member who helped establish the team as the machine learning center of excellence for the firm from 2015- April 2018. Worked with other LOBs to help them grow and establish their own datascience functions during that time period.   
   
Redesigned firm audio handling for speech recognition capabilities (virtual assistants, signal processing, speaker diarization, audio transcription (with acoustic and language model adaptation), classification of speech).   
   
Projects (past and present) are frequently included in quarterly business updates to the operating committee

**Data Science Fellow**

**The Data Incubator**

Washington, DC

June 2015 to August 2015

The Data Incubator (http://www.thedataincubator.com/) is a top-tier data science program    
   
• Selected in top 2% from over three thousand PhD applicants (based on statistical and programming ability and data analysis skills) to participate in a rigorous data science program.   
   
• Predicted Yelp business ratings using machine learning algorithms and natural language processing (NLP) (Scikit learn/sklearn: Random Forest, K-nearest neighbors KNN, multiple regression techniques).   
   
• Developed predictive models for NYC restaurant inspection grades using machine learning algorithms and NLP   
   
• Analyzed 8GB of Centers for Medicare and Medicaid (CMS) Provider Utilization and Payment Databy regional networks to view cost distribution for services and created D3 compatible visualizations (major packages: Pandas, Vincent).    
   
• Performed social network analysis on NYC social diary unstructured data photo captions (Web Scraping, Beautiful Soup, Regex, NetworkX). Calculated PageRank based on social graph analysis. Who is the most popular person in NYC? Which couples are most likely to be seen together?   
   
• Generated predictive models for weather data using time-series methodology. What is the chance of rainfall at 8pm in NYC on Mondays in July?   
   
• Used Hadoop and Map Reduce to analyze >10GB of Wikipedia data. What are the top 100 words used in English Wikipedia pages?   
   
• Developed financial models for Capital One customers to predict risk   
   
Skills Used   
Technical skills: Python (major packages: Pandas, NetworkX, NumPy, SciPy, SymPy, BeautifulSoup, Multiprocessing, Matplotlib), R, Hadoop Map Reduce.   
   
Soft skills: Communication (written and oral), expertise with common office software packages (Microsoft Office - PowerPoint, Excel, Word), and graphics generation

**Computational Biophysicist / Research Scientist**

**Johns Hopkins University**

Baltimore, MD

August 2009 to June 2015

• Wrote software packages for biophysical data analysis and visualization   
• Developed platform for global (shared parameter) non-linear and linear least squares regression of biophysical titration and time series data - resulted in 10-fold increase in data analysis efficiency   
   
• Generated custom statistical mathematical models (based on probability theory) to be used in fitting routines    
   
• Improved F-statistics and Bootstrap methodology for obtaining parameter confidence limits. My bootstrapping procedure led to a 12-fold speed enhancement through the use of multiprocessing.    
   
• Platform speaker at the 59th Biophysical Society Meeting (the world’s largest Biophysics conference with greater than 7000 attendees http://www.biophysics.org/)   
   
• Organizing committee member and speaker at the NSF Protein Folding Consortium (meeting for the most advanced protein folding research and analytical techniques in the world http://proteinfoldingconsortium.com/)   
   
• Co-founder - Johns Hopkins University Computational Graphics Club (Founder of a club to share visualizations and analysis code in multiple programming languages)    
   
• Invited seminar speaker in the Department of Biochemistry and Molecular Biology, University of MN    
   
Skills Used   
Technical skills: Python (major packages: NumPy, SciPy, SymPy, lmfit, MultiProcessing, Matplotlib), Matlab, and Mathematica.   
   
Soft skills: Communication (written and oral), expertise with common office software packages (Microsoft Office - PowerPoint, Excel, Word), and graphics generation

## Education

**Doctor of Philosophy in Computational and Molecular Biophysics**

**Johns Hopkins University**

Baltimore, MD

2009 to 2015

**Bachelor of Science in Biochemistry/Molecular Biology**

**University of Minnesota Duluth**

Duluth, MN

2009

**Bachelor of Arts in Chemistry**

**University Minnesota**

Duluth, MN

2009

## Skills

Python (5 years), SQL (2 years), Non-linear and Linear Regression (6 years), Data Science (4 years), Natural Language Processing (NLP) (2 years), Neural Networks (1 year), Javascript (Less than 1 year), Java (Less than 1 year)

## Links

<http://www.jakemarold.com>

[https://www.linkedin.com/in/jdmarold](https://www.indeed.com/url?q=https%3A%2F%2Fwww.linkedin.com%2Fin%2Fjdmarold&h=e3e36341)

# Indeed.com-12

# BIG DATA/DATA SCIENCE - PRINCIPAL INSTRUCTOR

## BIG DATA/DATA SCIENCE - PRINCIPAL INSTRUCTOR - Think Big Analytics, a Teradata company

Saratoga Springs, UT

Vision-driven change agent with career-long record of data science, big data management, and marketing success for leading organizations   
   
Proven talent for aligning business strategy and objectives with established data science and machine learning paradigms to achieve maximum operational impacts with minimum resource expenditures. Growth-focused thought leader with expertise spanning data analysis & management, big data, process improvement, business automation, web development, tool design, statistical analysis, datavisualization, marketing, technological growth strategy, trading algorithms, management development, training and mentoring, technology solutions, cross-functional team leadership, performance assessment, client relationship management, and project management. Exceptional administrator with keen interpersonal, communication, and organizational skills, as well as policy development and resource allocation expertise.

Authorized to work in the US for any employer

## Work Experience

**BIG DATA/DATA SCIENCE - PRINCIPAL INSTRUCTOR**

**Think Big Analytics, a Teradata company**

South Jordan, UT

May 2015 to Present

South Jordan, UT, May 2015 to Present   
Think Big Analytics is a Big Data/Data Science pure play open source consulting firm. Think Big Analytics is a Teradata company.   
BIG DATA/DATA SCIENCE - PRINCIPAL INSTRUCTOR   
• Function as a Principal Technical Instructor for the Think Big Academy, providing Big Data and DataScience training for Think Big internal staff and Think Big clients.   
Key Accomplishments   
• Designed and presented curriculum for the Academy, including hands-on development exercises in Hadoop, Hive, Pig, Spark, R, Machine Learning, TensorFlow, and other Big Data/Data Sciencetechnologies as a Think Big Boot Camp (employee on-boarding) and client-focused course producer and instructor.   
Skills Used   
• Data Science (R programming, statistical inference, linear regression, machine learning, and dataproducts production), and Data Engineering (Hadoop, MapReduce, Hive, Pig, Presto, HDFS, Storm, Spark, Java, Scala).

**SENIOR DEVELOPER - PROJECT LEAD**

**CSS Corp**

Draper, UT

March 2013 to April 2015

• Furnished website development and maintenance services for Estee Lauder Corp. (LAMP stack).   
• Performed as the senior developer and project lead for CSS Corp in their Draper offices, providing back-end engineering support for two of the premiere brands under the Estee Lauder (NYC) umbrella.   
• Drove and coded on-going back-end development for Bobbi Brown, North America (http://bobbibrown.com).   
• Steered and coded on-going back-end development for Prescriptives (http://prescriptives.com).   
• Oversaw front-end implementation work for both sites (Javascript, jQuery, and CSS).

**CTO/FOUNDER**

**Kostizi, Inc**

Salt Lake City, UT

2010 to November 2012

• Directed all technology / strategic development with five direct reports and an extended staff of 10. Led design, production and operations of all aspects of the firm's technologies and web offering.   
• Secured $700,000 in angel investment by developing a technology and business growth strategy with value proposition that defined market need and how to fill it - positioning firm for launch.   
• Bolstered revenues from $0 to $2+ million annually over two-year period by building and expanding user base from 400 users to 12,000+ users and merchant offers from 30 to more than 400.   
• Enhanced firm's value proposition and attractiveness to investors by successfully filing 3 software patents for less than $8,000, saving $27,000 over original quotes by reducing counsel's time demands.   
• Expanded one media partner in one market to 9 media partners in 11 markets in 18-month period by envisioning product suites and authoring B2B, promotional, and website materials.   
• Created and installed first-of-its kind web-based operating system within 6 weeks of launch to enable 50+ merchants to track redemptions electronically, while saving them paperwork and costs.

**CONSULTANT/FOUNDER**

**Jabberwolf, Inc**

Franklin, TN

2008 to 2010

• Supervised all aspects of a new FOREX trading services company from developing trading algorithms, a website, and marketing plan to overseeing sales efforts.   
• Utilized strategic positioning to enable the firm to offer competitive FOREX trading services to customers by developing and implementing 5 viable and profitable FOREX trading algorithms targeting a widely-used retail trading platform.

**CHIEF MARKETING OFFICER**

**Digital Reasoning Systems**

Brentwood, TN

2005 to 2007

• Governed all product marketing and commercial strategic analysis for startup technology firm. Supervised two reports.   
• Pinpointed a new viable $2 billion global market by developing a strategic growth plan to establish a solid commercial direction for a small government contractor.   
• Increased revenue from $100,000 to $2.5+ million annually by designing marketing campaign for company and its flagship product that included coaching and mentoring field sales staff to deliver results.   
• Procured $100,000 payment by crafting a strategic plan to launch company's largest product (Synthesys SDK) as part of satisfying requirements for existing government contract.

**CONSULTANT**

**MKG Ventures**

New Albany, MS

2003 to 2005

• Counseled small businesses across the region in strategic development including business valuation, strategic analysis, documentation, and due diligence.   
• Obtained $4 million angel investment within 6 months by creating a strategic business plan to secure capital funding for a cold forging manufacturing startup in Mississippi.   
• Wrote a textbook based on research on distributed computing, published by Prentice Hall in 2004. The textbook is currently used by at least 5 different computer science courses around the world today.

**JAVA TECHNOLOGY EVANGELIST**

**Sun Microsystems**

Palo Alto, CA

1994 to 2002

• Built critical mass of support for java platform, establishing it as technical standard in global marketplace.   
• Filed 3 software patents through Sun Labs while supporting company's marketing organization. Reached more than 100,000 Software Developers worldwide by proposing and supporting efforts to produce Sun's "Technology Days" events that covered new software products / protocols.   
• Contributed to boosting number of active Java Software Developers worldwide from 50,000 to 5+ million in a key role by writing numerous published articles and blogs advocating java technology to targeted Software Developers.

## Education

**Master of Science in Accountancy**

**University of Phoenix**

Phoenix, AZ

2010

**Doctor of Philosophy (Ph.D.) in Theoretical Computer Science**

**City University of New York**

1998 to 1999

**Master of Business Administration in E.M.B.A**

**University of San Francisco**

1992

**Bachelor of Science in Computer Science**

**Fordham University**

New York, NY

1982

## Skills

JAVA (10+ years), APACHE HADOOP MAPREDUCE (3 years), DATA SCIENCE (3 years), Hadoop (2 years), HADOOP (3 years)

## Links

<https://github.com/maxgoff>

## Additional Information

Technical Proficiencies: Big Data and Data Science technologies and platforms; major business computer applications and platforms; software- and web-development methodologies; major operating environments (Windows, UNIX, Linux, Apple OSX, MacOS), HTML5, Javascript, jQuery, PHP, Perl, J2EE, Java, C/C++, MySQL, NoSQL, MS SQL, Hive, Meta Trader 4, Java, R, R Studio, python, Scala, PHP, PhotoShop, Microsoft Office Suite (Excel, PowerPoint, Word), MVC, CSS3, Hadoop, MapReduce, NiFi, bash

# Indeed.com-13

# Data Scientist

## Data Scientist, Manager, IT Architect

Middleboro, MA

A long-standing passion for the collection, exploration, and analysis of digital data and for the design and architecture of the systems used to store and analyze such data has characterized my diverse career. I have managed a wide range of successful projects in this area and written software for this purpose. I have twice hired, mentored, and supervised teams of data professionals, a team of six datascientists at an educational management company and a team of five system and database administrators at a major business school. I am seeking a position in analytics and data science and enjoy leading new efforts in this area.

Authorized to work in the US for any employer

## Work Experience

**Data Scientist**

**McGraw-Hill Education**

Boston, MA

January 2016 to January 2018

I was a senior member of the data science team and I was also responsible for the design and selection of the computing infrastructure used for data science and analytics, including both hardware and cloud-based resources (Databricks).    
   
Most recently my focus has been on spaced practice and adaptive learning and on using data scienceto enhance student performance in those areas. I was part of a team which developed a mobile adaptive learning product, including developing the data API for that product and doing all of the analytics on the the resulting data. This included doing both the business analytics and the learning analytics for this suite of products.   
   
I lead a successful effort to implement an automated essay evaluation system using NLP. I was the architect of the data infrastructure for this system, which used lambda architecture.

**Senior Director of Research Computing and Data Science**

**Data Science group**

January 2014 to January 2016

• Recruited, managed, mentored the Data Science group (6 Data Scientists)   
• Specified and selected computing infrastructure for Data Science and Big Data including dedicated hardware and cloud services (AWS and Databricks)   
• Managed project to implement MongoDB-based system for an MHE product   
• Managed project to implement a system for essay evaluation in AWS using lambda architecture which included Kinesis, HDFS, and Postgres   
• Designed the algorithm for an adaptive flashcard system based on learning science research   
• Designed the data schema for messaging-based analytics data collection   
• Analyzed data to measure effectiveness of different question modes in an adaptive flashcard system using BKT and PFA   
• Five Peer-reviewed conference presentations   
• Author of chapter in book on Frontiers in Cyberlearning

**Associate Director of Infrastructure, Operations and Security**

**MIT Sloan School of Management**

Cambridge, MA

2005 to 2014

• One of five directors who administered Sloan Technology Services   
• Responsible for the server and storage infrastructure, including systems architecture, vendor management, purchase, and operation for 300 hardware and VMWare virtual servers, EMC SAN storage, and VMWare virtual desktops   
• Had eight years with no unscheduled outages in the 300+ systems managed   
• Hired and managed a team of five system and database administrators   
• Designed, selected, purchased, installed, and operated the research computing system for the Sloan School

**Unix Systems Administrator/Oracle DBA**

**EMC SAN storage, and VMWare**

2000 to 2005

• Linux and Unix system administrator   
• Oracle database administrator of 10 Oracle databases

## Education

**Ph.D. in (Earth Sciences) Scripps Institution of Oceanography**

**U. of California**

San Diego, CA

**MS in (Earth Sciences) Scripps Institution of Oceanography**

**U. of California**

San Diego, CA

**Physics**

**Princeton University, Princeton**

## Skills

Data Science (4 years), databases (5 years), MongoDB (2 years), Oracle (5 years), Apache Spark (3 years), Databricks (2 years), Unix Administration (10+ years), Python (4 years)

## Certifications/Licenses

**edX Verified Certificate for Distributed Machine Learning with Apache Spark**

edX Course on Apache Spark using the Databricks notebook environment.

**edX Verified Certificate for Big Data Analysis with Apache Spark**

edX Course on using Apache Spark for Big Data analysis using the Databricks notebook environment.

**edX Verified Certificate for Introduction to Apache Spark**

Introductory course on Apache Spark using the Databricks notebook enviromment.

**edX verified certificate in Big Data and Education**

edX course on using Big Data in Education, including clustering, classification, Bayesian Knowledge Tracing, and Logistic regression.

**Oracle Certified DBA**

January 2000 to January 2004

**GIAC Certified Unix Security**

March 2003 to March 2007

## Additional Information

Areas of Expertise   
• Adaptive learning systems, including algorithm development, knowledge tracing, and question mode evaluation   
• Logistic regression and hidden Markov modeling (BKT and PFA)   
• Linear regression and non-linear least squares   
• Time series analysis and spectral analysis   
• Apache Spark (Databricks) for parallel computing, particularly Spark SQL   
• Design and selection of computing systems for analytics and research   
• Data analysis with SQL, Spark SQL, Python, and PySpark   
• Hiring, supervising, and mentoring of Data Scientists, particularly for education   
• Extensive project management experience in Data Science, IT, and teaching   
• Relational and NoSQL databases, including Oracle, Postgres and MongoDB   
• Clear oral and written communicator including eight years teaching college physics   
• Ability to learn new technologies and academic material quickly   
• Design, specification, selection, purchase, installation, and administration of computing and storage for enterprise and analytics systems   
• Management of sensitive and confidential data, including student data   
• Programming in SQL, Python, PySpark, PL/SQL, Fortran, Java, MATLAB, shell script, C, HTML, XML, LaTex

# Indeed.com-14

# Hadoop Big Data Engineer / Data science

## Hadoop Big Data Engineer

Plano, TX

Sponsorship required to work in the US

## Work Experience

**Hadoop Big Data Engineer / Data science**

**Global Data Mart Inc**

Plano, TX

July 2015 to Present

" Extensively used Pig for data cleansing.   
" Created partitioned tables in Hive.   
" Managed and reviewed Hadoop log files.   
" Involved in creating Hive tables, loading with data, and writing hive queries which will run internally in MapReduce way.   
" Used Hive to analyze the partitioned and bucketed data and compute various metrics for reporting.    
" Developed Pig Latin scripts to extract the data from the web server output files to load into HDFS.   
" Developed queries based on the business analysis using Hadoop ecosystems such as PIG or HIVE.   
" Developed the Pig UDF'S to pre-process the data for analysis.    
" Created Scripts for Apache Falcon to automate derivation and send to different clusters.   
" Creating Hive tables and working on them using Hive QL.   
" Analyzed large amounts of data sets to determine the optimal way to aggregate and report on it.   
" Handled importing of data from various data sources, performed transformations using Hive, MapReduce, and loaded data into HDFS.   
" Ingesting the data into the lake.   
" I have created Bridge files (Universal mobility bridge file and Address Bridge file)   
" Publishing data to data router.   
" Maintenance of the ingestion tool and writing scripts for automation.   
" Coordinating audit activities based on the requirement to meet compliance.   
" Developing data for business such as internal threat analysis.   
" Ingesting data and developing a tool for automation.   
" Identify, analyze, and interpret trends or patterns in complex data sets.   
" Profiling the data.   
" Analyze data network traffic on 4G, 3G & 2G and predict which all towers need to be upgraded or downgraded.   
" Create profile of towers.   
" Cluster towers based on volume to identify the behavior of the tower.   
" Analyze browsing behaviors on-hand devices for marketing and advertising based on location-based algorithms and highly frequently visited websites.   
" Creating scrubbing logic for business to exclude the Opted -Out millions of customers from advertisement to meet the compliance requirement.   
" Analyzing data and predicting results using Machine Learning and Artificial Intelligence.   
" Saved 1+ billion dollars in tower upgrade.   
" Saved 200+ million in internal threat analysis.   
" Saved 300+ million in tethering threat analysis.

**Kiran Electronics Company**

**Production Engineer**

Bangalore, Karnataka

July 2012 to August 2013

" Developed a dynamic and interactive website that ensured high traffic page views and User Experience, resulting in an increase in sales.   
" Fixed bugs from existing websites and implemented enhancements that significantly improved web functionality and speed.   
" Used Netbeans IDE to develop the project   
" Create, Update and delete accountants detail after login   
" Can a search branch wise accounts.   
" Generation of customer ID

**Intern**

**BOSCH**

Bangalore, Karnataka

January 2012 to May 2012

" Researched on the sensors single and testing the National Instrument cards   
" For the acquiring and design purpose used LabView and Matlab   
" The main program for the working on the design was written in JAVA   
" The design was installed and test the results were successful

## Education

**PhD. in Information Technology**

**University of Cumberland's**

Williamsburg, KY

May 2018 to Present

**Master's in computer and Information Security in Systems**

**University of Cumberland's**

Williamsburg, KY

August 2016 to April 2018

**Master's in electrical engineering in electrical engineering**

**University of Dayton**

Dayton, OH

August 2013 to May 2015

**Bachelors in Telecommunication in Telecommunication Engineering**

**Visweshwaraiah Technological University**

Belgaum, Karnataka

August 2008 to May 2012

## Skills

APACHE HADOOP HDFS (Less than 1 year), APACHE HADOOP MAPREDUCE (Less than 1 year), APACHE HADOOP OOZIE (Less than 1 year), APACHE HADOOP SQOOP (Less than 1 year), Hadoop (1 year), Data Science (2 years), Machine Learning (2 years)

## Links

<http://www.linkedin.com/in/tejasviv>

## Certifications/Licenses

**HDP Certified Developer**

February 2017 to Present

Hortonworks, Inc.

## Additional Information

Having 4+ Year of strong IT experience Designing, implementing, and supporting technical solutions. keen technical insights and a leadership record of accurate, timely business solutions that increase efficiency, cut costs and accommodate robust business expansion.   
• Has more than 2+ years of experience in implementing Big Data projects and Artificial Intelligence with Machine Learning using Hadoop and other open source tools/technologies.   
• Has a strong understanding of Distributed Storage and Parallel processing framework.   
• Has good knowledge on architecture, working principles, development, writing validation, datacleansing of all the Hadoop Ecosystem components like HDFS, YARN, MapReduce, Pig, Spark, Hive, HBase, Oozie, Sqoop etc.   
• Worked on bulk data ingestion of terabytes of data sets from external relational databases like Teradata and Oracle to Hadoop using SQOOP.   
• Excellent interpersonal skills backed up-by-sound professional ability to work independently as well as a team member.   
• Highly organized and dedicated to positive Attitude, possess good time management and organizational skills with the ability to handle multiple tasks with a positive attitude   
• Possess effective command over written, verbal communication and presentation skills,   
• Strong analytical, problem solving and troubleshooting skills, willingness, and ability to quickly adapt to new environments and learn new technologies   
• Has good experience in testing like unit, Integration, System, Case Planning, Design, and Execution.   
• Bachelors in Telecommunications, which covers all Networking, Telecommunication, and computers science courses such as C, Assembly level language, Programming Microprocessor and Microcomputer, Analog and Digital Signal Processing.   
• Achieved Master's degree Knowledge and skills in Computer Science in Machine Learning Algorithms, data and pattern analysis which were a part of the curriculum in University of Dayton, Dayton, Ohio, USA   
• Second master's in Computer Science and Information Security Systems, which covers all the compliance and audit and maintenance of servers, and systems.   
• Has a strong knowledge of analysis tools such as Tableau, Microstrategy, Zeppelin, Panda, Numpy, R, paramiko, matplot, Matlab, sci-kit, google core, Microsoft Azure   
• Has also worked on AGILE environment.   
   
SKILLS   
• Big Data Ecosystems: Hadoop, MapReduce, HDFS, HBase, Zookeeper, Hive, Pig, Sqoop, Oozie, Falcon, Spark   
• Traditional Ecosystems: Oracle, Teradata, Vertica, Informatica.   
• Analysis Tools: Tableau, Microstrategy, Zeppelin, Panda, Numpy, R, paramiko, matplot, Matlab, sci-kit, google core, Microsoft Azure, Tensorflow, Keras,   
• Programming Languages: C, C++, Java, Shell Scripting, SQL, and Assembly level Language   
• Scripting Language: JSP & Servlets, PHP, JavaScript, Python, Scala   
• Tools: Eclipse, JDeveloper, Scala IDE, IntelliJ   
• Platforms: Mac, Windows (7/8), Linux   
• Testing Tools: NetBeans, Eclipse, Scala IDE, IntelliJ.

# Indeed.com-15

# Technical Director of Data Science

## Technical Director of Data Science

Glendale, CA

Data science leader with over 15 years’ experience in predictive analytics and data science, utilizing industry-standard tools as well as the latest technologies/languages. Extensive strategic and hands-on experience shaping the role that data analysis plays in organizational business decisions – ranging from being a thought-leader of teams, getting management buy-in, and developing unique solutions to deploying newly architected systems that collect and analyze key business data. Known to be passionate about ensuring organizations are collecting data using the appropriate techniques, as well as advocating for the utilization of data in new and unique ways to solve particular business challenges. Possesses a strong collaborative management and leadership style, with distinct skill at improving relationships and communication among team members, while also developing team members’ strengths and advancing their professional goals.

Authorized to work in the US for any employer

## Work Experience

**Technical Director of Data Science**

**SYMANTEC**

2015 to 2015

Provided strong technical direction to a new Data Science team, including:   
   
• Introduced and implemented a Big Data analytics platform   
• Developed custom algorithms to optimize sales offers within the support channel, combining machine learning and statistics.   
• Devised an original recommendation algorithm for website to drive customer retention   
• Provided technical mentorship and direction to team members

**Director of Analytics**

**SYMANTEC**

2013 to 2015

Led team of fifteen analysts, providing business intelligence and analytics, as well as maintained the data warehouse and interactive dashboards (including data from Adobe Analytics):   
   
• Restructured the team from a flat organization, with no sub-teams, into five groups, aligned with the business, each led by a single-point-of-contact   
• Integrated two new functions into the group: Data Warehouse Architecture and Predictive Analytics, including attracting high-profile new hires from top-tier universities   
• Led the development of a custom algorithm to better direct customers through an existing retention flow, resulting in saving approximately $500k/year with virtually zero increase in customer churn

**Optimization Engineer and Data Warehouse Architect**

**TALUS LABS**

Los Angeles, CA

2012 to 2013

Led development of custom bid and budget optimization algorithms for a Search Engine Marketing platform serving thousands of SMB accounts; devised and implemented data warehousing strategy and execution   
   
• Developed original Search Engine Marketing (SEM) bid/budget optimization engine, using statistical modeling combined with portfolio optimization   
• Led major data warehouse development initiative, including design, architecture, and tooling

**Associate Director of SEM Bid Optimization and Performance**

**AT&T INTERACTIVE**

2008 to 2012

Responsible for Search Engine Marketing bid optimization system, setting bids on thousands of SMB accounts   
   
• Developed original Search Engine Marketing (SEM) bidding & budgeting algorithm that optimizes individual account lead goals, while satisfying portfolio-level budget constraints   
• Performed statistical analysis of performance data, as it related to churn (logistic regression)   
• Statistical modeling to predict which media sources drove the most leads for businesses   
• Conducted a full-scale code and algorithm review of the entire code base of the above tool (in order to continue to enhance bidding algorithms and add features, such as the ability to manage bids on more sources)   
• Developed a Data Warehouse in SQL Server from the ground up (which allowed the group to move from Excel-based reporting on thousands of rows of data to fully automated SQL reports on hundreds of millions of rows)

**Data Analyst**

**AT&T INTERACTIVE**

2007 to 2008

Conducted daily and monthly audits of traffic data for YP.com site and all partner and extended distribution sites   
   
• Investigated sources of traffic anomalies by drilling down from summary to detail tables using custom SQL   
• Analyzed website traffic patterns and wrote a daily brief summary of trends and anomalies for executives   
• Developed custom Excel spreadsheets for business users with dynamic data capabilities

## Education

**M.S. in Clinical Psychology**

**Florida State University**

Tallahassee, FL

**B.A. in Psychology**

**University of Missouri**

Columbia, MO

## Skills

Data Analysis (10+ years), Database Management (10+ years), Data Mining (10+ years), Databases (10+ years), Statistical Analysis (10+ years), Statistics (10+ years), Excel (10+ years), SQL (10+ years), Sql Server (10+ years), PostgresSQL (2 years), Data Warehousing, Data Warehousing (10+ years), Data Warehouse (10+ years), Analysis, Analysis (10+ years), Analytical (10+ years), Analytics (10+ years), Analytical Skills (10+ years), Office Suite (10+ years), Reporting (10+ years), R (2 years), Python (2 years), Research (10+ years), Management (3 years), Management Skills (3 years), Manager (3 years), Management Experience (3 years), Spss (10+ years), Predictive Analytics (10+ years), Data Science (10+ years)

## Links

<http://about.me/scottedwards2000>

[http://linkedin.com/in/scottedwards2000](https://www.indeed.com/url?q=http%3A%2F%2Flinkedin.com%2Fin%2Fscottedwards2000&h=d0b2a4f3)

[http://twitter.com/scottedwards200](https://www.indeed.com/url?q=http%3A%2F%2Ftwitter.com%2Fscottedwards200&h=f05be8ae)

# Indeed.com-16

# Data Science Analyst

## Looking for Summer Internship in the field of Data Analytics | Data Science

Apt, AR

Authorized to work in the US for any employer

## Work Experience

**Data Science Analyst**

**Infosys, India**

Tampa, FL

February 2014 to July 2017

Worked for one of the US based health care provider, CareFirst Inc.   
The objective of this program was to reduce administrative complexity,   
improve efficiency, reducing costs and ultimately improving the quality.   
Interpret data from primary and secondary sources using statistical   
techniques using statistical tool R and provide ongoing reports.   
Implemented business process, database retrievals, access of information.   
Installed, configured and hosted the Tomcat app servers and MySQL DB   
servers on physical servers (Linux, Mac, Windows), and Amazon AWS virtual   
servers (Linux).   
Developed Complex database objects like Stored Procedures, Functions,   
Packages and Triggers using SQL.   
Building, publishing customized interactive reports and dashboards, report   
scheduling using Tableau server.   
Compile and validate data; reinforce and maintain compliance with   
corporate standards.   
Developed optimized data collection and qualifying procedures   
using Python.   
Leveraged analytical tools to develop efficient system operations.

**Software Programmer Intern**

**Xonware Technologies Pvt. ltd**

September 2012 to January 2014

Worked for a Banking firm to develop a product that supports handling their   
delinquent customers at different levels. The product also depicted the   
customer relationship with the bank.   
Successfully interpreted data to draw conclusions for managerial action and   
strategy.   
Used Oracle SQL server to manage the database. Involved in writing unit test   
cases.   
Involved in creating database objects like tables, views, procedures, triggers,   
and functions using PL/SQL to provide definition, structure and to maintain   
data efficiently.   
Designed and developed stored procedures, queries and views necessary to   
support SSRS reports.   
Worked with Java as Back-end.   
Created use case diagrams to explain the flow of events and how different   
software systems interact with each other.   
Conducted workflow, process diagram and gap analysis to derive   
requirements for existing systems enhancements.

## Education

**MASTERS in BUSINESS ANALYTICS AND INFORMATION SYSTEM**

**University of South Florida**

Tampa, FL

August 2017 to December 2018

**BACHELORS in COMPUTER SCIENCE**

**University of Pune**

Pune

August 2008 to May 2012

## Skills

Data Science (2 years), Java (2 years), MS SQL SERVER (2 years), Python. (2 years), SQL (2 years)

## Links

<http://www.linkedin.com/in/Aishwarya-Gaikwad>

## Additional Information

Tools: Toad, WEKA, SQL Server Management Studio, Jupyter(Anaconda), SSRS, Eclipse, R Studio, SAS, Tableau.   
Languages: R, Python, SQL, Java (Seed certification in Core and Advance Java), JavaScript, C , C++.   
Packages and Libraries: Keras (RNN), Scikit-Learn, TensorFlow, Numpy, Matplotlib, Pandas.   
Operating System: Unix, Windows.   
Databases: Oracle, Mongo DB, SQL Server, Redshift.   
Coursera Certifications in Python for Data Science and Machine Learning, Business Statistics and Analysis, R.   
Well versed in English and has an excellent communication skills.

# Indeed.com-17

# Data Scientist

## Data Scientist - Sykes Enterprises, CT

Sunnyvale, CA

## Work Experience

**Data Scientist**

**Sykes Enterprises, CT**

December 2017 to Present

• Deep Learning: Predicted type of customer service request by implementing deep learning model, on call center telemetric data to reduce an employee's average request handling time when assisting customers   
• Recommender Systems: Built and deployed a recommender system on Microsoft Azure to reduce agent's average handle time of issue and increased the number of calls answered in a quarter by 5%   
• People Analytics: Planned strategic workforce scaling by implementing employee attrition (Microsoft Azure, SQL, and R) predictive model; suggested recommendations to reduce effect of job abandonment by 10%

**Data Science Consultant**

**Limra International Insurance**

July 2017 to December 2017

• Insurance Analytics: identified trends that affect length of service (LOS) of agents and estimated LOS using Linear regression; suggested business proposals that increased LOS   
• Predicting Insurance Agent Churn: Identified the factors impacting agent churn by building logistic regression model and provided insights to develop agent retention programs that decreased churn by 5%   
• Consumer Analytics: Built customer retention prescriptive models for subscription based products using advanced machine learning algorithms. This resulted in decreasing customer churn by approximately 8%

**Data Science Analyst**

**IBM**

January 2015 to December 2016

• Retail orders Analytics: Created classification model (Logistic Regression, SQL, and R) to analyze customer complaints on incorrect orders and performed Root Cause Analysis; reduced incorrect orders by 75%   
• HR Analytics: Predicted the absenteeism of employees by modelling employee behavior data with classification model, assisted the management to reduce the absenteeism   
• Software Engineering: Delivered Financial Solution to fix incorrect monthly bill cycle and correctly billed approximately $1 million towards client and won recognition from Subject Matter Expert   
• Dashboarding: Designed interactive dashboards in Tableau to track Key Performance Indicators (KPIs) that assist in meeting service level agreement; proposed business strategies leading to increasing of team performances   
• Reporting: Extracted data (SQL), prepared insightful reports and identified performance hindering metrics; minimized percent of aged incidents by 60% in 4 months

## Education

**Master in Business Analytics and Project Management**

**University of Connecticut School of Business**

May 2018

## Skills

SQL (2 years), DATA SCIENCE (2 years), CLUSTERING (Less than 1 year), DATA MANIPULATION (Less than 1 year), DATA MINING (Less than 1 year)

## Additional Information

SKILLS   
Data Science: Classification, Regression, Time Series Forecasting, Machine Learning, Data mining, Feature Engineering, Data Scraping, Data Manipulation and Visualization, Text mining, Clustering, Support Vector Machine, Random Forest, A/B Testing, Market Basket Analysis, Natural Language Processing   
Technical Skills: SQL, R, Python (NumPy, Pandas, Scikit), SAS, JMP, Tableau, MS Excel, Hive, Azure

# Indeed.com-18

# Senior Data Science Analyst

## Senior Data Science Analyst - CVX TECHNOLOGIES

New Jersey

Microsoft and Sollers Certified Data Science professional with 6 years of experience in all phases of diverse technology projects specializing in Data Science, Python, SQL Server and Tableau. Team builder with excellent communications, time & resource management & continuous client relationship development skills.   
   
● Extensive experience in Data Mining, Machine Learning and Data Science technologies ensuring project completion on time, on budget and with the desired results. Highly creative, innovative, committed, intellectually curious, business savvy with good communication and interpersonal skills.   
   
● Analytical, performance-focused, and detail-oriented professional, offering in-depth knowledge of data analysis and statistics   
● Equipped with experience in utilizing statistical techniques which include Correlation, Hypotheses modeling, Inferential Statistics as well as data mining and modeling techniques using Linear and Logistic regression, clustering, decision trees, and k-mean clustering   
● Expertise in building Supervised and Unsupervised Machine Learning experiments using Microsoft Azure utilizing multiple algorithms to perform detailed predictive analytics and building Web Services models for all types of data: continuous, nominal, and ordinal.   
● Expertise in using Linear & Logistic Regression completed multiple projects applying these techniques   
● Mitigated risk factors through careful analysis of financial and statistical data. Transformed and processed raw data for further analysis, visualization, and modeling.   
● Increased customer visibility by developing real-time insights to sales people and sales managers using Tableau reports, resulting in boosting the revenue by 10%   
● Proficient in research of current process and emerging technologies which need analytic models, data inputs and output, analytic metrics and user interface needs.   
● Good Knowledge in Proof of Concepts (PoC's), gap analysis and gathered necessary data for analysis from different sources, prepared data for data exploration using data munging.   
● Good industry knowledge, analytical &problem solving skills and ability to work well with in a team as well as an individual.

Authorized to work in the US for any employer

## Work Experience

**Senior Data Science Analyst**

**CVX TECHNOLOGIES**

June 2016 to Present

● Acquire, clean using Talend and structure data from multiple source including external and internal databases   
● Perform data extraction, Manipulation, analysis and data mining using SQL   
● Perform data extraction, manipulation, cleaning, analysis, modeling and data mining using R programming in R Studio   
● Designed 10+ dashboards in Tableau for sales managers with instant access to personalized analytics portal, so they can access key business metrics such as time to close opportunity, delay-to-contract, resulting in increased customer satisfaction and improving client’s standing in the Sales Performance management industry   
● Develop and execute processes for accurate data capture across all clients to obtain key insights and relationships to overall business objectives using Statistical Hypotheses Modeling   
● Exploratory data analysis using R to deep dive into internal and external data to diagnose areas of improvement to increase efficiency   
● Utilized Booted Decision Tree, Linear and Bayesian Linear Regression Machine Learning models in Microsoft Azure to develop and implement interactive Webservice predictive models    
● Lead the company's machine learning and statistical modeling effort including building predictive models and generate data products to support customer segmentation, product recommendation and allocation planning; prototyping and experimenting ML/DL algorithms and integrating into production system for different business needs.    
● Designed, built and deployed a set of R modeling APIs for customer analytics, which integrate multiple machine learning techniques for various user behavior prediction (CLTV, marketing funnel propensity models etc.) and support multiple marketing segmentation programs.    
● Built models using decision trees, segmentation, regression and clustering intelligent decision models to analyze customer response behaviors, interaction patterns and propensity

**Senior Data Analyst**

**Nationwide, OH**

June 2014 to July 2016

● Performed data analysis by gathering, analyzing and deploying customer centric models from its pristine form to derive insurance related conclusions.   
● Perform Data extraction, Manipulation and Retrieve data from Data base   
● Used Tableau to publish visualizations.   
● Developed category segmentation using R which provides customizable view of market share and led to decreased labor cost by 50%.   
● Documented Functional / Nonfunctional requirements, process workflows, data maps, table layouts and their validations.    
● Actively pursued data quality compliance, assessed risk factors and generated models and scenarios for forecasting operational risk.   
● Administered cross-training sessions with other departments to improve workflow and streamline project completion.   
● Analyzed Learning Management System (LMS) for client and implemented improvements for process variation thereby reducing lead time by 30%.    
● Developed a dashboard using Tableau to provide the management with an overall understanding of resource optimization, attaining incremental revenue worth $1.2 million

**Data Analyst**

**Global Logic Technologies, NJ**

January 2013 to June 2014

● Acquire, clean using Talend and structure data from multiple source including external and internal databases   
● Perform data extraction, Manipulation, analysis and data mining using SQL   
▪ Developed R programs for manipulating the data reading from various Oracle data sources and consolidate them as single CSV File, and update the content in the database tables   
▪ Created monthly and quarterly business monitoring reports by writing complex SQL queries to include System Calendars, Inner Joins and Outer Joins to retrieve data from multiple tables   
▪ Designed easy to follow visualizations using Tableau software and published dashboards on web and desktop platforms   
▪ Worked on new ideas to develop new mapping techniques and faster the work and predicted the stats based on Severity and Quality.   
▪ Created numerous processes and flow charts using Visio to meet the business needs and interacted with business users to understand their data needs   
▪ Monitored dashboards & application performance.   
▪ Made recommendations to management for the coordination of the daily workflow of the mapping stage and establish standard performance benchmarks for the timely processing of core mapping stages.   
▪ Involved in analysis, design and documenting business requirement specifications so as to build datawarehousing extraction programs, end-user reports and queries    
▪ Worked closely with Associates to find the problems and getting solutions on the tool.

**Data Analyst**

**Wise Financial, NYC**

May 2010 to December 2012

● Acquire, clean using Talend and structure data from multiple source including external and internal databases   
● Perform data extraction, Manipulation, analysis and data mining using SQL   
   
λ Consolidated & visualized million row of data set and developed complex queries/scripts on SSIS, T-SQL, Excel Macro to perform ETL and data analyses   
λ Utilized exploratory & confirmatory data analytics to investigate classified/clustering data sets on daily basis.    
λ Helped team director maintain coefficients/parameters and performed ad hoc analyses like multivariate regression analysis, time series & machine learning techniques to build business intelligence   
λ Maintained risk factors/coefficient models, adjusted risk testing scales and back-tested trading/portfolios data on VAR models to manage quantitative risks on R, SAS, Matlab, Mathematica

## Education

**BS in ELECTRICAL ENGINEERING**

**Yaba College of Technology**

May 2002 to December 2008

## Skills

DATA SCIENCE (5 years), CLUSTERING (1 year), DATABASES (5 years), MACHINE LEARNING (1 year), ACCESS (Less than 1 year), SQL (5 years), Tableau (5 years)

## Additional Information

Core Competencies   
   
➢ Microsoft Certification in Essentials of Data Science   
➢ Descriptive and inferential statistics   
➢ Linear and logistic regression models   
➢ Machine Learning - K-Mean Clustering, Principal Component Analysis (PCA), Factor Analysis, Time Series   
   
Programming and Tools   
➢ R Programming, R Studio   
➢ Python   
➢ Azure Machine Learning   
➢ Tableau   
➢ Talend   
➢ Github   
   
IT Project Lifecycle:   
➢ Requirements Analysis   
➢ BA Analysis and Monitoring   
➢ Requirements Elicitation   
➢ Project Scheduling   
➢ Testing/QA   
   
Databases   
➢ MS Access   
   
Value-Added Leadership:   
➢ Cross-Functional Supervision   
➢ Team Building & Mentoring   
➢ Client Relations & Presentations   
➢ Business & IT Planning   
➢ Vendor Management

# Indeed.com-19

# Data Scientist

## Data Scientist - Sprint, Kansas

San Francisco, CA

• 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 working with data modeling tools like Erwin, Power Designer and ER Studio.   
• 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.   
• Experienced in Data Profiling, Analysis by following and applying appropriate database standards and processes, in definition and design of enterprise business data hierarchies.   
• 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 with 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, DataValidation in ETL   
• Experienced in creating and documenting Metadata for OLTP and OLAP when designing a systems.   
• Expertise in managing entire data science project life cycle and actively involved in all the phases of project life cycle including data acquisition, data cleaning, data engineering, features scaling, features engineering, statistical modeling (decision trees, regression models, neural networks, SVM, clustering), dimensionality reduction using Principal Component Analysis and Factor Analysis, testing and validation using ROC plot, K-fold cross validation and data visualization.   
• Excellent Knowledge of Ralph Kimball and BillInmon's approaches to Data Warehousing.   
• Expertise in synthesizing Machine learning, Predictive Analytics and Big data technologies into integrated solutions.   
• Extensive experienced in working with structured data using Hive QL, join operations, writing custom UDF's and experienced in optimizing Hive Queries.   
• 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 Rand python like ggplot2, caret, dplyr, Rweka, gmodels, RCurl, tm, C50, twitteR, NLP, Reshape2, rjson, plyr, pandas, numpy, seaborn, scipy, matplotlib, scikit-learn, Beautiful Soup, Rpy2.   
• Experienced using query tools like SQL Developer, PLSQL Developer, and Teradata SQL Assistant.   
• Excellent in performing data transfer activities between SAS and various databases and data file formats like XLS,CSV,DBF,MDB etc.   
• 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 datato/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.   
• Strong experience and knowledge in Data Visualization with Tableau creating: Line and scatter plots, Bar Charts, Histograms, Pie chart, Dot charts, Box plots, Time series, Error Bars, Multiple Charts types, Multiple Axes, subplots etc.   
• Excellent understanding and working experience of industry standard methodologies like System Development Life Cycle (SDLC), as per Rational Unified Process (RUP), AGILE Methodologies.   
• Experience in source systems analysis and data extraction from various sources like Flat files, Oracle […] IBM DB2 UDB, XML files.   
• Proficiency in SQL across a number of dialects (we commonly write MySQL, Postgre SQL, 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**

**Sprint, Kansas**

August 2017 to Present

Description: Sprint Corporation, commonly referred to as Sprint, is an American telecommunications company that provides wireless services and it an internet service provider.   
   
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 ApacheSpark 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.   
• 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.   
• 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 Map Reduce.   
• Let the implementation of new statistical algorithms and operators on Hadoop and SQL platforms and utilized optimizations techniques, linear regressions, K-means clustering, Native Bayes 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 AWSRedshift staging environment and Worked on migrating of EDW to AWS using EMR and various other technologies.   
• Worked on TeradataSQL queries, Teradata Indexes, Utilities such as Mload, Tpump, Fast load and FastExport.   
• 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 source 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 client.   
   
Environment: Erwin9.6.4, 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**

**Pariveda Solution**

Dallas, TX

May 2016 to July 2017

Description: Dallas serves as the original and largest Pariveda office, consistently grown over the years by leveraging our strong relationships within our community and delivering work that meets and exceeds our clients' expectations.   
   
Responsibilities:   
• Provide data architecture support to enterprise data management efforts, such as the development of the enterprise data model and master and reference data, as well as support to projects, such as the development of physical data models, data warehouses and data marts.   
• Create new data designs and make sure they fall within the realm of the overall Enterprise BI Architecture and Building relationships and trust with key stakeholders to support program delivery and adoption of enterprise architecture.   
• Implemented HQL Scripts in creating Hive tables, loading, analyzing, merging, binning, backfilling, cleansing using hive.   
• Used R, SQL to create Statistical algorithms involving Multivariate Regression, Linear Regression, Logistic Regression, PCA, Random forest models, Decision trees, Support Vector Machine for estimating the risks.   
• Used R for Exploratory Data Analysis, A/B testing, Anova test and Hypothesis test to compare and identify the effectiveness of Creative Campaigns.   
• Providing technical leadership, mentoring throughout the project life-cycle, developing vision, strategy, architecture and overall design for assigned domain and for solutions and responsible for the development of target data architecture, design principles, quality control, and data standards for the organization   
• Developed and maintains data models and data dictionaries, data maps and other artifacts across the organization, including the conceptual and physical models, as well as metadata repository   
• Performed extensive Data Validation, Data Verification against Data Warehouse and performed debugging of the SQL-Statements and stored procedures for business scenarios.   
• Used Spark Data frames, Spark-SQL, Spark MLLib extensively and developing and designing POC's using Scala, Spark SQL and MLlib libraries.   
• Working on a Map RHadoop platform to implement Bigdata solutions using Hive, Map reduce, shell scripting and Pig.   
• Worked with cloud based technology like Redshift, S3, AWS, EC2 Machine, etc. and extracting the data from the Oracle financials and the Redshift database.   
• Used Teradata OLAP functions like RANK, ROW\_NUMBER, QUALIFY, CSUM and SAMPLE.   
• Involved in designing and developing Data Models and Data Marts that support the Business Intelligence Data Warehouse.   
• Extensively used Aginity Netezza work bench to perform various DML, DDLetc operations on Netezza database.   
• Developed multiple MapReduce jobs in java for Data Cleaning and pre-processing analyzing data in PIG.   
• Worked on predictive and what-if analysis using R from HDFS and successfully loaded files to HDFS from Teradata, and loaded from HDFS to HIVE.   
• Designed the schema, configured and deployed AWS Redshift for optimal storage and fast retrieval of data.   
• NLTK, Stanford NLP, RAKE to preprocess the data, entity extraction and keyword extraction.   
• Transforming staging area data into a STAR schema (hosted on Amazon RedShift) which was then used for developing embedded Tableau dashboards   
• Developed SQL scripts for loading data from staging area to Target tables and worked on SQL and SAS script mapping.   
• Performed transformations of data using Spark and Hive according to business requirements for generating various analytical datasets.   
• Performed Multinomial Logistic Regression, Random forest, Decision Tree, SVM to classify package is going to deliver on time for the new route and Performed data analysis by using Hive to retrieve the data from Hadoop cluster, Sql to retrieve data from Oracle database.   
• Proposed the EDW data design to centralize the data scattered across multiple datasets and Worked on migrating of EDW to AWS using EMR and various other technologies.   
• Worked on the development of Data Warehouse, Business Intelligence architecture that involves data integration and the conversion of data from multiple sources and platforms.   
• Let the implementation of new statistical algorithms and operators on Hadoop and SQL platforms and utilized optimizations techniques, linear regressions, K-means clustering, Native Bayes and other approaches.   
• Created mapreduce running over HDFS for data mining and analysis using R and Loading & Storage data to Pig Script and R for MapReduce operations.   
• Worked on Teradata SQL queries, Teradata Indexes, Utilities such as Mload, Tpump, Fast load and Fast Export.   
• Used Metadata tool for importing metadata from repository, new job categories and creating new data elements and proficiency in SQL across a number of dialects (we commonly write MySQL, PostgreSQL, Redshift, SQL Server, and Oracle).   
   
Environment: Oracle 12c, SQL Plus, Erwin 9.6, MS Visio, SAS, Source Offsite (SOS), Hive, PIG, Windows XP, AWS, QC Explorer, Share point workspace, Teradata, Oracle, Agile, PostgreSQL, DataStage, MDM , Netezza, IBM Infosphere, SQL, PL/SQL, IBM DB2, SSIS, Power BI, AWS Redshift, Business Objects XI3.5, COBOL, SSRS, QuickData, Hadoop, MongoDB, HBase, Hive, Cassandra, JavaScript.

**Data Architect/Data Modeler**

**EDAC Technologies Corp**

Farmington, CT

January 2015 to April 2016

Description: EDAC Technologies Corporation, together with its subsidiaries, provides design, manufacturing, and services for tooling, fixtures, molds, jet engine components, and machine spindles in the aerospace and industrial markets.   
   
Responsibilities:   
• Architect and design, solutions for complex business requirements, including data processing, analytics and ETL and reporting processes to improve performance of data loads and processes.   
• Develop a high performance, scalable data architecture solution that incorporates a matrix of technology to relate architectural decision to business needs.   
• Participated in the design, development, and support of the corporate operation data store and enterprise data warehouse database environment.   
• 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 detail level.   
• Owned and managed all changes to the data models, Created data models, solution designs and data architecture documentation for complex information systems.   
• 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.   
• Analyze change requests for mapping of multiple source systems for understanding of Enterprise wide information architecture to devise Technical Solutions.   
• Worked on AWS Redshift and RDS for implementing models and data on RDS and Redshift.   
• Worked with SME's and other stakeholders to determine the requirements to identify Entities and Attributes to build Conceptual, Logical and Physical Data Models.   
• Provided data sourcing methodology, resource management and performance monitoring for dataacquisition.   
• Designed and implemented Near Real Time ETL and Analytics using Redshift database   
• Supported and followed information governance and data standardization procedures established by the organization. Documents reports library as well as external data imports and exports.   
• Prepared Tableau reports and dashboards with calculated fields, parameters, sets, groups or bins and publish on the server.   
• Developed mappings to load Fact and Dimension tables, SCD Type 1 and SCD Type 2 dimensions and Incremental loading and unit tested the mappings.   
• Extensively used Netezza utilities like NZLOAD and NZSQL and loaded data directly from Oracle to Netezza without any intermediate files.   
• Performed analysis of data sources and processes to ensure data integrity, completeness and accuracy.   
• Created a logical design and physical design in Erwin.   
• Enforced referential integrity in the OLTP data model for consistent relationship between tables and efficient database design.   
• Implemented Hive Generic UDF's to incorporate business logic into Hive Queries and Creating Hive tables and working on them using Hive QL.   
• Developed DataMapping, DataGovernance, and transformation and cleansing rules for the Master Data Management Architecture involving OLTP, ODS and generated ad-hoc reports using OBIEE.   
• Responsible for migrating the data and data models from SQL server environment to Oracle environment.   
• Analysis and designing the ETL architecture, creating templates, training, consulting, development, deployment, maintenance and support.   
• Created SSIS Packages which loads the data from the CMS to the EMS library database and Involved in data modeling and providing technical solutions related to Teradata to the team.   
• Build a real time event analytic systems using dynamic Amazon redshift schema.   
• Designed the physical model for implementing the model into Oracle 11g physical data base and Developed SQL Queries to get complex data from different tables in Hemisphere using joins, database links.   
• Wrote SQL queries, PL/SQL procedures/packages, triggers and cursors to extract and process datafrom various source tables of database.   
• Created Hive Tables, loaded transactional data from Teradata using Sqoop and created and worked Sqoop jobs with incremental load to populate Hive External tables.   
• Developed LINUX Shell scripts by using NZSQL/NZLOAD utilities to load data from flat files to Netezza database.   
• Worked with cloud based technology like Redshift, S3, AWS, EC2 Machine, etc. and extracting the data from the Oracle financials and the Redshift database.   
• Used Erwin to create logical and physical data models for enterprise wide OLAP system and Involved in mapping the data elements from the User Interface to the Database and help identify the gaps.   
• Designing and customizing data models for Data warehouse supporting data from multiple sources on real time. Requirements elicitation and Data analysis. Implementation of ETL Best Practices.   
• Generated comprehensive analytical reports by running SQLqueries against current databases to conduct data analysis.   
• Created data models for AWS Redshift and Hive from dimensional data models.   
• Developed complex SQL scripts for Teradata database for creating BI layer on DW for Tableau reporting.   
• Extensively used ETL methodology for supporting data extraction, transformations and loading processing, in acomplex EDW using Informatica.   
• Created Active Batch jobs to load data from distribution servers to PostgreSQL DB using \*.bat files and worked on CDC schema to keep track of all transactions.   
   
Environment: Erwin 9.5, MS Visio, Oracle 11g, Oracle Designer, MDM , Power BI, SAS, SSIS, Tableau, Tivoli Job Scheduler, SQL Server 2012, DATAFLUX 6.1, JavaScript, AWS Redshift, PL/SQL, SQL/PL SQl, SSRS, PostgreSQL, Data Stage, SQL Navigator Crystal Reports 9, Hive, Netezza, Teradata, T-SQL, Informatica.

**Data Architect/Data Analyst/Data Modeler**

**Dorman Products Inc**

Colmar, PA

May 2013 to December 2014

Description: Dorman Products, Inc. supplies automotive replacement parts, fasteners, and service line products primarily for the automotive aftermarket.   
   
Responsibilities:   
• Design and develop data warehouse architecture, data modeling/conversion solutions, and ETL mapping solutions within structured data warehouse environments   
• Reconcile data and ensure data integrity and consistency across various organizational operating platforms for business impact.   
• Define best practices for data loading and extraction and ensure architectural alignment of the designs and development.   
• Used Erwin for effective model management of sharing, dividing and reusing model information and design for productivity improvement.   
• Involved in preparing Logical Data Models/Physical Data Models.   
• Worked extensively in both Forward Engineering as well as Reverse Engineering using datamodeling tools.   
• Involved in the creation, maintenance of Data Warehouse and repositories containing Metadata.   
• Involved using ETL tool Informatica to populate the database, data transformation from the old database to the new database using Oracle and SQL Server.   
• Identifying inconsistencies or issues from incoming HL7 messages, documenting the inconsistencies, and working with clients to resolve the data inconsistencies   
• Resolved the data type inconsistencies between the source systems and the target system using the Mapping Documents and analyzing the database using SQL queries.   
• Extensively used both Star Schema and Snow flake schema methodologies in building and designing the logical data model in both Type1 and Type2Dimensional Models.   
• Worked with DBA group to create Best-Fit Physical Data Model from the Logical Data Model using Forward Engineering.   
• Worked with Data Steward Team for designing, documenting and configuring Informatica DataDirector for supporting management of MDM data.   
• Conducting HL7 integration testing with clients systems that is testing of business scenarios to ensure that information is able to flow correctly between applications.   
• Extensively worked with MySQL and Redshift performance tuning and reduced the ETL job load time by 31% and DW space usage by 50%   
• Developed Data Migration and Cleansing rules for the Integration Architecture (OLTP, ODS, DW).   
• Used Teradata SQL Assistant, Teradata Administrator, and PMON and data load/export utilities like BTEQ, Fast Load, Multi Load, Fast Export, and Tpump on UNIX/Windows environments and running the batch process for Teradata.   
• Created Dashboards on Tableau from different sources using data blending from Oracle, SQL Server, MS Access and CSV at single instance.   
• Used the Agile Scrum methodology to build the different phases of Software development life cycle.   
• Documented logical, physical, relational and dimensional data models. Designed the data marts in dimensional data modeling using star and snowflake schemas.   
• Created dimensional model based on star schemas and designed them using ERwin.   
• Performed match/merge and ran match rules to check the effectiveness of MDM process on data.   
• Carrying out HL7 interface unit testing aiming to confirm that HL7 messages sent or received from each application conform to the HL7 interface specification.   
• Used tools such as SAS/Access and SAS/SQL to create and extract oracle tables.   
• Data modeling and design of data warehouse and data marts in star schema methodology with confirmed and granular dimensions and FACT tables.   
• Developed SQL Queries to fetch complex data from different tables in remote databases using joins, database links and Bulk collects.   
• Enabled the SSIS package configuration to make the flexibility to pass the connection strings to connection managers and values to package variables explicitly based on environments.   
• Responsible for Implementation of HL7 to build Orders, Results, ADT, DFT interfaces for client hospitals   
• Connected to Amazon RedShift through Tableau to extract live data for real time analysis.   
• Developed SQL Queries to fetch complex data from different tables in remote databases using joins, database links and Bulk collects.   
• Developed Slowly Changing Dimensions Mapping for Type 1 SCD and Type 2 SCD andUsed OBIEE to create reports.   
• Worked on data modeling and produced data mapping and data definition specification documentation.   
   
Environment: Erwin, Oracle, SQL server 2008, Power BI, MS Excel, Netezza, Agile, MS Visio, Rational Rose, Requisite Pro, SAS, SSIS, SSRS, Windows 7, PL/SQL, , SQl Server, MDM, Teradata, MS Office, MS Access, SQL, SSIS, MS Visio, Tableau, Informatica, Amazon Redshift.

**Data Modeler/Data Analyst**

**Dabur India Ltd**

Ghaziabad, Uttar Pradesh

December 2011 to April 2013

Description: Dabur (Dabur India Ltd.) is India's largest Ayurvedic medicine & natural consumer products manufacturer.   
   
Responsibilities:   
• Designed logical and physical data models for multiple OLTP and Analytic applications.   
• Involved in analysis of business requirements and keeping track of data available from various datasources, transform and load the data into Target Tables using Informatica Power Center.   
• Extensively used the Erwin design tool &Erwin model manager to create and maintain the DataMart.   
• Extensively used Star Schema methodologies in building and designing the logical data model into Dimensional Models   
• Created stored procedures using PL/SQL and tuned the databases and backend process.   
• Involved with Data Analysis primarily Identifying Data Sets, Source Data, Source Meta Data, DataDefinitions and Data Formats   
• Performance tuning of the database, which includes indexes, and optimizing SQL statements, monitoring the server.   
• Developed Informatica mappings, sessions, workflows and have written Pl SQL codes for effective and optimized data flow coding.   
• Wrote SQL Queries, Dynamic-queries, sub-queries and complex joins for generating Complex Stored Procedures, Triggers, User-defined Functions, Views and Cursors.   
• Created new HL7 interface based on the requirement using XML, XSLT technology.   
• Experienced in creating UNIX scripts for file transfer and file manipulation and utilized SDLC and Agile methodologies such as SCRUM.   
• DataStage jobs were scheduled, monitored, performance of individual stages was analyzed and multiple instances of a job were run using DataStage Director.   
• Led successful integration of HL7 Lab Interfaces and used expertise of SQL to integrate HL7 Interfaces and carried out detailed and various test cases on newly built HL7 interface.   
• Wrote simple and advanced SQL queries and scripts to create standard and ad hoc reports for senior managers.   
• Involved in collaborating with ETL/Informatica teams to source data, perform data analysis to identify gaps   
• Used Expert level understanding of different databases in combinations for Data extraction and loading, joining data extracted from different databases and loading to a specific database.   
• Designed and Developed PL/SQL procedures, functions and packages to create Summary tables.   
   
Environment: SQL Server, UML, Business Objects 5, Teradata, Windows XP, SSIS, SSRS, Embarcadero, ER studio, Erwin, DB2, Informatica, HL7, Oracle, Query Management Facility (QMF), SSRS, Data Stage, Clear Case forms, SAS, Agile, Unix and Shell Scripting.

**Data Analyst/Data Modeler**

**Kotak Mahindra Bank**

Mumbai, Maharashtra

May 2009 to November 2011

Description: Kotak Mahindra Bank is an Indian private sector bank headquartered in Mumbai, Maharashtra, India. In February 2003, Reserve Bank of India (RBI) gave the licence to Kotak Mahindra Finance Ltd   
   
Responsibilities:   
• Developed Data Mapping, Data Governance and transformation and cleansing rules for the Master Data Management Architecture involving OLTP, ODS.   
• Created new conceptual, logical and physical data models using ERWin and reviewed these models with application team and modeling team.   
• Performed numerous data pulling requests using SQL for analysis and created databases for OLAP Metadata catalog tables using forward engineering of models in Erwin.   
• Enforced referential integrity in the OLTP data model for consistent relationship between tables and efficient database design.   
• Proficient in importing/exporting large amounts of data from files to Teradata and vice versa.   
• Developed Data Mapping, Data Governance, and Transformation and cleansing rules for the Master Data Management Architecture involving OLTP, ODS.   
• Identified and tracked the slowly changing dimensions, heterogeneous sources and determined the hierarchies in dimensions.   
• Utilized ODBC for connectivity to Teradata &MS Excel for automating reports and graphical representation of data to the Business and Operational Analysts.   
• Extracted data from existing data source, Developing and executing departmental reports for performance and response purposes by using oracle SQL, MS Excel.   
• Extracted data from existing data source and performed ad-hoc queries and used BETQ to run and Teradata SQL scripts to create physical data model.   
   
Environment: UNIX scripting, Oracle SQL Developer, SSRS, SSIS, Teradata, Windows XP, SAS datasets.

## Education

**Bachelor of Computer Science in Science & Technology**

**Informatica Power Centre**

## Skills

data science