Santosh Dighe
Technology Architect
E: sdighe1405@gmail.com ▪ M: +91 9004380181 ▪ Profile: https://santoshdighe.com

SYNOPSIS

Over 14 years of experience in various phases of IT Business Implementation related to Data and Business Intelligence, including Design, Product development, Strategy, General Management, Business Development, and Sales & Marketing. Expertise in executing large-scale big data, machine learning, and advanced analytics projects involving technology and cloud in varied industries, including Retail, Financial Services, Human Resources, and Insurance.

EXPERIENCE SUMMARY

- Experience in Business Development, Cloud, Data Transformation, Computer Vision, Machine Learning, and Leading and managing technology effectively. Worked in major corporates in the U.S. and successfully used technology for business benefit. Have had a successful track record in using technology to assist and transform Businesses in an efficient manner.
- End-to-end Machine learning/Deep learning pipeline implementation with data collection, data preparation/labeling, model building, training, testing, inference with web hosting using techniques and technologies like python, TensorFlow, Keras, PyTorch, OpenCV, and data augmentation.
- Expertise in implementing large-scale big data enterprise platform and analytics projects primarily using Cloudera CDH, Cloud Computing, Amazon Web Services, Machine Learning, Hadoop Ecosystem technologies, Apache Spark, Impala, Kafka, NoSQL MongoDB, and Linux technologies.
- Proximity marketing strategy using various location technologies like Bluetooth Low Energy (BLE) Beacons (Eddystone/iBeacon) and Geofencing, IoT, and retail and location analytics.
- Extensive expertise in implementing large-scale multi-dimensional data warehousing projects primarily using Big Data Hadoop and PeopleSoft EPM platform with IBM InfoSphere DataStage, Unix, and Oracle database in conjunction with various phases of Software Development Life Cycle (SDLC) model.
- Web Development & hosting with AWS, Web Service REST API, Apache HTTP & Tomcat, Flask, Django, VueJs, HTML, CSS, and Javascript.

CERTIFICATIONS/PROFESSIONAL DEVELOPMENT

Certification Program	Institute/Verified By	Month/Year
Neural Networks and Deep Learning	DeepLearning.AI/Coursera	Feb 2019
Big Data Using Hadoop	DePaul CDM IPD, Chicago, IL	Spring Quarter 2015
Data Science for Business	DePaul CDM IPD, Chicago, IL	Spring Quarter 2015
The Data Scientist's Toolbox	Johns Hopkins University/Coursera	Jan 2015

R Programming	Johns Hopkins University/Coursera	Feb 2015
Getting and Cleaning Data	Johns Hopkins University/Coursera	Mar 2015
IBM Big Data Fundamentals Technical Professional v1	IBM/Prometric	Sep 2013

EDUCATION

Degree/Diploma	University / Institute	Description	Year
PG Diploma in Advanced Computing (PG-DAC), C-DAC	Institute for Advanced Computing and Software Development, Pune, India	Center for Development of Advanced Computing, Advanced Computing Training School.	2006
B.E Computers	University of Pune, Maharashtra, India	Bachelor of Computer Engineering.	2004

TECHNICAL PROFICIENCIES

Type	Skills
Industries	Retail ▪ Finance & Banking ▪ Human Resource ▪ Insurance
Technologies	Deep Learning ▪ Python ▪ TensorFlow ▪ Keras ▪ PyTorch ▪ OpenCV ▪ Flask ▪ Cloudera CDH Enterprise ▪ Amazon Web Services (AWS) ▪ Hadoop Ecosystem ▪ Apache Spark ▪ Beacons ▪ IoT ▪ Impala ▪ Apache Kafka ▪ Apache Flume ▪ Mahout ▪ Sqoop ▪ Elasticsearch ▪ R ▪ Java ▪ HTML ▪ CSS ▪ Javascript ▪ Django ▪ VueJS ▪ Wordpress Web Designing ▪ Service Oriented Architecture (SOA) ▪ Oracle Service Bus ▪ IBM InfoSphere DataStage ▪ PeopleSoft EPM
Visualization/Modeling	Jupyter Notebook ▪ D3JS ▪ R ggplot ▪ Tableau ▪ Microsoft Visio Pro
Version Controlling	GitHub ▪ Microsoft TFS ▪ PVCS ▪ SVN
Database	MySQL ▪ NoSQL MongoDB ▪ PostgreSQL ▪ Oracle ▪ Microsoft SQL Server
Operating Systems	CentOS Linux ▪ Unix ▪ Mac ▪ Windows

WORK EXPERIENCE

Deep Learning Specialist (self-work)

Tasks:

- End-to-end implementation & execution of deep learning projects with data collection, data preparation/labeling, model building, training, testing, inference with web hosting.
- Setting up GPU Deep Learning infrastructure with NVIDIA CUDA Toolkit on Linux platform hosted on Amazon Web Services.
- Training Deep Learning Convolutional Neural Networks for image object detection, classification, and segmentation problems.
- Web hosting of trained models for real-time image object detection and classification using Python Flask web framework, HTML, CSS, and JavaScript.

Project Work:

1. **Fashion Items Detection & Classification**
   - Deep CNN Model in Python, Keras/TensorFlow
   - YOLOv3 based custom trained model
   - Category level items detection & classification

- Python flask framework for web hosting for real-time detection using web API

2. **Social Distancing Monitoring**
   - PyTorch with Yolov5
   - TensorFlow/Keras with Yolov3 & OpenCV
   - TensorFlow Object Detection Models (CenterNet/EfficientDet)

3. **Human Pose Detection**
   - Deep CNN for person detection
   - Pose Guided Person Image Generation
   - TensorFlow CenterNet HourGlass-512*512-kpts
   - Multi-Person Pose Estimation & Keypoints detection

4. **Image Data Augmentation for Computer Vision**
   - When enough training data is not available
   - Helps avoid overfitting
   - More data variety and volume
   - Flipping, Cropping, Rotating, sharpening, Embossing, Color Jittering, Edge Enhancement, etc.

5. **Content-Based Recommendation System**
   - Build historic data store in HDFS Hive/Impala from a relational database.
   - Data preparation from data store with item metadata.
   - Create similarity indexes using the Mahout RowSimilarity algorithm.
   - Map user/item integers to IDs, convert to Elasticsearch format.
   - Hadoop HDFS, Hive/Impala, Apache Spark, Mahout, Java, Python, & Elasticsearch.

Jeevan Sangharsh Foundation (part-time work)
Role: Co-Founder & Technology Lead May 2020 – Present

Description:
Jeevan Sangharsh is an online donation-based crowdfunding platform that enables people in India to raise funds for healthcare, education, disaster relief, and other personal causes with an easy-to-operate interface.

PinCarts Private Limited **Apr 2015 – Feb 2020**
Designation: Co-Founder

Description:
PinCarts is a technology-driven company with a vision to provide a seamless shopping experience to its valuable customers throughout their shopping journey using futuristic retail technologies like proximity marketing, personalization, advanced retail analytics, indoor navigation, contact-less checkout, and virtual and augmented reality. PinCarts' business model for brand retailers and shopping malls is meant to benefit the shoppers and retailers both. This technology helps with comprehensive cloud-based pre-sales and post-sales analytics and facilitates key decision-making.

Core Responsibilities:

- Strategic planning for the company's future growth, including business operations, marketing, and sales strategy development.
- Design and build efficient and secure big data, mobile, and web cloud infrastructure using Amazon web services to support multiple workloads.
- Build a centralized Enterprise Data Platform with data from discrete sources like real-time, near-real-time, and batch processes.
- Design, Develop and maintain advanced big data analytics and machine learning applications to support the company's vision of providing personalized and real-time location-based shopping experience to our valued customers and operational efficiency for retailers.
- Proximity marketing strategy using various location technologies like Bluetooth Low Energy (BLE) Beacons (Eddystone/iBeacon) and Geofencing.

Hexaware Technologies Inc. **May 2006 – Jul 2016**
Designation: Technical Architect

Project: Global Accounting Hub and Operational Data Store
Client: The Northern Trust Company (TNT), Chicago, IL, U.S.A
Work Location: Chicago, IL Role: SOA Designer May 2015 – May 2016

Description:
Global Accounting Hub is one of the strategic accounting projects under the Asset Servicing program. The goal of strategic accounting is to consolidate the posting of transactions from the various systems in use today into the new accounting system. The current technological state of accounting within the organization is complex and redundant with the direct point-to-point connections between different systems. This project will make accounting functionalities available as services and remove tight coupling accounting and its interfacing systems.

Enterprise data warehouse/operational data store is a Hadoop-based storage and processing management system. The purpose of this project is to isolate all the source systems from reporting requirements.

Environment:
Apache Hadoop, Apache Impala, Java, IBM CDC, IBM InfoSphere DataStage 11.3, UNIX, Oracle Service Oriented Architecture (SOA) suite, IBM Mainframes and DB2

- Implemented operational data store for accounting data in Enterprise Data Warehouse using Hadoop and Apache Impala/Hive metastore framework.
- Designed and implemented strategy for providing accounting data required for Wealth management system processing.
- Designed overall data transformation strategy for publishing the real-time transactions into message queue using IBM CDC and DataStage.
- Designed and Implemented service-oriented architecture for accounting business units.
- Integrated with upstream and downstream systems/applications with web services interface.
- This helped remove inter-system dependencies by defining ISO messaging standards.

DePaul Course Project: Recommender System **Apr 2015 – June 2015**
Course: Big Data Using Hadoop (Spring Quarter 2015, Chicago, IL, U.S.A)

- Implemented functionality to provide real-time, personalized product/item recommendations to customers using technologies Collaborative and Content Based Filtering, Cloudera CDH 5.x, Apache Spark, Apache Impala, MLlib, Mahout, Spark Streaming, Python, Flume, HDFS and Elasticsearch.
- Collaborative/used based filtering approach with Alternating Least Square algorithm for matrix factorization using Apache Spark MLlib and Python.
- Content/user-based filtering approach using Mahout Spark-Item/row similarity algorithms using Mahout/Spark technologies.

Project: Comprehensive Capital Analysis and Review (CCAR)
Client: The Northern Trust Company (TNT), Chicago, IL, U.S.A
Work Location: Chicago, IL Role: Automation Architect Oct 2014 – Mar 2015

Description:
The Comprehensive Capital Analysis and Review (**CCAR**) is an annual exercise by the Federal Reserve to ensure that institutions have robust, forward-looking capital planning processes that account for their unique risks and sufficient capital to continue operations throughout times of economic and financial stress.

Environment:
IBM InfoSphere DataStage 8.5, IBM InfoSphere Server Manager and ISTOOL, WinSCP, UNIX, Oracle 11g

- Designed overall automation strategy for production execution of all CCAR schedules, which ensures fully automated execution and rerun ability and email notifications.
- Implemented Business-As-Usual (BAU) ad-hoc data load process with fully automated execution mechanism using Unix shell scripts and DataStage dsjob utility.
- Designed and implemented Retail-Other schedule with 99% data accuracy in system integration testing with IBM InfoSphere DataStage 8.5 and Oracle database in Unix environment.
- Reduced time spent on quality assurance and decreased product defects by enhancing development procedures.

Project: Policy Administration System
Client: Trust Mark Insurance, Lake Forest, IL, U. S. A
Work Location: Lake Forest, IL Role: Data Modeler July 2014 – Sep 2014

Description:
The objective of this project is to follow the database guidelines to create a new relational database that will make all the data from the source data spreadsheets available to end-users. With this new relational database, we can standardize the data storage and normalize the table/entity relationships up to a certain extent to eliminate data redundancy and manipulation anomalies.

Environment:
Microsoft SQL Server 2008, Visio Professional 2013

- Ensued database guidelines to create a new relational database in Microsoft SQL Server that will make all the data from the source data spreadsheets available to end-users.

- Standardized data storage and normalized the table/entity relationships up to a certain extent to eliminate data redundancy and manipulation anomalies which helps automation test data users to analyze the input test data easily with simple SQL queries.
- Documented all business requirements (BRD) for logical data model preparation for future references.
- Prepared physical data model and DDL scripts for object creation using MS Visio professional.

Project: Global Data Warehouse (GDW)
Client: American Insurance Group, U. S. A
Work Location: Mumbai/Pune, India Role: Team Lead Nov 2011 – March 2014

Description:
This project was for implementing a global data warehouse system for the human resource department under Global Workforce Reporting & Analytics (GWR&A) initiative. GDW gets data from 4 source systems including domestic and international feeds and applies calculations to enable fact-based, consistent, and transparent analytical and operational reporting and other ad-hoc packages in order to facilitate decision-making.

Environment:
IBM InfoSphere DataStage 8.5, PeopleSoft EPM 8.9, UNIX, Oracle 11g, PeopleSoft Scheduler

- Delivered a First-of-a-kind multi-dimensional data warehouse for HR analytics and operational reporting with IBM InfoSphere DataStage and Oracle as a back-end database in Unix environment.
- Automated and scheduled data load processes using Unix shell scripts and PeopleSoft scheduler with interactive email notifications and error log distributions.
- Improved Data Quality and standardized definitions for metrics and data elements using common rules engine using DataStage routines.
- Streamlined data processing for downstream applications and security using the global security Model.

Project: Northern Financial Objective Reporting, and Management Project (nForm)
Client: The Northern Trust Company (TNT), Chicago, IL, U.S.A
Work Location: Chicago, IL Role: Team Lead July 2006 – Oct 2011

Description:
To re-engineer profitability and trust asset reporting, costing, planning, and forecasting to drive better decisions, accountability, and profit. This system gets data from roughly 80 source systems and applies calculations to derive organization, client & product profitability. It modernizes our financial technology platform and updates methodologies to enable fact-based, consistent & transparent reporting and planning of profitability and other financial metrics in order to facilitate decision-making. Financial methodologies like Cost Allocation, Revenue Allocation, Capital Allocation, and Fund Transfer Pricing along with the Balance-sheet/Income Statement reporting are used to drive the organizational profitability.

Environment:
PeopleSoft EPM 8.9, IBM InfoSphere DataStage, People Tools 8.4, Windows XP 2002, UNIX, Oracle 10g, Control-m, Rapid SQL

- Implemented multi-dimensional financial data warehouse with IBM InfoSphere DataStage 7.5 and Oracle 11g as a back-end database in Unix environment.
- Spearheaded a development team of over 15 members disseminated across onsite and offshore locations.
- Delivered organization, client profitability, trust asset reporting and planning and forecasting data marts with performance efficient design, which ensured on-time delivery of financial reports to senior management.
- Designed and implemented multi-dimensional security for operational reports to ensure highly secured information governance using PL/SQL procedures.
- Migrated IBM DataStage version from 7.5 to 8.5.
- Automated and streamlined execution notification processes for uninterrupted schedule flow Using technologies like IBM InfoSphere DataStage Universe/XMETA Repository, UNIX, XML, XSLT, Java, Java XML Parsers (DOM and JAX), Apache Web Server, Java & VB Scripting, Oracle PL/SQL, Microsoft DOS Batch programming and FTP techniques.
