

## Santosh Dighe

### Solution Architect – Data, Cloud & AI

E: sdighe1405@gmail.com ▪ M: +91 9004380181

Profile: <https://santoshdighe.com> ▪ Pune, India 411057

LinkedIn: <https://www.linkedin.com/in/santoshdighe>



---

## EXPERIENCE SUMMARY

- Over 15 years of experience in various phases of technology projects and business execution related to **Data** and **Business Intelligence**, including design, planning, and development in varied industries, including **Financial Services**, **Human Resources**, and **Retail**.
- Strong **Problem-solving**, **Analytical**, **Business Development**, and **Communication** skills with the ability to design and implement innovative technology solutions efficiently.
- Experience in **Cloud Architecture & Migration**, **Big Data**, and **Data Science** technologies. Worked for major corporate clients in the USA and successfully used technology for business benefit.
- Extensive expertise in implementing **Cloud** & On-premise **Data Warehouse** projects using various **ETL**, **Data Analysis & Processing**, **Database**, and **Visualization & Reporting** tools and services.
- End-to-end implementation of **Data Analytics** & **Machine/Deep Learning** pipelines with Data Collection & Preparation, Model Building, Training, Testing, and Deployment with web hosting.

## WORK EXPERIENCE

### PinCarts Private Limited

Sep 2016 – Present

Designation: Technology Architect – Data, Cloud & AI & Co-founder

### Core Responsibilities:

- Strategic planning for the company's future growth, including technology and business operations, marketing, and sales strategy development.
- Designed and implemented an end-to-end analytical **Cloud Data warehouse** and **BI** solution with **AWS Redshift** lake house architecture with **Data Export**, **Streaming**, and **Query** services for **ETL** and **Data Analysis**, **AWS S3**, **Relational Databases**, and **NoSQL** MongoDB for data storage and **Tableau** for **Data Visualization** and **Reporting**.
- Built an end-to-end **Cloud Data Lake/Mart** and **Analytics System** with data from varied data sources (structured/unstructured) using various technologies like **Hadoop HDFS**, **Cloudera**, **Spark**, **Kafka**, **Impala** & **StreamSets** **ETL**, **Relational** & **NoSQL** DB, and **AWS S3** & Block storage.
- Working knowledge about **Microsoft Azure Data Lake**, **Synapse Analytics**, **Data Factory** and **Snowflake Data Cloud** technologies.
- Designed and implemented Content-Based Fashion **Recommendation System** using Hadoop HDFS, Hive/Impala, Apache Spark, Mahout, Java, Python, & Elasticsearch.
- Business-to-business Sales - On-boarded over 25 retail customers with end-to-end system integration for product inventories to provide digital and in-store navigation of products.
- Design, develop and maintain advanced data and machine learning applications to support the company's vision of providing personalized and real-time location-based shopping experience.
- Proximity marketing platform using various location technologies like BLE Beacons (Eddystone/iBeacon) and geofencing for tracking customers in real-time.

**Hexaware Technologies Inc.**  
**Designation: Technical Architect**

**May 2006 – Jul 2016**

**Project: Global Accounting Hub and Operational Data Store**

**Client: Northern Trust Corporation, Chicago, IL, U.S.A**

Work Location: Chicago, IL

Role: Technology Architect

May 2015 – Jul 2016

- Implemented operational data store for accounting data in **Enterprise Data Warehouse** using Hadoop and **Apache Impala/Hive Metastore Framework**.
- Designed and Implemented service-oriented architecture for accounting business units.
- Integrated with upstream and downstream systems/applications with web services interface.

**Project: Comprehensive Capital Analysis and Review (CCAR)**

**Client: Northern Trust Corporation, Chicago, IL, U.S.A**

Work Location: Chicago, IL

Role: Automation Architect

Oct 2014 – Mar 2015

- Designed overall automation strategy for production execution of all CCAR schedules, which ensures fully automated execution and rerun ability and email notifications.
- Reduced time spent on quality assurance and decreased product defects by enhancing development procedures.

**Project: Policy Administration System**

**Client: Trustmark Insurance, Lake Forest, IL, U. S. A**

Work Location: Lake Forest, IL

Role: Data Architect

July 2014 – Sep 2014

- Ensured 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.

**Project: Global Data Warehouse (GDW)**

**Client: American International Group (AIG), U. S. A**

Work Location: Mumbai/Pune, India

Role: Team Lead

Nov 2011 – March 2014

Multi-dimensional Data Warehousing Implementation:

- 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.

**Project: Northern Financial Objective Reporting, and Management**

**Client: Northern Trust Corporation, Chicago, IL, U.S.A**

Work Location: Chicago, IL

Role: Team Lead

July 2006 – Oct 2011

- Implemented multi-dimensional financial data warehouse with IBM InfoSphere DataStage and Oracle as a back-end database in the Unix environment.

- Spearheaded a development team of over 15 members disseminated across onsite and offshore locations. Schedule execution automation with rerun ability.

## PROFESSIONAL PROFICIENCIES

Type	Skills
Areas	Data Warehousing ▪ Cloud Computing ▪ Data Engineering ▪ Data Science Deep Learning ▪ Big Data ▪ IoT ▪ Computer Vision ▪ Business Development
Technologies	AWS Cloud ▪ Hadoop ▪ HDFS/YARN ▪ Apache Spark ▪ Hive/Impala ▪ Cloudera CDH ▪ Apache Kafka ▪ Apache Flume/Sqoop ▪ Python ▪ TensorFlow ▪ Keras PyTorch OpenCV ▪ Flask ▪ StreamSets ETL ▪ R ▪ Java ▪ HTML/CSS ▪ Javascript Service Oriented Architecture ▪ IBM InfoSphere DataStage ▪ ETL ▪ PeopleSoft EPM ▪ REST/Webservice APIs ▪ Microservices ▪ Docker ▪ Elasticsearch
Cloud Data Lake/Warehouse	AWS Redshift ▪ Azure Synapse Analytics ▪ Snowflake ▪ Oracle ▪ MySQL
Databases	Oracle ▪ MySQL ▪ PostgreSQL ▪ Microsoft SQL Server ▪ NoSQL MongoDB
Application Management	GitHub ▪ PVCS ▪ MS TFS ▪ SVN ▪ Agile/Scrum ▪ Jira
Industries	Financial Services ▪ Human Resource ▪ Retail

## PROFESSIONAL DEVELOPMENT/CERTIFICATIONS

Certification Program	Institute/Verified By	Month/Year
Neural Networks and Deep Learning	DeepLearning.AI/Coursera	Feb 2019
Data Science for Business	DePaul University CDM IPD, Chicago	Spring Quarter 2015
Big Data Using Hadoop	DePaul University CDM IPD, Chicago	Spring Quarter 2015
The Data Scientist's Toolbox, Getting and Cleaning Data, and R Programming	Johns Hopkins University/Coursera	Jan 2015
IBM Big Data Fundamentals Professional v1	IBM/Prometric	Sep 2013

## EDUCATION

Degree	University/Institution	Major	Year
C-DAC	Institute for Advanced Computing and Software Development, Pune	Post Graduate Diploma in Advanced Computing (PG-DAC)	2006
Bachelor of Engineering	University of Pune, India	Bachelor of Computer Engineering	2004