

SANDIP DATTA

Phone	+91-9886172752
Email	sandipdatta.job@gmail.com
Personal Blog	http://dattasandip.blogspot.in/
Drawings	http://pencilsbysandipdatta.blogspot.in/

Github	https://github.com/mesandipdatta
Linkedin	https://in.linkedin.com/in/dattasandip
Kaggle	https://www.kaggle.com/sandipdatta
TechBlog	https://sandipdatta.wordpress.com/

Profile Summary

- ❖ Big Data Analytics Lead with Overall 11+ years of IT Experience & 5.5+ years of Fully-Hands-on Big Data\Machine Learning Experience - working with Global Clients in Banking\Finance , Retail & Manufacturing domains to Build ,Enhance & Implement Enterprise Business applications in Cloud & Non-Cloud Environments.
- ❖ Financial Domain Knowledge – OFSAA Product, AML , Basel, IFRS, CASA, Overdrafts etc.
- ❖ Experienced in IoT & Non-IoT Solutions & Applications with Big Data Stack , Data Science Techniques and visual analysis frameworks using various tools & open source libraries.
- ❖ Worked on Multi-Node Cluster Systems processing TBs of Structured , Unstructured & Semi-Structured Data.
- ❖ Experienced in Conceptualization, Coding , Consulting ,Guiding & Leading teams for Projects, POCs , POVs, Business Use Cases & Pre-Sales Activities.
- ❖ Worked in Data Science Analytics, Batch\Online\Streaming \Web-REST based Applications , Data Warehouse & Migrations.
- ❖ Held additional responsibilities of Stakeholder Management, Liaising with Clients\Product Owners\Geographical Teams, Task Prioritization, Dependency Identification, Risk & Issue Evaluation, Scrum Meetings , Customer Interface , Business & Management Presentations etc.
- ❖ Well versed in Onshore-Offshore & Discreet-Geographical Model of IT engagement with 8-Month Onsite Client Experience in UK & Thailand as a Lead.
- ❖ Result-driven professional, recognized for taking on new initiatives, adapting to rapidly changing environments and resolving mission-critical issues to ensure bottom-line success.
- ❖ Possess Good communication, Interpersonal, Analytical & Team Management skills.
- ❖ Hard Working & Love to Scale up & Explore New Technologies & Open Source Prospects.
- ❖ Personally – I am a Blogger , Cartoonist , Calligrapher and an occasional Kaggle Participant.

Experience

- ❖ Working as a **Data Analytics Manager** in **Oracle** from March 2019 – Till Date.
- ❖ Study-Break for Masters (Business Analytics) from March 2018 – Feb 2019.
- ❖ Worked as an **Analytics Lead Practice Consultant** in DellEMC (Previously EMC) from Sep-2014 - Feb-2018.
- ❖ Worked as a Tech Lead in ITC Infotech from Nov 2010 – Sep 2014.
- ❖ Worked as a Senior Software Engineer in Fidelity Information Service from Mar 2010 – Oct 2010.
- ❖ Started career as a Software Engineer (Campus Recruit) in Dell (previously Perot Systems) from Jun 2006 – Jan 2010.

Educational Qualifications

M.Sc.	Business Analytics	Birla Institute of Technology & Science, Pilani – Pursuing (2018 –)
B.E	Electrical & Electronics Engineering	Visveswaraiah Technological University
PUC	Science	Cotton College,Guwahati

Skill Summary

Platform	Linux , Windows	Visualization	Plotly , Dash , Tableau
Big Data Stack	Hadoop , HDFS, Spark, Spark-Streaming, Kafka, Cassandra, Druid , Map Reduce ,Hive, Tez, Flume, Yarn, Zookeeper, Sqoop	Languages	Python , Scala\Java (Basic), SQL ,PERL , Shell Scripting , R (Basic)
Hadoop Distro	Cloudera, Hortonworks	Miscellaneous	Github , Tomcat, Postman , Docker
Cloud Platforms	AWS(EMR,EC2,S3) , GCP	Databases	PostgreSQL , DB2 , Sql Server
Machine Learning	Regression, Classification, Clustering, Dimensionality Reduction , Time Series Forecasting	Machine Learning\ NLP Toolkits	Scikit Learn , H2O, Spark ML , NLTK , BeautifulSoup
Web\REST API Framework	Django , Flask	Product Knowledge	OFSAA

Roles & Responsibilities

- ❖ Involved in Pure Technical as well as Functional , Business & Lead roles.
- ❖ Creating Data Pipelines,Process Engines,Ingestion,Analytics(Machine Learning) & Web Framework Solutions.
- ❖ Hands-on Coding & Development,Guiding & Supporting Teams
- ❖ Participate and Contribute to Design & Architectural discussions.
- ❖ Developing Visualizations, Dashboards , Reports.
- ❖ Involved in System Evaluation, Analysis and Recommendations.
- ❖ Support Product Owners in Requirement Gathering\User story creation & Work Prioritization.
- ❖ Involved in Client Relations, Business Development, Technical\Functional Analysis
- ❖ Active Member of Innovation Group to Explore & Implement New Technologies to build Quick-Try-Out\Discovery Solutions.
- ❖ Supporting Pre-Sales team with POCs , Presentations & Generating Process Layouts.
- ❖ Guide and Train Team member in New Tech Stacks through Trainings & Webinars.
- ❖ Senior Co-ordination roles for Development & Requirement Gatherings across Multiple Geographical Teams.

Training & Certifications

- ❖ 6 Weeks Big Data\Analytics\Data Science Bootcamp\Workshop – EMC Analytics Group
- ❖ Attended Big Data Workshops in ITC Infotech
- ❖ IIBF® Basics of Banking Certification | ITIL® 2011 Foundation Certification
- ❖ IBM Certified DB2 UDB V8.1 Family.
- ❖ Intro to Descriptive Statistics - Udacity
- ❖ Workshop in Probability & Statistics (22 Hrs Course)| Statistics for Business Analytics (7 Hrs Course) – Udemy
- ❖ Machine Learning A-Z (22 Hrs Course) - Udemy
- ❖ IBM Certified Badge - Big Data Hadoop Foundations(Hadoop, HDFS)
- ❖ IBM Certified Badge - Big Data Programming (MapReduce & Pig)

Key Projects Summary

Project	Description
OFSAA Oracle Banking Clients	OFSAA is Financial Analytics Big Data product which caters for Enterprise Market Risk , Performance Management , Customer Insight & Financial Crime & Compliance Analytics. Stack – Spark , Hive, Livy, HDFS, Zeppeline,Sqoop, Classification & Clustering Models
Smart Factory European Client	It is a Manufacturing & Industrial IoT solution built on top of Open source Big Data & Data Science components. Works as a Smart factory mode to use various Sensor data through Gateways and building Real-Time Analytics to handle & minimize production issues through Alert Generation & Anomaly Detection for Prescriptive & Predictive follow-ups. Stack – Spark ,Kafka, Druid, HDFS, Clustering Models , Superset , AWS
Fault Analytics Platform US Client	Predictive Maintenance Big Data Analytics Platform for anticipating failures and taking pre-emptive actions for servers running high-consumption & high-volume consumer or transactional data. Batch process on “ Dark Data ” using Spark ML algorithms to check server health & identify fault-prone servers which would need immediate supervision & maintenance. Stack – Hadoop, Spark ,Classification & Clustering Models
Log Analysis US Client	Business Rule based Log Analysis Solution is developed to Parse , Analyze & Store Variety of Logs belonging to various EMC server family – VxBlock\VxRack\VxRail\VMAX\VN\Isilon\XtremeIO etc. Idea was to offer a Business driven comprehensive overview of usage , power, processing, aggregation by various factors, performance as well as downtime. Stack – Cassandra, Hive , Text Analysis , Spark
Casper – Data As a Service EMC Internal	Casper is a profile based data service & administration, metadata management platform. Also features Sandbox provisioning, Hadoop based workspace enablement, Data cataloguing & search.Built to facilitate data governance & analysis, query, data & metadata management from different access level perspective. Also connects to external services – Zeppelin , Jupyter , Hive , Solr etc . All services are exposed as Rest Api accessible through browsers. Stack – Spark, Zeppelin, Django , Python
Cyber Security Platform Malaysian Govt.	Big Data , Machine Learning & Text Analytics Platform for Cyber-crime & Fraud Analytics. Solution built with NLP & Predictive analytics(Machine Learning) pipelines implemented on Big Data platform as advisory solutions to imply possible fraudulent probabilities & threat classification using H2O.ai & Spark ML\MLLIB. Stack – Hadoop , H2O, Python , Java , R , Spark, Classification & Clustering Models

Service Ticket Router Framework US Client	Automatic Routing Service to streamline Service Tickets to appropriate Business group in the support team . Text Processing & Analytics done in Big Data platform and subsequently combined with Machine Learning Algorithms to classify the tickets to service groups. Results exposed as Django based Rest API for interval quality check publishing & viewing. Also connected to Tableau . Stack - Spark, Java, Python ,Django ,Classification Models
Data Sanitation Framework US Client	The Project consists of EMC Data Lake Architecture ,built on Isilon boxes , for ingestion with other tools from the Hadoop ecosystem to build the platform. Converting Queries to SparkSQL-Hive . Data processing using Spark . Stack – Hadoop, SparkSql , Hive
Product Feed App US Client	Application - developed on Big Data and Machine Learning stack to intimate & inform customers with personalized recommendations & feeds on products based on their online purchases. Data Processed , cleaned for Historical analysis. Worked as a developer to build Big Data pipelines Helped to build and develop the framework . Stack – Cassandra, Spark , Hive
Legacy Migration US Retail	Converted & Migrated Terabytes of Legacy Mainframe data to Hadoop using Custom EMC\Tiburon tools. Developed Spark Batch Jobs to process those data & create Custom reports. Written Hive Queries to Analyze data. Stack – Sqoop, Spark, HDFS
Banking Applications US\UK\Denmark Banking Clients	Data Warehouse & ETL , Application Development & Maintenance Of Large Scale Enterprise Platform . For Fortune 500 Banking Services providers. Developing Front-end, Backend & Core Business Logics - Real-time Online system as well Batch Applications \Transaction systems to process very high volume of data (Billions of records, TBs of Data) which generates Daily\Weekly\Monthly Analytical Insights\Reports. Assisted Customers & Business Stakeholders to prioritize Business Requirements. Stack – ETL Tools, Java, Sql , DB2, Sql Server
Miscellaneous Internal & Project Specific	Developing Shell scripts & PERL for automating tasks. PERL regex scripts for quick data clean-up and checks Generate Cross Reference and Migrate Data from DB2 to Sql-Server\MySQL\PostgreSQL etc. using – EMC Proprietary ETL Toolsets . Write SQL queries to analyze Payment data from DB2 and generated Dashboards using Tableau . Generating various aggregate\summary data , File interdependency reports using Excel (vlookup)

Personal Projects Summary

Project	Description
Interactive Forecasting	Browser-based interactive tool which can do Forecasting using different algorithms & packages and publish actual & forecast results as an interactive graph plots. Used ACF\PACF , ARIMA , AIC & BIC Stack – Python, Django, Prophet, Statsmodels
Visualization Framework	EDA Analysis system to analyze various feature attributes – Correlation , Outliers , terminal cases etc. End goal was to produce Interactive Data Visualization spread across timelines. Stack – Plotly , Dash, Python , Django
Web Scraping	Web-Scrap\Crawl various websites data as CSV format and hosted as a Django Rest API . Stack – BeautifulSoup , Jupyter Notebook, Python , Django
Graphical Web App	Built a framework to publish Interactive Graphs\charts on browsers or Web apps. Stack – Python , Plotly-Dash
Similarity Finder	Simple NLP Application - Reads various document contents & analyzes content similarity . exposed as a web-based tool. Stack – Python , Sklearn , Cosine-Similarity , Django
Topic Summary	Crawl Online Text & apply NLP techniques to perform Text analytics & build Topic Modeling . Stack – Gensim, NLTK, Jupyter , Beautifulsoup, Python , Tableau

Achievements

- ❖ Stood 6th in Regional Maths Olympiad 1999 , Assam
- ❖ Received 6 Times - EMC 'Silver Award' , 'Bronze Award' & 'Cause for Applause' Award for Excellent Performance
- ❖ SPOT Award from – Program Manager, ITC Infotech – for excellent performance in project.
- ❖ GAME ON Award from Account Director , Experian , UK – a client of DELL for all round performance
- ❖ SPOT Award - from Vertical Head , DELL - for being acknowledged by client for effective problem resolution
- ❖ Stood First in School in 10th with **Highest Marks in Maths , Advanced Maths & English**