SUNDARESAN MANOHARAN

(510) 304-8141 / mvsundaresan@gmail.com

SUMMARY

Mr. Sundar is a certified Azure/AWS Cloud Architect and a seasoned data enthusiast having 20+ years of experience in Traditional & Modern Data Management and Big Data Analytics, having worked at AT&T, Airbnb, Hewlett Packard and Wipro Technologies. He is based in Dallas and has collaborated with diverse companies in sectors such as Telecom, Online Media, Transportation, Energy & Utilities, Retail, Finance, Health & Managed Care, Government, Insurance and Industrial on a wide range of initiatives. Mr. Sundar's primary focus lies in the realm of Data, specializing in Cloud and Big Data Architecture utilizing various platforms and tools including AWS, Azure, Databricks, Snowflake, Cloudera, Python, Spark, Vertica, Tableau, Qlik Sense, IBM Cognos Suites, Informatica and Ab Initio. His expertise extends to Presales, Strategic Planning, Enterprise Architecture and Infrastructure Architecture for Cloud, Big Data, Machine Learning, Data Science, Business Intelligence and Analytics platforms along with Project/Program and Client Management.

PROFESSIONAL EXPERTISE

- Experience working in common Cloud Data Platform(s) – AWS, Azure, GCP, Databricks and Snowflake
- Expertise in Cloud Architecture, Big Data Analytics, Data Integration, Data Science, Machine Learning, Deep Learning, Predictive Modeling, NLP, LLMs, Web Scraping, Business Intelligence and Visualizations
- Oversaw Assessments, IT Strategy, Scope Definitions, Project Schedules, and Solutions Architecture for Cloud & Big Data Projects.
- Orchestrated the design of Cloud Architecture for Data Ingestion and Analytics across various clients.
- Expertise in handling large scale Structured, Unstructured and Semi-structured data pipelines/streams
- Expertise in Hadoop - HDFS, MapReduce, Hive, NoSQL, Spark, PySpark, Python, R, Java and Scala
- Expertise in Feature Engineering, Regression, Classification, Clustering and Dimensionality Reduction
- Experience in implementing end to end large-scale data lake / warehousing / data ingestion solutions
- Experience in Relational and Dimensional Database Modeling (Logical and Physical)
- Experience in implementing projects and working alongside with Multi-shore Development teams
- Consulting experience in Analytical solution deployments for tactical, operational as well as strategic initiatives and has provided project management and technical leadership to teams
- Strong research, analysis, communication and presentation skills and a strong aptitude for learning

KEY ACCOMPLISHMENTS

- Patent Pending – IDN Homoglyph Attack Detection - Transfer Learning and Siamese Neural Network
- Deep Learning Kaggle Competitions Facial Key Points Detection, Detect Distracted Drivers
- Telco - Repair Job Prediction – Demand vs Supply Prediction
- Predicted US Senate 2014 midterm election in favor of Republicans total # of seats 50-53 (Won 53)
- Configured VPC in the AWS Cloud to process 1.5 TB of AT&T DNS Data feed daily using EMR
- Worked on large relational database of 30+ TB with a high growth rate (Bloomberg Pattern Mining)
- Processed 50+ TB of machine "Telemetry" generated "Big" data in the Hadoop environment
- Sentiment Analysis using stream of DirecTV Twitter handles to prevent customer churn

CERTIFICATIONS

- Certified AWS Solution Architect – SAA-C03
- Certified Microsoft Azure Data Scientist
- Certified Hortonworks Developer: Enterprise Apache Spark
- Certified in HP Vertica Big Data Analytics (Columnar Database)
- Microsoft Certified Technology Specialist – Microsoft SQL Business Intelligence Suite
- Certified Teradata Design Architect – Teradata Corporation
- Certified Data Warehousing Data Architect – Wipro Technologies

EDUCATION

- MS in Analytics – Data Science, Georgia Tech, Atlanta, Georgia, USA
- Data Science – Harvard Extension School, Boston, Massachusetts, USA
- MCA - Master of Computer Application, IGNOU, New Delhi, INDIA
- BCA – Bachelor of Computer Application, IGNOU, New Delhi, INDIA

AWARDS

- AT&T Connection Award – 2021 & 2022 for Threat Detection and Alarm Prioritization
- HP Analytics & Data Management "Consultant of the Year" & "Leading the Way" – 2013 & 2014

TOOLS & TECHNOLOGIES

Cloud Computing	Amazon Web Services (AWS), Azure, Databricks, Snowflake
Databases	MS SQL Server, Oracle, Postgres, MySQL, Redshift, Aurora, Azure Syapnse
MPP Database	HBase, MongoDB, Cassendra, HP Vertica, Teradata
Data Integration	Glue, Ab Initio, Informatica, Microsoft SSIS, TalenD, Analytix Mapping Manager
Business Intelligence	PowerBI, Tableau, Microsoft SSAS, SSRS, Custom MDX Queries, Crystal Enterprise
Programming	Python, Scala, Java, Shell Scripts, SQL, PL/SQL, VBA, JavaScript, HQL
Data Science	iPython, Pandas, Scikit-Learn, Matplotlib, R / R Studio, NLTK / NLP, D3.js
Big Data	Hadoop, YARN, Map Reduce (v1/2), Pig, Hive, Sqoop, Mahout, Spark, PySpark
Data Modeling	ER Win, MS Visio
Methodology	SDLC, Agile - SCRUM

PROFESSIONAL EXPERIENCE

MAY'20 TO PRESENT – PRINCIPAL Cloud Architect, AT&T Cybersecurity, PLANO, TX
JUL'16 TO MAY'20 – Principal BIG DATA Architect, AT&T SERVICES INC, PLANO, TX
JAN'08 TO JULY'16 – ENTERPRISE DATA ARCHITECT, HP ENTERPRISE SERVICES, PLANO, TX
- HPES IT - INFRASTRUCTURE, PALO ALTO, CA – BIG DATA ARCHITECT
- HP IMAGING & PRINTING SERVICES, PALO ALTO, CA – DATA ARCHITECT
- CVS CAREMARK, RICHARDSON, TX – ETL ARCHITECT
- STATE OF SOUTH CAROLINA, SC – DATA INTEGRATION ARCHITECT
- SOUTHWEST AIRLINES, DALLAS, TX – ETL ARCHITECT
- CVS CAREMARK, NORTHBROOK, IL – ETL ARCHITECT
- BLOOMBERG FINANCIAL MARKETS, PRINCETON, NJ – ETL / BI ARCHITECT
- GEORGIA CANCER COALITION, ATLANTA, GA - BI ARCHITECT
- PARADIGM MANAGEMENT SOLUTIONS, CONCORD, CA - ETL ARCHITECT
- DIRECT ENERGY, HOUSTON, TX - BI ARCHITECT

OCT'03 TO JAN'08 – DATA WAREHOUSE DESIGN ANALYST, WIPRO TECHNOLOGIES, MOUNTAIN VIEW, CA
- RAPT INC. (ACQUIRED BY MICROSOFT ADVERTISING), SAN FRANCISCO, CA – ETL ARCHITECT
- 3M (MINNESOTA MINING AND MANUFACTURING CO.), SAINT PAUL, MN – TERADATA ARCHITECT
- EASTMAN KODAK COMPANY, ROCHESTER, NY – DATA ANALYST

APR'03 TO SEP'03 – SENIOR SOFTWARE ENGINEER, INFOGAIN INDIA (P) LTD, NOIDA, UP, INDIA
- KIA MOTORS AMERICA, IRVINE, CA – SENIOR SOFTWARE ENGINEER

AUG'02 TO MAR'03 – SOFTWARE CONSULTANT, OTIS ELEVATORS CO INDIA LTD, MUMBAI, MH, INDIA
NOV'00 TO JUL'02 – DATABASE DEVELOPER, PINNACLE SOFTWARE SYSTEMS (P) LTD, CHENNAI, TN, INDIA
APR'98 TO JUL'00 – SOFTWARE ENGINEER, THIRD WAVE COMPUTERS CONSULTANCY, MADURAI, TN, INDIA

PROJECTS

PRINCIPAL CLOUD ARCHITECT / DATA SCIENTIST
AT&T CYBERSECURITY- AT&T SERVICES INC. (MAY 2020 TO PRESENT) PLANO, TX

ALIENVAULT – ALARM PRIORITIZATION / SUPPRESSION, PLANO, TX
- Use case: Prioritize and Auto suppress Alarms using BERT Embeddings
- Extract & Load ELT alarms activities data from USM Anywhere
- Transform / Engineer features to detect anomalies using Pretrained Language Models
- Used the BERT Embeddings to find the Semantic Textual Search
- Ranked the alarms using Cosine Similarity Score
- Number of alarms reduced by 10x
- SOC focuses on most severe alarms and auto suppress low severe/repeated alarms
- Collaborated with USM Engineering team to implement end to end Solution into PROD
- Improved model performance using SOC / Customer feedback loop
- Tools: ElasticSearch, AWS EMR, AWS RDS, Spark, Python, Apache Kafka

ALIENVAULT – UEBA THREAT DETECTION, PLANO, TX
- Use case: Detect Threat using User Entity Behavior Analysis
- Collect login activities log event data from various sources / applications

- Standardize the event log data into Common Record Format
- Engineer features to detect anomalies using Machine Learning Models
- Explored HBOS and iForest unsupervised algorithm to generate outlier score
- Collaborated with USM Engineering team to overarching end to end PoC Solution
- Improved model performance using SOC / Customer feedback loop as reinforcement learning
- Identified data issues in Geo IP Metadata and fixed the data issues
- Received an "***AT&T Connection Award***" for successful end to end PoC demonstration
- Tools: AWS EMR, AWS RDS, Spark, Python, Kafka

ALIENVAULT – DOMAIN SIMILARITY THREAT DETECTION, PLANO, TX
- Use case: Detect IDN Homoglyph Attack using Domain Names
- Collect AT&T DNS Resolver data from Data Router
- Configured AWS VPC and EMR Cluster to process 1.5 TB per day of continuous data feed
- Convert the "*.csv.gz" to Parquet for faster distributed data processing further in the data pipeline
- Profile and Clean domains with Domain Whitelists and Known domain lists
- Identify Newly observed domain every day and send it to AlienVault Researchers for investigation
- Collaborated with Cybersecurity researchers to find domain similarity with AT&T Customers domain
- Performed various quick PoC on OCR, Language Detection Algorithms, Fonts, Image Generations
- Explored CNN VGG16 Transfer Learning and Fastest Nearest Neighbor search algorithms
- Researched and Implemented a Siamese Neural Network Architecture to compare domains
- Submitted for **US Patent – Homoglyph Attack Detection (2021-0129) "Pending"**
- Tools: AWS EMR, Spark, Python, Tensorflow, Keras, FAISS

PRINCIPAL BIG DATA ARCHITECT
CHIEF DATA OFFICE - AT&T SERVICES INC. (JULY 2016 TO MAY 2020) PLANO, TX

MOBILITY – CALL INTENT TAGGING
- Use case: Classify call intent(s) using agent notes
- Pre-process agent notes "free form text" data, ~10 Mil calls a month
- Created a dictionary to correct for abbreviations, shorthand notation, products and synonyms
- Created rules with key terms to identify/tag each agent notes aligned with business needs
- Clustering techniques used to cluster untagged message to enrich business rules
- Improved business forecasting on call volume and intents of the calls
- Tools: Hortonworks HDP, HDFS, Spark 1.6, Spark MLLib, Scala

DIRECTV – CHURN DETECTION USING TWEETS
- Use case: Detect Customer Churn using Tweets
- Collected historical tweets using Sprinklr
- Pre-process tweets to remove URLs, Emoticons, #hashtags and @users
- Used LDA topic modeling, Graph analysis to find insights
- Using sentiment analysis and proprietary logics to hand label Churn indicator
- Trained various Classification models to measure the performance
- Helped pro-actively reach out to customer to save them from potential Churn
- Tools: Hortonworks HDP, HDFS, Spark, Python, Scikit-Learn, TensorFlow

MOBILITY – CREDIT DECISION ENGINE
- Use case: Reduce credits given to customers
- Extracted Mobility consumer business Credit transactions from various sources
- Cleansed and Created model ready dataset
- Exploratory data analysis to find the high credit reasons
- Created DecisionTree Regressor to find an optimal credit amount
- Optimized / Reduced the credit amounts given by 25%
- Tools: Hortonworks HDP, HDFS, Spark, Python, Scikit-Learn

MOBILITY – INTRADAY RETAIL FORECAST
- Use case: Forecast Intraday Labor requirements
- Collected retail transaction data, store traffic data, store schedule
- Exploratory analysis on Transactional data and Shopper Track Traffic data
- Used number of transactions to forecast traffic and labor forecast
- Tools: H2O.AI, Python, HDFS, Hive, Sqoop

MOBILITY – CITY OF ATLANTA. GA – TOGETHER FOR SAFETY ROADS
- Use case: Prevent accidents on North Ave, Atlanta, GA
- Collected "zendrive" data from Mobile Phone users
- Detects Harsh breaking, Aggressive Acceleration, Speeding, Phone use and Accidents
- Visualized location of each event on the map
- Combined with City data, Traffic data and 911 data for correlation analysis
- Provided insights to City of Atlanta officials
- Tools: Hortonworks HDP, HDFS, Spark, Python, Tableau

ENTERPRISE DATA ARCHITECT
HEWLETT PACKARD COMPANY (JANUARY 2008 TO PRESENT) PLANO, TX

HPES IT - INFRASTRUCTURE, PALO ALTO, CA – BIG DATA ARCHITECT
- Size and Configure Hadoop clusters for Text Analytics
- Categorize support desk ticketing data using unsupervised machine learning algorithm K-Means
- Develop configurable solution to discover the pattern(s) in the ticketing data (human) and server logs (machine)
- Create Business Rules to identify discovered patterns and enrich rules after every iteration
- Develop predictive model using Naïve Bayes algorithm to predict the future ticket category
- PoC on multiple technology stack
  - Apache Mahout & MapReduce (MRv1)
  - Hadoop streaming API using Python NLTK (Natural Language Processing) and NGrams
  - Apache Spark on YARN
- Managed 5 member technical team
- Implemented Apache Spark solution in the production environment
- Tools: MapR distribution, HDFS, Spark MLLib, Scala, Python - NLTK

HP IMAGING & PRINTING SERVICES, PALO ALTO, CA – DATA ARCHITECT
- Pilot on 50+TB of ink usage information generated by Printer Telemetry sources across the world
- Data collected in Data Lake (Hadoop – MapR distribution)
- Evaluate tools to parse the incoming semi unstructured data (ini, XML, JSON) generated by printers
- Evaluate Analytix Mapping Manager to define and document the source to target functional / technical mapping specifications from business requirements
- Reverse engineer the existing MapReduce java code and create functional design specifications
- Assess and suggest Big Data Integration for the data flow (Landing, Staging, CRF, and Reporting)
- Tools: HDFS, HBase, MapReduce, TalenD, Analytix Mapping Manager 5.2, ERWin 9.2

CVS CAREMARK, RICHARDSON, TX – ETL ARCHITECT
- Responsible to manage the DW / BI development life cycle to produce planned deliverables, timeline, and cost
- Provide technical guidance and directions to designers and review
- Manages staff assignments and ensures that staff has adequate training and education for the role
- Manages all project planning, progress tracking, and team coordination
- Anticipates project issues / risks, formulates actions to address and tracks to its completion
- Manages a continuous stream of team actions and communication to adjust client's expectations
- Captures, documents, and takes action to resolve all changes to project scope
- Responsible to coordinate mixed HP / client teams to operate cohesively
- Responsible for directing team leads to calculate estimate-to-complete status
- Writes and delivers to HP EM the HP weekly/monthly status reports
- Ensures that a formal weekly or bi-weekly status meeting takes place
- Gather, synthesize and present lessons learned
- Provide timely performance feedback and assist in performance assessment of team members

STATE OF SOUTH CAROLINA, SC – DATA INGESTION ARCHITECT
- Analyze current state of data conversion
- Re-strategize data conversion approach
- Re-design rollout strategy (47 source systems – Region vs Vendor)
- Optimize cutover window (62.5 hours)
- Pre-load & PoC of new data conversion approach
- Proposed Master Data Management solutions
- Proposed Data Profiling Tool solution
- Data Masking / Data Sanitizing
- Develop plan, estimate and schedule
- Tool solutions and vendor management
- Manage & Provide technical guidance to the team
- Re-organize the team structure and staffing
- Co-ordinate with the cross functional team leaders
- Identify potential risks and create and execute plans to mitigate them
- Prepare weekly status reports for Business Stakeholders

SOUTHWEST AIRLINES, DALLAS, TX – ETL ARCHITECT
- Created functional & technical source to target mapping documents
- Desgined data migration from different source systems
- Data integration between two major airlines system (Southwest & Airtran Airlines)
- Created High Level and Detailed design documentation of ETL Logic / work flow
- Development of complex ETL graphs using Ab Initio 3.0
- Engaged in creation of business test cases and validation of data migration/integration

- Manage & Co-ordinate onsite / offshore team activities

CVS CAREMARK, NORTHBROOK, IL – ETL ARCHITECT
- Involved in requirement gathering & business analysis
- Desgin and development of ETL using Informatica 8.1.6
- Worked on Audit, Balance and Control model
- Created High Level and Detailed technical documentation of ETL Logic / work flow
- Engaged in creation of business test cases and validation
- Manage & Co-ordinate onsite / offshore team activities

BLOOMBERG FINANCIAL MARKETS, PRINCETON, NJ – ETL ARCHITECT
- Desgin and development of ETL using Informatica 9.0
- Parse unstructured data files
- Metadata Management (MDM)
- Worked extensively on large data volumes (~30 TB of data loaded)
- Worked extensively on partitions (Oracle and Informatica)
- Combined both Partition and Parallelism to get highest throughput
- SQL queries tuned and optimized for the maximum performance
- Worked on Audit, Balance and Control model

BLOOMBERG FINANCIAL MARKETS, PRINCETON, NJ – ETL/BI ARCHITECT
- Proof of Concept designed & developed in Microsoft BI Platform
- Loaded 1.5 TB of data in SQL Server 2008
- Parse unstructured data files in SSIS 2008
- Cube designed & built on SSAS 2008 for analytical and adhoc reporting

GEORGIA CANCER COALITION, ATLANTA, GA - BI ARCHITECT
- Designed, developed and implemented a dimensional model
- Designed and developed OLAP cubes using SSAS
- Worked extensively on custom MDX queries
- Customize dimension filters in Dashboard for easy navigation
- Implemented Dashboards using PerformancePoint Server and SharePoint Server 2008
- Experience of PerformancePoint Server configuration in both Native and SharePoint Mode
- Knowledge of Kerberos configuration for Single Sign On

PARADIGM MANAGEMENT SOLUTIONS, CONCORD, CA - ETL ARCHITECT
- Reverse engineer and document metadata for source systems
- Perform domain analysis and deliver data matrix, source to target data mappings
- Work with team members to incorporate and design for enterprise data architecture
- Assist in building corporate ETL standards and best practices
- Developed comprehensive documentation of data source definitions and data mapping
- Defined best approaches for procuring source data and implemented optimal ETL logic / workflow processes using SSIS
- Designed and developed auditing logic, ETL packages using SSIS
- Defined data acquisition and extract policies and implemented optimal ETL logic and workflow processes
- Trained team of developers in Kimball warehousing methodology fundamentals, ETL workflows and overall design and development principles
- Assisted Quality Assurance teams in developing test plans to validate data marts, reports, views and schemas before production deployment
- Provided detailed translation of ETL logic / work flow in to a technical documentation

REFERENCES

Available upon request

LinkedIn Profile: http://www.linkedin.com/in/SundaresanManoharan
Refer to my LinkedIn profile to see other's recommendations about my previous works.

I Declare that all the information given above are true.