

Manisha Sethi

Data Architect

Email: sethi.mani89@gmail.com

Phone: +65 89382471

Professional Summary

- Around **12 years** of experience in design development and architecting distributed systems ,data lakes . data platforms (cloud -native/on-prem) & Data Mesh using technologies like Hadoop, Spark , Structured Streaming , AWS , Databricks, Snowflake, AZURE , ML , CD4ML , MLOps etc along with Data Governance & Data discovery.
- Demonstrated capability in understanding business requirements and translated them into execution & delivery
- **Data community & Capability Lead and Service Line leader** in solutioning and pursuits for Data projects.
- Experience working on Public and private cloud like AWS, Azure and GCP and GovTech GCC cloud environments.
- **Cloudera Certified Hadoop Developer, Cloudera Certified HBASE Developer, Hortonworks Certified Hadoop Developer, Map-R Certified Spark Developer , AWS Certified Solution Architect**

Competencies

- **BigData Technologies:** NoSQL Databases(HBase , Dynamo And Cassandra), Storm, Kafka ,Hadoop, Streaming ,MapReduce, YARN, Zookeeper, Cascading, Driven SPARK ,Impala, Spark SQL , Presto, Zeppelin , docker , Airflow , SageMaker , Azure ML , Databricks etc
- **Development Tools:** Eclipse, IntelliJ, PyCharm etc.
- **Database:** MySQL, NoSQL: Cassandra, Hbase, Dynamo DB , CosmosDB ,Solr, Lucene.
- **Programming Languages:** Scala, Java
- **Cloud Tech :** AWS – EMR | EC2| Glue | Athena | SQS | Kinesis etc , AZURE – HDI , Databricks , ADLS , Synapse , Unity Catalog ,Purview etc IaaS- Terraform etc

Education Summary

Qualification	Board/University	Percentage
M.Tech (2013-15)[Work Integrated]	Kurukshetra University, Kurukshetra	80%
B.Tech(2007- 11)	Kurukshetra University, Kurukshetra	84 %
12th (2006-07)	CBSE Board	90 %
10th (2004-05)	HBSE	87%

Work Experience

Dates	Organisation	Designation(Role)
July-2011 –Dec 2013	Impetus Technologies	Software Engineer (R&D wing)
Jan-2014-Aug 2015	Xebia IT Architects pvt. Ltd.	BigData Consultant
Aug-2015 t-May 2018	Pitney Bowes India pvt. Ltd.	Advisory Engineer
June-2018 till Date	ThoughtWorks	Lead Consultant (Data Architect)

Key Projects and Accomplishments

Project : HPB PHDH [SINGAPORE]

Duration : Aug 2021 – till Date

Skills : Java, Scala ,Spark Streaming ,Azure, Terraform, Data security and Governance ,spark SQL , Azure Databricks , GovTech , Unity Catalog , Vault , Modernisation

Brief: Data Architect

Project involves building and evolving a Azure native data platform to ingest Singapore citizen's health data and other day to day activity data from wearables and apps to the platform and processing it for deriving insights out of the data.

Responsibilities and Contributions:

- Designing Data platform in **Government Cloud for Singapore government**, Migration of existing flows to secure restricted GCC environment
- Managing the IM8 compliance and other regulations implementation as per SG govt protocols.
- Managing the stakeholders and working with Devs , BA & QA in the team to ensure timely delivery.
- Working as part of **pre-sales data service line** for proposals on **Smart Automated Marketing** on Singapore citizens via personalization.
- **Managing TWDU program globally for data enthusiasts, Hosting and Leading Data Community & Capability**

Project : Planet Fitness Data Platform [US]

Duration : July 2018-May 2021

Skills : Java, Spark Streaming, Kinesis, Presto ,AWS, Terraform, Data security and Governance spark SQL, Tableau , Airflow

Brief: Team Lead and Consulting Data Partner

Project involves integrating the bigdata capabilities to a big gymming chain and franchisee in US. To create a marketing intelligence data mart. Integrate data from different sources and process and run campaigns to maximize revenue and profits.

Responsibilities and Contributions:

- **Team Lead** for Data Platform in managing data platform Team , design and roadmap •
- Moved to **Data Architect role gradually** , managing technical vision and evolution of Data Platform

Project : Spectrum at Pitney Bowes [INDIA]

Duration : Aug 2015-May 2018

Skills : Java, MapReduce, Spark, Apache, Cloudera & Horton flavors of Hadoop, Shell scripting, Hadoop File Types : RC/ORC/Avro/Sequence/Parquet , Hive [UDF, UDAF, UDTF], oozie, Hue ,Security : Kerberos ,Ranger and Sentry, Encryption Zones Brief:

Project involves integrating the bigdata capabilities to existing spectrum platform and known as BigData Integration(BDI), here Hadoop file systems and types are integrated to spectrum client. Other part is BigData Quality, doing data quality and address quality using loqate, Address Doctor and own addressing and data quality engines for address Matching and validation jobs to be used by e-commerce and banking clients of spectrum, to whom product is delivered as SDK.

Responsibilities and Contributions:

- Writing Java code for BDI to integrate Hadoop file systems.
- Writing MR, Hive UDAF and spark applications For data and address quality
- Exploring and doing tasks for workflow integration with Hue and oozie
- Setting up the apache, CDH and HDP clusters for dev and product certification tasks
- Collaborations with Cloudera and Horton for certifying the spectrum product, on their latest version, included new security features like Sentry, Ranger and encryption zones as per their requirement along with Kerberos.

Project : Advertising Intelligence [INDIA]

Duration : Feb-14-July 2015

Client : Pubmatic , Pune

Skills : Kafka, zookeeper,Avro, Cascading, Driven, AWS, Junit, Amazon s3, amazon EC2 and EMR

Brief:

Project involves getting the business data as impressions and getting the data stored in kafka in avro form and do the processing using Cascading and its insights over amazon to generate reports over the processed data stored in Hbase in hourly processing period. The jobs get deployed through quartz over EMR.

Responsibilities and Contributions:

- Setup the complete ingestion cluster setup over kafka and zookeeper
- Cluster Design for ingestion
- Data processing from avro being read and analyzed using Cascading and Driven • Cluster deployment over EMR and writing Junit Testing for processing flow.
- Using S3 as filesystem and migration of processing over AWS to hadoop2.

Project : Volly (Digital Mail Box) – Platform Services & API , Reporting [INDIA & US]

Duration : Oct 2011 till Dec 2013

Client : Pitney Bowes

Skills : Spring, Spring-Batch, Web Services, Cassandra, JDBC, BIRT, MySQL, Localization, XML Parser, File SFTP, Solandra, Core Java Brief:

Volly is an Application which went live just few months ago in Australia, It is a consumer centric application where various Service Providers and end users can register for digital Delivery of any kind of Bills/Notes and where User is given the option to choose various providers to add as per matching criteria and thus got almost all of its bill gets delivered in to single account in Volly.

Responsibilities and Contributions:

- Using Spring Batch to Process Persisted Events through various steps and producing output files and getting it delivered to the Subscriber's SFTP location.
- Writing Purge Batch Jobs to delete processed events or expired events from Cassandra with Multiple Threads and using Transactions to speed up the deletion Process.
- Development and testing of Services to subscribe and the code to read and authenticate Input files and Deliver the output file as per matched by supplied input.
- Explored and configured solandra in the system and thorough testing on the cluster for Replication factor and consistency level, when used in cluster.

Personal Details

Date of Birth	10-SEP-1989
Sex	FEMALE
Nationality	INDIAN
Marital Status	SINGLE