



Mayank Mishra

Indore, India 
+91 9669688999 
mayankmishra01@gmail.com

To tremendously empower an organisation's ability to transform their customer's businesses.

Career Summary

- Over **20+ years** of expertise in Service and Product IT enterprises
- **Build 1000+ team sized service lines from scratch for Data & Analytics**
- **Fortune 100 and Fortune 500** - Implementations, Solutions and Architecture consulting
- **14+ years** of Commercial Product development
- **5+ years** of expertise in Machine Learning, Deep Learning solutions
- **~2 year experience in Generative AI**

Technology Skills

- **Big Data Lake, Warehouse and LakeHouses**
- **Comprehensive set of services on Cloud - AWS, Azure and GCP**
- Languages - Java, Python, Rust and Querying
- Stores - Cassandra, HBase, Neo4J, Janusgraph, Redshift, DynamoDB, BigTable, BigQuery
- In memory - Redis, Presto AWS Athena, Impala, Hive LLAP
- AI - LLM, RAG, Azure AI Search, hypernear, LangChain, Azure Open AI, Amazon Q
- Analytics - Spark, PySpark, Hive, Tez, Glue
- **Data Platforms - DataBricks, Snowflake, AWS EMR, Azure HDInsight, GCP DataProc, CDP, CDH, HDP, MapR, Denodo**
- Enterprise - Camel, MuleSoft, Kafka, RabbitMQ, ActiveMQ, Airflow, Jupyter

Skills

- **Technology Thought leadership**
- **Customer Architecture, Consultation**
- Strong Product Development & ISV Development expertise
- Demonstrated ability to sell solutions to customers
- Innovativeness
- Influencing C level executives
- **Demonstrated ability to build, hire, mentor and entire service/practice teams**

Employment History

- Principal Engineer: Wipro-Jun24 onwards
- **Director of Technology:** Ontic-Mar23 to Aug23
- **Director of Technology:** Impetus-Jul18 to Feb22
- **Sr. Architect:** Impetus - Jul14 to Jun18
- **Architect:** Impetus -Jul12 to Jun 14
- 14+ years expertise in Data Lake Analytics, Big Data and Cloud technologies
- Consulted Data Architectures & implemented use cases for 20+ customers - BFSI, Healthcare, Travel, Security, Manufacturing
- Helped Pre sales for RFPs, RFIs of 50+ Fortune

Licenses & Certifications

DataBricks Generative AI Engineer - Issued Jan 2025, Databricks

AWS AI Practitioner - Issued Jan 2025, Amazon Web Services (AWS)

Neo4j Certified Professional - Issued Aug, 2024, Neo4J

Large Language Models: Foundation Models from the Ground Up - a course of study offered by Databricks. - Issued May 31, 2024, *edX and Databricks* verified certificate

Patents Granted

509008 - **"SYSTEM AND METHOD FOR TRACKING FLOW OF BIG DATA IN A DISTRIBUTED FILE SYSTEM"**
ISSUED ON 09-FEB-2024

506640 - **"DYNAMICALLY OPTIMIZING PERFORMANCE OF A DISTRIBUTED COMPUTING FRAMEWORK"**
ISSUED ON 02-FEB-2024

Experience

JUN 2024 - TILL NOW

Principal Architect, Wipro

- Working for a leading global bank offering Credit Cards, Wealth Management and Investment services.
 - Helped customer in building Architecture strategies for modernised data platforms.
 - Hired and built 100+ engineering team with Architects, Leads, Data engineers for customer.
 - Spearheading Wipro in person meetings with key customer's stakeholder visits.
- Heading Delivery Assurance for entire North America region for Data Analytics & AI opportunities
- Architectural & technical guidance to all bank accounts
- Technologies - Spark, PySpark, Scala, Cloudera, Neo4J, Snowflake, Java, AWS

AUG 2023 – MAR 2024

Principal AI Architect

- For an American multinational chain retailer, one of the world's largest in it's segment, Architected a next generation Generative AI platform on Azure Cloud.
 - Built data flow strategy involving company knowledge base, data admins and business users involving text splitters, embeddings, Indexers and Indexing, language model task decomposition, Memory model, moderation filters and more.
 - Architecture provided a secure and trusted environment that strictly limits the exfiltration of data from the secure area for data admins, resting data, business users and all AI, Data interacting components.
 - Included LLMOps strategy for including multiple LLM models, model monitoring, LLM metrics such as Performance, ROI, Cost, Safety, etc.
 - *Technologies - Azure AI Search, Langchain, Azure Open AI, Azure Data Factory, Azure Blob Storage, Moderation Filter, Azure Portal Web App, Azure Bastion, Azure Key Vault, Azure Policy, Virtual Network Gateway and more*

- Built a RAG System to solve hallucination of LLM by using a external long term memory along with GPT 3.5, to enhance the ability to Question Answering with OpenAI
 - Created Dense Vector embeddings to retrieve data that is semantically related to queries
 - Used Open AI GPT natural language model with 4096 context window
 - Modified prompt to provide context to LLM
- Implemented a Retrieval Augmented Generation (RAG) technique with source information chain using LangChain
 - Ingested vectors from movie data set
 - Implemented namespace to reuse data processing pipeline but queries a subset of movies.
 - Search for similar vector embeddings only among those items that match the filter
- Currency Matching - using OpenCV using a fusion of FAST key point detector and BRIEF descriptor
- Distance measurement using OpenCV and Single Shot Detection as an alternative to YOLO
- Training on Vector DB internals and fine tuning
- Technologies - RAG, OpenAI's ada-002 model (1536 dim), gpt-3.5-turbo-instruct(alt. text-davinci-003), pinecone, langchain

MAR 2022 – AUG 2023

Director of Technology/ Ontic Technologies, India

- **Architected Data Lake, Implemented Data Lake in Stage and Prod and built team Data Lake & Analytics from scratch**
- **Accelerated Zero Shot Classification task of Robustly Optimised BERT model with transformer based neuron architecture on AWS Inferentia neurons**
- **Implemented entity relationship intelligence using Neptune based graph model connection establishment**
 - Instantiated, Configured and created graph model for business domain
 - Hydrated Neptune from S3 data, both for bulk ingestion and incremental loads
 - Queried hydrated graph network using Gremlin to get hidden and multi legged relationships between entities
- Laid down the foundation of AWS cloud based data lake
- Implemented AWS Glue, PySpark, AWS Lambda based data pipeline
- Implemented data ingestion and data reconciliation on object store
- ***Created Machine Learning based business use cases for achieving business imperatives -***
 - Observed person's state change detection,
 - Change point detection in time series data of observed signals,
 - Ensemble based Anomaly Detection in observed signals
 - Named Entity Recognition - BART, BERT model and Zero shot classification with Hugging Face
- Implemented **Material Views for OLAP on AWS Redshift**
- Technologies - *AWS Glue, PySpark, AWS Lambda, AWS S3, AWS Redshift, Apache Hudi, AWS Neptune, ML and Deep Learning, AWS Glue Crawler, AWS Athena, Graphana*

JUL 2019 – FEB 2022

Director of Technology / Impetus Technologies, India

Founding member of Big Data consulting arm for Impetus, created engineering stream from scratch. Laid down solution practices, technologies, platforms. Consulted, Architected, and implemented solutions for multiple fortune-100 customers at all stages of their Big Data, Cloud and Digitalisation journey.

- For one of the biggest US based semiconductor industry giant **Modernised Production Data lake** running in HortonWorks based data lake to GCP data lakes
 - Discovery, Assessment, Migration Inventory Prioritisation, Capacity Planning, TCO and Migration planning for 2 PROD clusters running at the scale of hundreds of TBs

- Workload transformation for Spark, Hive, HBase, MapReduce, Tez, PySpark applications
- **HDP—>GCP**
 - Hive on Tez->Spark, Spark 1.6->Spark2.4, HBase->BigTable, MapReduce->Spark2.4
- For one of the biggest US based transport truck manufacturing giant **Modernised Production Data lake** running in Cloudera CDP 7 based data lake to Azure HDInsight and Synapse based lake house
 - Discovery, Assessment, Migration Inventory Prioritisation, Capacity Planning, TCO and Migration planning for PROD clusters running at the scale of hundreds of TBs
 - Done data usage audit for entire data lake
 - Workload transformation for Spark 1.6 and 2.4, Hive, Impala, PySpark applications to Azure technologies
 - **CDP —> Azure HDInsight + Azure Synapse**
 - Hive -> Spark 3.1 (Synapse), Spark 1.6 & Spark 2.4-> Spark 2.4.4, Impala -> Spark 2.4.4 (HDInsight), PySpark 2.0 -> PySpark 3.1
- For one of the Canada based bank **Modernised Data lake** running in Hortonworks HDP Platform to Cloudera CDP
 - Discovery, Assessment, Migration Inventory Prioritisation, Capacity Planning, TCO and Migration planning for PROD clusters
 - Workload transformation for Hive on Tez, HiveLLAP, HBase, Spark, MapReduce, PySpark applications to CDP technologies
 - **HDP —>CDP**
- For a US based bank holding company performed assessment of existing Cloudera cluster to migrate to AWS ecosystem
 - Discovery, Assessment, Migration Inventory Prioritisation, Capacity Planning, TCO and Migration planning
 - Created high level solution architecture and migration plan
- For an American speciality retailer Fortune 500 company **Modernised Data lake** running in MapR Platform to Snowflake platform
 - Migrated Spark Jobs, Hive QLs and Sqoop jobs to Snowflake platform
 - Converted DDLs, Teradata EDW BTEQ scripts to snowflake compatible code and model
 - Recommended architecture, configuration and best practices
 - **MapR Hadoop -> Snowflake**
 - BTEQ ->SnowSQLs, HiveQLs->SnowSQLs, Sqoop -> Snowpipe
- For one of the American multinational computer software company done **AIOps** based optimisation of Cloudera CDH cluster
 - Scale of thousands of nodes, TBs of memory and ten thousands of cores
 - Figured out inefficiencies for slow running applications, data skewness and cluster chocking
 - Recommended optimisations at application, cluster, Hadoop to assist DataOps and DevOps team

NOV 2011 – FEB 2022

Product Creator, Product Head, Architect, Developer **Jumbune (now LeapLogic)/** Impetus Technologies, India

An AI powered Data Lake Modernisation and Observability platform. Product helped customers with analytics and executing strategies to modernise and transform their analytics. Product has numerous niche features in space of Data Lake modernisation, AIOps, DataOps and APM, leveraged by many enterprise customers.

- **Visioning:** Incepted the product, laid down vision, core layers of the product, strategized offerings
- **Development:** Implemented many core and challenging functionalities throughout 11 years - *Spark, Hadoop, Java, Hive, Tez, Machine Learning, AWS (S3, EMR, Redshift, Athena, Glue, Lambda), Azure (HDInsight), GCP (DataProc), Cloudera CDH, CDP, HortonWorks, DataBricks, MapR, Python*
- **Lead:** Lead the teams throughout lifecycle: development, Quality, UI, UX, Releases, Roadmap
- **Evangelism:** Covered the product across all Social Media platform, Events, Webinars, Blogs
- **Pre-Sales & Sales:** Converting leads/prospects to customers
- **Management:** Covered Competitor analysis, critical thinking, Operating space, Customers Deliveries, Recruitment, Team management
- **Customer Success:** Heading the PS team to ensure Paid and POC customers succeed with us.
- **Partnerships:** Created partnership with hyperscalers, SI partners and data platform leaders; AWS, Databricks, IBM, Zensar, TCS

Customers: **Micron, Navistar, Adobe, L Brands, United Airlines**

JUL 2015 – JUN 2019

Sr. Technical Architect / Impetus Technologies, India

Helped customers succeed their strategic vision through technical leadership. Mentored Architects. Helped customers in visioning future modernised use cases through technology.

- **Established** Big Data Strategy of a USA based Intercontinental Airline Company
 - Implemented Hortonworks based Data Lake for business analytics
 - Implemented Named Entity Recognising (NER) for customer feedback forms during travel
 - Using Stanford NLP library to tag who, where, when
 - Implemented Sentiment Analysis for travel journeys across different sectors, different travel seasons.
 - Implemented seat wise pricing recommendation analytics on customer data
 - - *HortonWorks HDP, Hive on Tez, Spark, Stanford NLP, Python* Customer: **United Airlines**
- **Formulated** Multi Cloud (Azure+GCP) & Hybrid Strategy for a Billion-dollar healthcare giant - *Analytics and Storage Services of Azure, GCP, CDH, Customer: McKesson*
- **Realised** migration of Data Lakes to Cloudera CDH Bare metal for a Billion-dollar healthcare giant - *Spark, Hadoop, MapReduce, CDH, Ranger, Hive, HBase, Customer: McKesson*
- **Architected** Deep Learning based solution for information extraction from P&ID document for a USA based Oil & Gas giant using TensorFlow - *RNN, TensorFlow, Python, Customer: Chevron*

JUL 2012 – JUN 2015

Technical Architect / Impetus Technologies, India

Architected robust, extensible technical solutions.

- **Established** Enterprise Information Architecture & Data Governance on Hybrid Data Lake for a Pharma Giant - *Cloudera Navigator, CDH, Customer: Bristol Myers Squibb*
- **Consulted** Security & Governance on Data Lake for a US based Telecom provider - *HDP, Atlas, Ranger, Customer: Verizon*
- **Implemented** an automated Azure based Platform as a Service for a healthcare giant - *Azure, Azure Data Lake, Azure Blob storage, IaaS, Azure Analytical services, HDInsight, Customer: Kaiser Permanente*
- **Architected** & Implemented Routing Engine based analytical platform for a Media company - *Apache Camel, Java, Customer: Springer*
- **Architected** & Implemented Use Cases & Data Pipeline architecture for a US based avatar based social networking company - *Pig, Hive, Machine Learning, Cloudera CDH, Customer - IMVU*
- **Architected** & Implemented Data Migration solution for a US based company that provides information for the internet, telecommunications. - *Cassandra, Java, Hadoop, Hive, Customer: Verizon*
- For scaling required **Architected** a solution for migrating Neo4J to Janusgraph to scale to millions of nodes

DEC 2010 – JUN 2012

Lead Software Engineer/ Impetus Technologies, India

Implemented, Lead production grade solutions

- **Architected** & Implemented Data Fabric based Virtualisation layer for a recurring revenue generation organisation - *Java, Apache Camel, RabbitMQ, Customer: Service Source*
- **Implemented** a Data Replication solution for UIDAI project of **Govt. of India** - *Hadoop, Hive*
- **Architected** & implemented scalable solution for analytics on number portability data using for a US based organization - *Hive, Hadoop, Base, Java, Customer: Neustar*
- **Architected** & Implemented an ESB based gateway solution over clustered storages for a US based business & financial software company - *Java, MuleSoft, Customer: Intuit*
- **Architected** & implemented a platform of a social enterprise-wide content publishing framework for a British multinational enterprise software company - *Java, JAXRS, JPA, MySQL, Customer: Sage Software*

JAN 2009 – DEC 2010

Lead Software Engineer/ Prokarma Softech, India

Customer: Union Pacific

Enterprise Shipment Services was responsible for consuming shipment request from customer, generating waybill and providing shipment services to other modules.

My key responsibilities were:

- Created highly concurrent service for ESS customers which can cater huge number of requests extracting data from large set of RDBMS tables from Shipment Master.

Integrated ESS to Enterprise Event Bus.

- Given prompt solutions to the issues faced by ESS XMF services clients.
- Migrated EJB based services from Java EE container to Linux based container

Technologies: XMF (Proprietary ESB and WS), Java SE 5, Java Concurrency, XML, spring

JAN 2007 – JAN 2009

Senior Software Engineer/ Pramati Technologies, India

PERFORMANCE IMPROVEMENT IN MESSAGE TRANSMISSION OPTIMIZATION MECHANISM (MTOM) PROTOCOL IMPLEMENTATION FOR APACHE SOFTWARE FOUNDATION (ASF).

MTOM defined by W3C to send attachments as Binary Content with XML data during Web Service communication. MTOM sends the data in binary format instead of Base64 format, hence saved around 30% on wire transmission content.

Technologies: Apache CXF, MTOM, XOP, XML, JAXB, Java SE 5

WEB SERVICE * IMPLEMENTATION AND INTEGRATION OF APACHE CXF WITH PRAMATI JAVA EE APPLICATION SERVER

My key responsibilities were:

- Integration of WSS4J with Java EE server
- Enhancing WSS4J to provide rich support for WS-Security
- Writing WS-Security Policy support in Java EE server
- Support for WS-Reliable Messaging in Java EE server
- Support for WS-Trust in Java EE server
- Resolving issues with JAXB and passing TCK

Technologies: Apache CXF, Apache WSS4J, WS-Security, WS-Security Policy, XSL, XML, WS, Java SE 5

IMPLEMENTATION OF JAVA EE 1.5 JSR: INTEGRATION OF JPA 1.0 IN PRAMATI JAVA EE APPLICATION SERVER 6.0

Java Persistence API 1.0 introduced in Java EE 1.5, provides POJO based data persistence API. This API supports development of persistence application on both Java SE and Java EE environment. For Java EE 1.5 certified application server, JPA 1.0 support in application server is mandatory. My key responsibilities were:

- Implementing the Java EE container responsibilities for JPA
- Providing support for development and deployment of JPA projects in Java EE container
- Container support for persistent context, transaction propagation
- Clearing all TCK certification successfully

JAN 2006 – DEC 2006

Member of Technical Staff / Sun Microsystems, India

Project was Reference Implementation (RI) of Web Services Interoperability Technologies having XML and Web Services Security. XWSS was about securing XML and Web Service communication. The project involved securing XML based SOAP message; it involved providing message level security to XML. One can encrypt and/or sign any content/element/header/body of soap message using WS Security and that can be configured using WS Security Policy specification. Over all, the XWSS project used 4 specifications: WS-Security, WS-Security Policy, WS-Trust, and WS-Secure Conversation.

My key responsibilities were:

- Implementing new features and JSR specifications
- Adding support for new security tokens specifications published by OASIS, namely, X509 1.1 profile, Username Token v 1.1 profile,
- Communicating with users on Forums, Bug fixation tasks, new feature development along with writing their unit test and end-to-end test,
- Participation in interoperability plug-fests with other WS vendors,
- Closely working with global development teams across world on WS* technologies.

Technologies: Java, WS-Security, WS-Security Policy, Glassfish Application Server, WS*

JUN 2004 – JAN 2006

Member of Technical Staff / Persistent Systems, India

Product provided Enterprise Information Integration (EII) capabilities inside an Enterprise Service Bus.

The subproject data service platform was intended to provide logical data model created over heterogeneous physical data sources with querying and publishing facilities over logical data model.

My key responsibilities were:

- Development of EJB and Web services to serve as backend
- Evaluation of XQuery for various scenarios
- Proposal of RIA framework to BEA Systems, implementation of prototypes
- Evaluation of competent EII products based on various scenarios and matrices.

Technologies: XScript, XPath, XQuery, Web Services, Java, EJB, BEA Weblogic 8.1, Composite Information Server, MetaMatrix

Education

JUN 2004

M.Tech. (Information & Communication Technology) /

DA-IICT, Gandhinagar, India

JUN 2002

B.E (Hons.) (Computer Science) / JIT, Khargone, India

Research Activities

- “Towards providing seamless services from mobile devices using CORBA Servants in Mobile Ad Hoc Environment” , IEEE/ACM Sigmobility International Conference on Communication Systems Software and Middleware (IEEE COMSWARE 2006), 8-12 January 2006, New Delhi, INDIA
Link: <http://www.comsware.org/Comsware2006/techprog.html#t3>
- “Design and Implementation of an architecture supporting Mobile CORBA Servants under Intermittent Connectivity Environment”, IEEE Wireless Communication and Networking Conference (IEEE WCNC-2005), 13-17 March 2005, New Orleans, LA, USA
Link: <http://ieeexplore.ieee.org/iel5/9744/30731/01424893.pdf?tp=&arnumber=1424893&isnumber=3073>
- “Architecture for Locating Mobile CORBA Objects in Mobile Wireless Environment”, Fourth International Symposium and School on Advanced Distributed Systems (ISSADS-2004), 24-30 January 2004, Guadalajara Jalisco, Mexico. LNCS Volume 3061, in Springer-Verlag Heidelberg
Link: <http://www.springerlink.com/link.asp?id=5bnad1lmy1jrpwb2>