Varun Verma

+91-9996470791 | vermavarun@outlook.com | linkedin.com/in/varunve | github.com/vermavarun

OBJECTIVE

Results-driven Full Stack Developer and Cloud Architect with 13+ years of proven expertise in designing, architecting, and deploying enterprise-scale AI/ML solutions and distributed systems. Specialized in MLOps, cloud infrastructure automation, and building high-performance systems. Track record of delivering business impact through innovation: reduced operational overhead by 97%, improved ML model relevance by 20-25%, architected systems serving 10M+ global users, and contributed $180K+ in business value.

EXPERIENCE

Shell, Bangalore
April 2022 – Present

Senior Technical Lead

- Architected and deployed production MLOps pipelines using AWS SageMaker and Apache Airflow, enabling automated model training and validation with 99.9% uptime SLA.
- Developed Python and .Net SDKs for Confluent Kafka APIs with OAuth2 authentication, processing 10M+ real-time events daily with disaster recovery capabilities.
- Created ML-powered extraction logic using AWS Textract and custom NLP models, converting unstructured PDFs to structured data with 94% accuracy, processing 100K+ documents monthly.
- Reduced operational toil by 97% (9 hours to 5 minutes) through automated infrastructure provisioning with Azure DevOps and GitHub Actions.
- Designed Apache Airflow pipelines integrating edge device data to centralized ML platforms, eliminating third-party costs and enabling real-time analytics.
- Developed automated AWS multi-account provisioning using infrastructure-as-code, reducing setup time from days to minutes across 50+ accounts.
- Built backend microservices with enterprise SSO (SAML 2.0, OAuth2) securing access for 1000+ users with 99.99% availability.

Deloitte, Gurgaon
February 2021 – April 2022

Manager (Architect)

- Designed AI-enhanced microservices architecture for fintech client, integrating ML-based fraud detection models with real-time payment processing pipelines.
- Architected multi-region ML model replication service with end-to-end encryption, achieving 99.99% availability across 5 geographic regions.
- Developed custom Terraform provider and modules codifying 500+ on-premises infrastructure resources, reducing manual configuration by 80%.
- Strategized and secured $180K business through innovative AI/ML consulting solutions and enterprise architecture recommendations.
- Automated GitHub and SonarQube onboarding using Azure Logic Apps, reducing setup time from 3 days to 30 minutes for 200+ developers.

Nagarro, Gurgaon
June 2014 – February 2021

Associate Staff Engineer

- Led architecture and implementation of ML-optimized search platform, improving relevance by 25% through custom ranking algorithms and embedding-based similarity search.
- Designed and deployed intelligent chatbots using MS Bot Framework and Google Dialogflow with NLP, handling 50K+ customer interactions monthly with 85% resolution rate.
- Mentored 4+ engineers in full-stack development, unit testing, ML fundamentals, and cloud-native architecture patterns.
- Automated data update processes using orchestration frameworks, reducing manual effort by 70% and improving data freshness to near real-time.

Amdocs India Limited, Pune
Aug 2012 – April 2014

Subject Matter Expert

- Configured and implemented enterprise SharePoint Search.
- Created web portals using front-end technologies.
- Developed WCF Windows Service to expose portal data to consumers.

PROJECTS

Created dotnet and python SDKs for Confluent Kafka with OAuth
- Developed SDKs to simplify integration with Confluent Kafka APIs for trading applications.
- Implemented features to enhance performance and reliability of data streaming.
- Disaster Recovery Capabilities with active cluster switching details.

Cross Organization Information Analysis
- Architected data integration platform ingesting from 20+ APIs, processing 500K+ records daily with 99.8% accuracy using Apache Airflow.
- Implemented ML-based data quality framework detecting anomalies with 92% precision using isolation forests and statistical methods.
- Developed transformation pipelines using Azure Data Factory and Databricks, handling multi-terabyte datasets with complex ETL operations.
- Built automated Power BI dashboards refreshing every 4 hours, providing real-time insights to 200+ stakeholders.
- Integrated Azure Translator APIs enabling document processing in 25+ languages, expanding reach to global markets.
- Implemented RBAC using Azure AD, securing sensitive data access for 500+ users across 5 business units.

Wind Farm Power Generation NLP
- Built end-to-end NLP pipeline extracting structured data from unstructured PDFs using AWS Textract and custom ML models with 94% accuracy, processing 10K+ documents.
- Developed SageMaker-based ML models for automated document classification, entity extraction, and validation, reducing manual data entry from 40 hours to 2 hours weekly.
- Implemented MLOps infrastructure with automated model training, validation, A/B testing, and deployment, reducing time-to-production from 2 weeks to 1 day.
- Built real-time monitoring dashboards with CloudWatch, achieving 99.5% SLA compliance for data pipeline availability.
- Integrated data quality checks using statistical methods and isolation forests, catching 95% of data quality issues before production.
- Streamlined quarterly reporting for 50+ wind farm clients, reducing manual reporting time by 85%.

Energy Savings Marketing Portal
- Designed ML-powered recommendation engine using collaborative filtering, increasing vendor match accuracy by 35%.
- Optimized API performance through caching and ML-based prediction, reducing latency from 800ms to 680ms (15% improvement).
- Implemented enterprise SSO with SAML 2.0 and ML-optimized search results, improving user conversion by 22%.
- Automated deployment using Azure DevOps, reducing deployment time by 30% and enabling 10+ daily releases.
- Enhanced security with OAuth2, TLS 1.3, and ML-based fraud detection achieving 98% precision.
- Achieved 99.95% service availability through multi-region active-active deployment and automated failover.

Hyper Automation Templates with AI/ML
- Engineered boilerplate ML project templates with pre-configured frameworks (TensorFlow, PyTorch), reducing setup time by 40% for 150+ developers.
- Integrated AWS Textract and custom OCR libraries (Tesseract, EasyOCR) for Python and Node.js across 20+ enterprise applications.
- Built end-to-end CI/CD pipelines with automated testing, model validation gates, and performance benchmarks using GitHub Actions.
- Implemented Infrastructure as Code templates for quick ML environment provisioning (notebooks, GPU clusters, monitoring).
- Created documentation adopted across organization, reducing developer onboarding from 2 weeks to 3 days.

Enterprise Search with ML Optimization
- Built large-scale search platform serving 10M+ monthly users globally, handling 100K+ concurrent searches with sub-100ms latency.
- Implemented ML-based ranking using LambdaMART and neural networks, improving relevance by 20% and user satisfaction by 28%.

- Developed custom connectors integrating 15+ heterogeneous data sources with automated sync pipelines.
- Built semantic search using BERT and sentence transformers, achieving 35% improvement over keyword search.
- Implemented faceted search, AI-powered autocomplete, and personalized recommendations using collaborative filtering.
- Achieved 99.9% availability using distributed Elasticsearch clusters across 3 regions with sophisticated caching reducing load by 60%.

TECHNICAL SKILLS

Languages: Python (primary), Go (Golang), C#, .Net, Java, C++, Node.js, TypeScript, JavaScript, R, SQL, HTML/CSS.

ML/AI Frameworks: TensorFlow, PyTorch, Scikit-learn, XGBoost, LightGBM, Keras, Hugging Face Transformers, spaCy.

ML Platforms: AWS SageMaker, Azure ML Studio, Google AI Platform, Databricks, Weights & Biases, mlflow, Kubeflow, Ray.

NLP/LLM: AWS Textract, Langchain, LLamaIndex, FAISS, embedding models, fine-tuning, RAG, prompt engineering, Transformers.

Cloud Platforms: AWS (SageMaker, EC2, S3, Lambda, RDS, Textract, DynamoDB, CloudWatch), Azure (ML, Data Factory, Functions, App Service), GCP (AI Platform, BigQuery).

Data Engineering: Apache Airflow, Apache Spark, Pandas, NumPy, PySpark, Delta Lake, Azure Data Factory, Databricks.

Containers/Orchestration: Docker, Kubernetes, EKS, AKS, Helm, Docker Compose.

Infrastructure as Code: Terraform, ARM Templates, Bicep, CloudFormation, Ansible.

Databases: PostgreSQL, MongoDB, Cosmos DB, MSSQL, Oracle, DynamoDB, Redis, Elasticsearch.

DevOps/MLOps: GitHub Actions, Azure DevOps, GitLab CI, mlflow, DVC, Prometheus, Grafana, ELK Stack.

Identity & Security: FIDO, WebAuthn, SSO, SAML 2.0, OAuth2, JWT, Azure AD, identity management protocols, MFA.

Web Frameworks: React, Angular, .Net Core, ASP.Net, Express.js, FastAPI.

Tools: Visual Studio Code, Git, Jira, Postman, TensorBoard, Jupyter, VS Studio, Confluence.

OS: Linux (Ubuntu, CentOS), macOS, Windows.

EDUCATION

MMU, Ambala, India
2008 – 2012

Bachelor of Technology [B.Tech] in Computer Science and Engineering

KEY ACHIEVEMENTS

Operational Excellence: Reduced operational overhead by 97% (9 hours to 5 minutes) through intelligent automation

ML Impact: Improved search relevance by 20-25% through advanced ML algorithms and semantic search

Scale & Performance: Architected systems serving 10M+ users with sub-100ms latency and 99.99% availability

Business Value: Contributed to $180K+ business revenue and cost savings of 60%+ through innovation

Data Intelligence: Built pipelines processing 500K+ records daily with 99.8% data quality

Team Leadership: Mentored 50+ engineers in cloud architecture, ML, and DevOps best practices

Technology Innovation: Designed and deployed 15+ production ML models with 80-98% accuracy metrics

PERSONAL INFORMATION

Father's Name: Late Sh. Baldev Raj Verma