

Pranali Chipkar

📍 Boston, MA 📞 8573985923 ✉ chipkar.p@northeastern.edu 🌐 in/pranali-chipkar 📄 /PranaliChipkar 📄 /Pranali_Chipkar

EDUCATION

Northeastern University, Boston USA

Master of Science in Computer Software Engineering

Expected December 2025

- Relevant Coursework: Application Engineering Development, Data Science Engineering Methods, Big Data Application and Intelligent Analytics, Business Process Engineering

University of Mumbai, Mumbai, India

Post Graduate Diploma in Information Technology

July 2022

- Relevant Coursework: Software Engineering, Advance Java, Linux, Web Programming, Networking

Bachelor of Science in Information Technology

Oct 2020

- Relevant Coursework: C, C++, Core Java, Database Management Systems, Python, Operating Systems, Embedded Systems

PROFESSIONAL EXPERIENCE

nChroma Bio, Boston, USA

Jan 2025 – Jun 2025

Data Science Co-op

- Developed 5+ automated bioinformatics pipelines processing 100+ GB of genomics data, reducing manual processing time by 80% through cloud-based DNA methylation analysis and protein structure prediction workflows
- Built scalable AWS Batch orchestration systems handling 50+ concurrent jobs with dynamic resource allocation, implementing Nextflow-based BAM sampling pipelines and AlphaFold protein folding workflows for drug discovery research
- Engineered automated data backup systems using AWS Lambda and Step Functions to seamlessly transfer sequencing data from BaseSpace to S3 with real-time notifications and 99%+ pipeline reliability
- Developed PostgreSQL-based Benchling integration for metadata retrieval across 200+ samples and implemented CloudFormation notification infrastructure monitoring 10+ job queues with automated email alerts
- Utilized Python, Bash scripting, Docker containerization, and bioinformatics tools (Bismark, Samtools, MultiQC) to integrate multiple platforms (BaseSpace, Benchling, GitLab, AWS) delivering cost-effective cloud computing solutions

INFOSYS, INDIA

Nov 2020 – Aug 2023

Senior Systems Engineer

- Engineered 40+ automated data processing scripts using PowerShell and Python, streamlining data migration workflows and improving processing efficiency by 25% for enterprise scale database systems
- Performed database administration activities including backup, restore, indexing, and performance optimization for SQL Server environments supporting 600,000+ user records at Siemens Gamesa
- Implemented comprehensive data verification and troubleshooting solutions, resolving data conversion issues and maintaining detailed issue resolution logs, resulting in 30% reduction in data-related errors
- Designed and maintained database objects to support data migration utilities and projects, improving data quality and operational insights by 20%
- Integrated multiple heterogeneous data sources (ServiceNow, SolarWinds, Nagios) using SQL and custom scripts, creating unified data solutions achieving 99.9% system reliability
- Developed data analysis solutions to identify data patterns, anomalies, and inefficiencies in data processing workflows, implementing automated validation scripts for improved data integrity

PROJECTS

Brazilian E-commerce Data Analytics Pipeline

Aug 2025

- Designed scalable Azure cloud data pipeline processing 500K+ Brazilian e-commerce records using Azure Data Factory, Databricks, and Azure Synapse Analytics, implementing medallion architecture for efficient data transformation and analytics
- Implemented hybrid data storage solution combining MySQL and MongoDB with automated PySpark transformations, achieving seamless data integration across 8 datasets and enabling advanced delivery performance analysis
- Created production grade data warehouse with external tables in Azure Synapse Analytics, supporting real-time business intelligence reporting and PowerBI dashboard integration for stakeholder decision-making

Research Pipeline and Multi-Agent System

Nov 2024

- Constructed end-to-end ML data pipeline using Apache Airflow for document processing and vector-based search, managing large-scale datasets with 40% faster retrieval performance
- Implemented scalable data ingestion strategy with automated document parsing and Pinecone vector storage, enabling efficient similarity search for automated ML-driven content analysis
- Developed multi-agent ML processing system using LangGraph with specialized agents for research paper analysis, demonstrating advanced ML framework integration and automated model orchestration

Intelligent Document Pipeline and Interactive Search System

Oct 2024

- Created comprehensive data migration solution using Apache Airflow to scrape and convert CFA Institute publications data, storing structured data in Snowflake and assets in AWS S3
- Designed and implemented database schema with appropriate tables, indexes, and constraints to support document metadata and content storage requirements
- Developed stored procedures and functions for data transformation and migration processes, ensuring data conversion solutions met technical specifications and client requirements.
- Implemented automated data quality checks and anomaly detection algorithms

TECHNICAL SKILLS

Languages: Python, SQL, Java, PowerShell, Bash Script, R, C++, JavaScript, JSON

Big Data & Analytics: Apache Airflow, Apache Spark, Snowflake, Google BigQuery, Databricks, Data Warehousing, Power BI, Tableau

Cloud Platforms & DevOps: AWS (S3, Lambda, Batch, Step Functions, CloudWatch, IAM, ECR), Azure (Data Factory, Synapse Analytics, Microsoft Fabric), Git, GitLab, CI/CD Pipelines, Docker, Containerization, API Development

Data Engineering: ETL Pipelines, Data Modeling, Data Transformation, dbt, Workflow Automation, Data Product Development, Data Validation

Databases: MySQL, PostgreSQL, MongoDB, Pinecone, Milvus, Benchling, Oracle, SQL Server

Libraries: Pandas, NumPy, scikit-learn, seaborn, matplotlib, PyTorch, Boto3, Streamlit, Statistical Analysis, A/B Testing, Bismark, Samtools, FastQC