SRIKANTHBELLARY

SR.DATAENGINEER/ARCHITECT(GCPCertifiedProfessionalDataEngineer)

# ProfessionalSummary

<https://www.linkedin.com/in/srikanth-bellary/>

varun.kothval@interaslabs.com

(614)-992-8701

* **15+ years** of consulting experience in **Business Systems Analysis**, **Data Engineering**, **Data Architecture**,**SolutionArchitecture**,**CloudComputing**,**MachineLearning**,**AI**,and**Functional Programming** with master’s degree in software engineering
* Expertisein**consolidating**,**integrating**,and**migrating**EnterpriseDatatoCloud-basedDataLakes
* Experiencedindesigninganddeveloping**Hadoopecosystem**solutionsforBigDataAnalyticsand Business Intelligence.
* Proficiencyin**AI/ML/LLMsolutions**(ChatGPT,Gemini,Claude)includingpromptengineeringand production integration for significant business optimization
* **ExpertinimplementingRAGarchitecture**withvectordatabasestoenhanceLLMcapabilitieswith custom use cases and internal guardrails in live production systems
* **DemonstratedsuccessindevelopingAgenticAIworkflows**thatsegregatebusinesslogicfrom technical features
* Strongexperiencedevelopingbigdatapipelinesusing**Spark**(Scala/PythonAPI)withRDD,SQL, Datasets, Data Frames, Streaming, and MLlib
* Hands-onwith**Hadoopecosystem**components(HDFS,MapReduce,Hive,Sqoop,Oozie)
* Experiencewith**SOA**,**Microservices**,and**ServerlessLambdaArchitecture**
* Deepunderstandingof**RDBMS**,**OLTP**,**OLAP**,**EDW**,**ETL/ELT**,**NoSQL**,and**DataGovernance**
* Experiencein**StrategicEnablement**includingCapabilityMaturityModeling(CMM)
* Skilledin**CloudMigration**and**PlatformAutomation**acrossmajorcloudproviders
* Effectivecommunicationskillswithabilitytoengageeffectivelywith**C-level**and**technicalteams.**

# CoreCompetencies

* **DataArchitecture&Engineering:**DataModeling,ETL/ELTPipelines,DataLakes/Warehouses,Data Governance
* **ArtificialIntelligence:**LLMImplementation,RAGArchitecture,AgenticAI,PromptEngineering, Vector Databases
* **MachineLearning:**ModelDevelopment,FeatureEngineering,ModelTraining/Inference,Ensemble Methods
* **CloudTechnologies:**Multi-CloudArchitecture(AWS,Azure,GCP),IaaS,PaaS,SaaSImplementation
* **BigDataEcosystem:**Hadoop,Spark,Kafka,StreamingAnalytics,BatchProcessing,Real-timeData Processing
* **DataAnalytics&Visualization:**BusinessIntelligence,PredictiveAnalytics,DashboardDevelopment
* **DatabaseManagement:**SQL/NoSQLSolutions,PerformanceTuning,DataMigration,SchemaDesign
* **EnterpriseIntegration:**APIDevelopment,MicroservicesArchitecture,Event-DrivenSystems
* **DevOps&MLOps:**CI/CDPipelines,InfrastructureasCode,Containerization,Orchestration
* **ProjectLeadership:**StakeholderManagement,TechnicalTeamLeadership,ArchitectureReview Boards

# TechnicalSkills

|  |  |
| --- | --- |
| **Gen AI/LLM** | ChatGPT,Llama,Gemini2.0Flash,Claude3.7Sonnet,Perplexity&DeepSeekR1 |
| **ML/AI** | PyTorch, TensorFlow,Keras, Random Forests,XGBoost, Clustering,NER,NLP |
| **Hadoop** | Hadoop,HDFS,MapReduce,YARN,Pig,Hive,Spark,Sqoop,Kafka |
| **CloudTechnologies** | **GCP:**GCS,BQ,DataProc,VertexAI,CloudCompose,DataFusion**AWS:**S3,EC2,EMR,EKS,Glue,Redshift,Kinesis,Lambda,Athena**Azure:**BlobStorage,ADL,DataFactory(ADF),Databricks,DeltaLake |
| **DevOps/Infra** | Jenkins, Maven,SBT,Terraform,Kubernetes,CloudFormation |
| **Programming** | SQL,Java,Scala,Python,Unix,HTML,CSS,JavaScript,XML,JSON,REST |
| **Tools** | VSCode,Cursor,Copilot,IntelliJ,PyCharm,Git, WinSCP,Putty, Power Shell |
| **Databases** | AWSRDS,Redshift,MongoDB,PostgreSQL,Oracle,MySQL,Teradata |
| **ETLandVIZ** | Streamlit,Tableau,Spotfire,Cognos,SAPBO/BI,Informatica |

**Certification**

* + **GoogleCloudCertifiedProfessionalDataEngineer**(2023-2025)

# Education

* + **MasterofScience(MS)**inSoftwareEngineering(2009)
	+ **BachelorofTechnology(BTech)**inMechanicalEngineering(2007)

**ProfessionalExperience**

# InterasLabs|Remote

Sr.Data,ML&AIEngineer–SolutionArchitect|*May2024–Present*

WorkingasatechnicalconsultantwithmultipleretailclientsatInterasLabs

* + LeadroletoserveasSolutionArchitectandProductOwnerImplementingML,NLP,LLM,andAI solutions in production for Retail clients to streamline the business processes
	+ SecuredGooglepartnershipforInterasLabsinGCPServicesandGoogleWorkspacedomains.
	+ ProvidedSolutionDesignforAI-enabledRetailattributemappingengineusingVertexAI
	+ Transformedlabor-intensiveattributemappingprocessesthroughAIautomation,reducing operational costs by 60% and increasing throughput by 3x
	+ BuiltscalableDataPipelinesusingGCS,BigQuery,DataProc,SparkandAirflowfordatasourcing
	+ BuiltoptimizedMLDataPipelinestocreateTraining,LabeledandFeatureEngineereddatasets
	+ BuiltintelligentAIPipelinestomaprawattributestotargetdictionarywithoutusingSQLorETL
	+ IntroducedConfidenceScoringmechanismandreasoningframeworksatindividualattribute level to support and build trust in AI/LLM based solutions
	+ Analyzedpromptengineeringtechniquestoenhanceconfidencescoringaccuracy
	+ UsedOCRenginetoextractretailproductdetailsfromproductimageswithconfidencescoring
	+ ImplementedNERandNLPalgorithmsaspreprocessingstepstomaptheattributestodictionary
	+ CreatedAgenticworkflowswithVertexAIRAGEngineformapping,inference,andvalidation
	+ DevelopedmultiplePOVs/POCstodemonstratesuccessfulintegrationofML,AI,GenAI,LLMand RAG based solutions in production along with business specific custom guard rails.

**Environment:**GCP,GCS,BigQuery,Spark,Airflow,VertexAI,LLM,RAG,MongoDB,Agents,Python,SQL, Golang, Streamlit, Claude 3.7 Sonnet, Lang chain

# ThermoFisherScientific|PersistentSystems|Pittsburgh,PA

Sr.ML/AIConsultant–SolutionArchitect|May2023–May2024

CustomerChannelsGroup(CCG)isThermoFisher’slargescaleinitiativetodrivereal-timedataforfaster action and insights and deliver value from the Legacy Mainframe data migrated to AWS cloud.

* + ProposedML/AIbasedsolutionsforCustomerChannelsGroup(CCG)atThermoFisherScientific to migrate Legacy Mainframe data to AWS cloud.
	+ Designedandimplementedrobustrelationaldatamodelstohostnon-relationalmainframes data which is hierarchical and sequential.
	+ DevelopedML/AIpipelinestomapcomplexmainframedataformatslikeVSAMandISAM,flat- file formats to ThermoFisher specific internal business entities.
	+ WorkedcloselywithLeadsfrommultiplechannelsatCCGtogathertheirbusinessusecases,

extractdomainspecificdatafromMFandthenprovidecleanandstructureddataasdeliverable

* + ExtensiveanalysisonCOBOLandDARSTRANlogicusingAI/LLMsolutionsandconvertitto SQL
	+ BuiltrobustdatavalidationframeworkforreconciliationofdatabetweenMFandAWSDataLake.

**Environment:**Mainframes,VSAM,DARSTRAN,AWS,PostgreSQL,SQL,Erwin,Glue,S3,IAM,EC2, Databricks, Athena, SageMaker, Kafka, Python

# CVSHealth|Woonsocket,RI

Sr.ML/AIAdvisor–DataEngineeringLead|Aug2019–April2023

RetailPharmacyArtificialIntelligence(RPhAI)isamajorenterpriselevelbusinessoptimizationinitiative byCVSHealthtoleverageMachineLearningandArtificialIntelligenceTechniquestoResolvePharmacy Claim Rejections by the PBM’s.

* + LeddataengineeringteamforRetailPharmacyArtificialIntelligence(RPhAI)initiativeto optimize pharmacy claim rejections using ML/AI solutions and implementations.
	+ LedMLDatasolutionsforRPhAIappsusingCDP,EDW,RxClaims,EPICEHRandClinicaldata
	+ SecuredARBapprovalsforsolutionswhichconsolidatepayer,provider,andpatientdata
	+ MigrateddataassetsfromCVS-Aetna'sAzureCloudPlatformtoGoogleCloudPlatform(GCP)
	+ Designedandimplementedsolutionsforcloudmigrationswithminimalcodechanges
	+ Developedandmaintainedend-to-endMLpipelinesandproductiondeployments
	+ WorkedexclusivelyonMLmodeltraining,generatinglabelleddataandfeatureengineereddata for model training and aggregating real-time look up data for retrieval and model inference
	+ BuiltandmaintainedlookupdatacollectionsinMongoDBusedformodelinference
	+ CollaboratedwithDataScientistsonmodelselection,hyperparametertuningandA/Btesting
	+ RolledoutregularML/AIenabledfeaturestotheRPHAIappslikeRxOptimizerandRxResolver
	+ BuiltandmaintainedDataandMLpipelinessupportingtheRPHAIappsformultipleyears.
	+ ImplementedOpticalCognitiveRecognition(OCR),NameEntityRecognition(NER)andNatural Language processing (NLP) techniques to extract data from Prescriber notes.
	+ BuiltCI/CDpipelinesforMLproductiondeploymentframeworksusingJenkins,Git,andAirflow
	+ Sourcedreal-timedatausingKafkaandSparkStreamingusedforModeltrainingandinference
	+ Integratedreal-timemodelfeedbackintoMLtraining,retrieval,andinference.

**Environment:**Azure,GCP,Spark,Kafka,Python,Scala,AdobeCDP,SaaS,EDW,Gitlab,Jenkins,Data Robot, Databricks, Synapse, ADF, Conda, Airflow+

# ChangeHealthcare|Chicago,IL

Sr.CloudDataEngineer/SolutionArchitect|Aug2017–Aug2019

TheIHDPputsCHCinauniquepositiontobeanauthoritativesourceforlargedatasetsfortheindustry by incorporating financial, clinical, and operational data combined with Artificial Intelligence and Machine Learning.

* + KeycontributortoIntelligentHealthcareDataPlatform(IHDP)combiningfinancial,clinical,and operational data with AI/ML.
	+ Translatedbusinessrequirementsintotechnicalspecificationsandapplicationcode
	+ DevelopedtacticalandstrategicsolutionsfortheIntelligentHealthDataPlatform
	+ Engineereddataingestionfrompayers,providers,institutions,andclearinghouses
	+ ProcessedlargevolumesofEDIclaimsdataforenterprisedatalakestorage
	+ BuiltSpark/ScalaapplicationstotransformEDItransactions(837,835,270/271)intotabular formats
	+ CreatedXSLTscriptstoconvertX12formatteddatatostructuredXML
	+ ImplementedDataRightsManagementandRBACmodelsforsecureclaimsdataaccess
	+ DevelopedAWSGlueETLjobsutilizingCrawlers,CatalogsandAthena
	+ ConductedresearchonMachineLearning,AIandBlockchaintechnologies

**Environment:**Spark,Scala,Hadoop,Hive,HDFS,AWS,S3,EMR,GLUE,SNS,Talend,SaaS,Confluence, Gitlab, Jenkins, Artifactory, Stylus Studio

# Cars.com|Chicago,IL

MachineLearningEngineer|Mar2017–Aug2017

Projectwastoresearch,experimentandproductionizeBigDataandMachineLearningtechniquesfor Predictive Analytics and Business Optimization.

* + LedBigDataMachineLearninginitiativesforPredictiveAnalyticsandBusinessOptimization.
	+ LedBigDataMachineLearningteamofDataAnalysts,ScientistsandEngineers
	+ IntegratedCustomerDataPlatformintoMLworkflowforcustomerbehavioranalysis
	+ ProductionizedMLModelsforproductionClouderaHDFScluster(CDH5.5.1)
	+ ImplementedRegressionandClassificationalgorithmsusingSparkandTensorFlow
	+ DevelopedensemblemodelsincludingRandomForest,XGBoost&GradientBoostedRegression
	+ DevelopedanddeployedproductiongradeMLpipelinesusingSparkScalaandPythonAPIs
	+ ImplementedCI/CDforMLpipelinesacrosstheenvironmentsforperiodicmodeltraining
	+ Builtcustomizeddatapipelinesfortrainingdata,labeleddataandfeatureengineereddata
	+ Experimentationwithhyperparametertuningtoimprovemodelefficiencyandaccuracy.

**Environment:**Cloudera,Kerberos,Spark1.6/2.0,SaaS,Scala,Python,Kafka,Hive,Hue,HDFS,AWS,S3, Sqoop, Couchbase, Confluence, Adobe CDP, Bitbucket, Jenkins, Artifactory

# McDonald'sCorporation|SapientRazorfish|Chicago,IL

LeadBigDataConsultant|Nov2016–Mar2017

StrategicEnablement(SE)andAccelerationforMcDonald’saspartoftheEnterpriseCloudMigrationand AutomationProgram. Key aspects includeCapabilityMaturityModel, BigDataArchitecture, Global Data Lake Setup, Streaming Data, IoT Data, Predictive Reporting (Machine Learning) and Global Market Solution for Rapid Roll-Out to over one hundred Countries.

* + StrategicEnablementforEnterpriseCloudMigrationandAutomationProgramtotransitionfrom Monolithic to Microservice architecture.
	+ LedBigDataTrackforCloudPlatformMigrationProgram
	+ CapturedCapabilityMaturityModelforCoreCustomerPlatformandGlobalDataPlatform
	+ ConductedStrategicEnablementworkshopswithChiefDataOfficeratMcDonald’s
	+ DevelopedPOCSparkApplicationsandreplicatedthemusingTalend/Spark,S3andEMR
	+ Evaluatedtechnologytools(WaterlineData,Paxata)forDataScienceandGlobalDataLake
	+ WorkedonSolutionsforGDPRRegulationstobeimplementedinEUregionbyMcDonalds
	+ DesignedframeworkforDataIntegrationadaptorsinMicroservices/APIEnvironment
	+ CreatedAgilestoriesforDataIngestion,GlobalDataLake,ETLMigration,andDataScience activities

**Environment:**AWS,S3,EMR,Kinesis,Talend,Tableau,Collibra,Waterline,Paxata,Teradata,Epsilon, Spark, Kafka, RabbitMQ, Java, ELK, Redshift, MapReduce

# NorthernTrustCorporation|Chicago,IL

Sr.BigDataConsultant|April2015–Nov 2016

TheDerivativesTransformationProgramatNorthernTrustaimedtoenhancethedataandtechnology infrastructure by incorporating cloud and big data technologies into the IT processes.

* + KeyconsultantforDerivativesTransformationProgramupgradingenterprisedataexchange format to ISO 20022 (Dodd Frank).
	+ LeddataintegrationofHadooptechnologieswithEnterpriseITApplicationsandDataLayer
	+ Engineereddatapipelinejobsfromproductioncluster(CDH5)toCentralDataLake(S3)
	+ DesignedEDWandDataMartarchitectureusingSqoop,Spark,Redshift,andTableau
	+ CreatedEnterpriseSecurityMonitoringdatapipelineusingAWSEMR,Spark,KafkaandS3
	+ DevelopedSpark2.0applicationsinScalausingSparkSQL,DataFrame,DatasetandDstream
	+ BuiltlargevolumedataprocessingSparkdatapipelinesfordatamigrationandenrichment
	+ CreatedPOCsforMachineLearningapplicationsinvolvinglinearregression,logisticregression, supervised learning, recommendation engines and predictive engines.
	+ WorkedonDevOps,CICD,RoleBasedAccessControls(RBAC)&DataRightsManagement(DRM)

**Environment:**Windows,Linux,Oracle11g,Mainframe,JDBC/ODBC,RESTfulAPI,AWS,Hortonworks, Cloudera,Spark1.6/2.0,SQL,NoSQL,HDFS,Talend,HBase,EMR,CSV,XML,JSON,Parquet,ScalaIDE, SBT, Oozie

# KeyBank|Cleveland,OH

SeniorConsultant–BigDataArchitect|September2014–April2015

EnterprisescaleinitiativebyKeyBanktoconsolidatetheEnterpriseDatafromvarioussourcesintothe Shared Foundation Data (SFD) and adopt big data technologies for Business Reporting, Management Reporting, Executive Reporting and Predictive Analysis.

* + ServedasaDataArchitectonSharedFoundationData(SFD)programusingBigDataTechnology
	+ Identifieddatasourcesforbatchprocessingandreal-timestreamprocessing
	+ ConfiguredClouderaManagerforclustersetupinStagingandTestenvironments
	+ DevelopeddatapipelineSparkapplicationsinPythonusingSparkContextandSparkSQL

**Environment:**Windows,Linux,Java,EclipseIDE,Hadoop,HDFS,Spark,Flume,Hive,Pig,Cloudera,AWS, EC2, EMR, Redshift, Teradata, JSON, Parquet, AVRO, ALM, Cognos, MS SQL Server

**PreviousProfessionalExperienceSummary**

# WellsFargo|DesMoines,IA

SeniorConsultant–WebServicesDataMigration|2013-2014

WFFDataMigrationProgram:TheprojectistomigratetheWellsFargoFinancialHomeLoanDatafrom Wells Fargo Supreme to Mortgage Servicing Big Data Platform.

* + LeddatamigrationinitiativesforWellsFargoFinancialHomeLoandatatoMortgageServicing Platform
	+ ManagedWebServices(50+)consumedbyvariousJava,.NET,andmobileapplicationsinSOA environments

# PrudentialFinancial|Shelton, CT

BusinessIntelligenceConsultant|2011-2013

PruInside-NewPolicyAdministrationPlatformtomanagestrategicmethodsandriskcompliancerules.

* + DesignedandimplementedBIreportingsolutionsandETLprocessesforPolicyAdministration Platform

# CognizantTechnologySolutions|Minneapolis,MN

BusinessSystemsAssociate|2009-2011

* + CreateddatawarehousingsolutionsusingKimballmethodologywithstarschemadesign

# CareerBuilderInc.|Reston,VA

ETLAnalyst|2008-2009

* + DevelopedETLmappings,SQLprocedures,andtransformationlogicforenterprisedatasystems

# UIAResearchandDevelopment|Hyderabad,India

ETLAnalyst|2006-2007

* + Ledcross-functionalteamsthroughallSDLCphasesandcollaboratedwithbusinessstakeholders

**Technologies:**Oracle,SQLServer,Informatica,ETL,SSRS,BusinessObjects,MicroStrategy,Data Warehousing, SOA, Web Services, UNIX, Windows