Srikanth Bellary

SR. CLOUD DATA ARCHITECT/ENGINEER (GCP Certified Professional Data Engineer)



www.linkedin.com/in/srikanth-bellary

rohitb@interaslabs.com

614-992-8709

Professional Summary

* Dedicated IT professional with 15+ years of consulting experience in Data Engineering/Architecture, Big Data, Solution Architecture, Multi-Cloud Computing, Machine Learning, Artificial Intelligence and Functional Programming with a Master of Science Degree in Software Engineering.
* Excellent communication skills, technical skills, managing skills and extensive client facing background in Healthcare, Pharmaceutical, Finance, Retail and .COM domains.
* Solid experience in consolidating, integrating, and migrating Enterprise Data to Cloud based Data Lake.
* Highly knowledgeable in architecting and engineering solutions powered by Hadoop ecosystem to embed ‘Big Data Analytics’ and ‘Business Intelligence’ into Technology Operations for Business Optimization.
* Proficient in Big Data processing, Real-Time Ingestion, Data Discovery/Exploration and Data Science.
* An AI enthusiast skilled in LLM, GenAI, Prompt Design, Prompt Engineering and Prompt Tuning.
* Experience in developing data pipeline Spark applications for production environments using Scala API & Python API with Spark modules like RDD, Spark SQL, Datasets, Data Frames, Spark Streaming and MLlib.
* Hands-on with Hadoop), HDFS, MapReduce, Hive, Sqoop, Spark, Kafka and Oozie.
* Hands-on with both manual and vendor supported – installation, configuration and maintenance of Big Data components and tools on enterprise grade - production clusters running on commodity hardware, public/private cloud infrastructure and hybrid cloud arrangements.
* Experience in SOA (SOAP), Micro-services (API’s) and Server-less Lambda Architecture Environments.
* Strong understanding of RDBMS, OLTP, OLAP, EDW, SQL, ETL/ELT, NoSQL, Big Data & Hadoop concepts.
* Experience in Strategic Enablement to the clients by means of Capability Maturity Modeling (CMM), providing Strategic solutions, Tactical acceleration and Thought leadership to adopt Design Thinking.
* Experience with Large-scale Enterprise Cloud Migration and Platform Automation Programs.
* Data Governance experience in financial regulations (Swift ISO 20022) and general regulations (GDPR).
* Ability to efficiently communicate, collaborate and engage both C-level and technology representatives.
* Experience in working with large teams and vendors on multiple concurrent technology projects.
* Good understanding of SaaS, PaaS, IaaS, DevOps and Continuous Integration (CI/CD)
* Well-developed troubleshooting skills, ability to maintain calm and perform under stressful conditions.

Technical Skills

|  |  |
| --- | --- |
| **Hadoop Ecosystem** | Hadoop, HDFS, MapReduce, YARN, Pig, Hive, Spark, Sqoop, Kafka |
| **ML/AI** | OpenAI, Spark ML, Spark MLlib, Tensorflow, Ensemble Modeling, GBT Regression, Random Forests, Natural Language Processing (NLP) |
| **Cloud Technologies** | AWS – S3, EC2, EMR, EKS, Glue, Redshift, Kinesis, Lambda and AthenaAzure – Blob Storage, ADL, Data Factory (ADF), Databricks, Delta Lake |
| **DevOps/Infra** | Jenkins, Maven, SBT, Terraform, Kubernetes, Cloud Formation etc. |
| **Programming/Scripting** | SQL, Java, Scala, Python, CLI shell scripting |
| **Web Technologies** | XML, HTML, XSLT, SOAP, REST, JavaScript and JSON |
| **Tools** | Git, BitBucket, WinSCP, Putty, mRemoteNG |
| **Databases** | AWS RDS, Redshift, MongoDB, PostgreSQL, Oracle 10g/11g, MySQL and Teradata |
| **ETL and Visualization** | Tableau, Spotfire, Cognos, SAP BO/BI and Informatica |

Certification

* [Google Cloud Certified Professional Data Engineer](https://google.accredible.com/3f2a6a59-b7e5-4bd9-b859-6c969bf213dc) *2023-2025*

Education

* Master of Science (MS) Degree in Software Engineering *2009*
* Bachelor of Technology (BTech) Degree in Mechanical Engineering *2007*

Professional Experience

**Persistent Systems/ThermoFisher Scientific**

**| Pittsburg, PA**

Sr. Cloud Data Architect May 2023 – Present

Persistent Systems is playing a key role in building and supporting the Customer Channels Group (CCG) at ThermoFisher Scientific. CCG is a large-scale initiative to drive real-time data for faster action and insights and deliver value from the Legacy Mainframe data migrated to AWS cloud.

Responsibilities:

* Played lead role in driving solution architecture to migrate Mainframe data to AWS Cloud platform.
* Developed multiple data pipeline templates specific to each business product and feature.
* Implemented various Machine Learning techniques like Regression, Classification and Clustering.
* Data architecture and data modeling for Operational Data Store (ODS) supporting new features.
* Created solutions and POC demos to secure architecture review approvals before deployment.
* Worked on AWS – S3, EC2, IAM, EMR, DMS, Glue, Athena and Databricks for pipeline development.
* Created frameworks for Data Quality, Auditing and Operational Reporting.
* Work with end-users and change requests based on UI changes/updates to advanced analytics reports.

**Environment:** AWS, PostGres, SQL, Glue, S3, IAM, EC2, Databricks, Athena, Confluence, Kafka, Python, Scala

**CVS Health**

**| Woonsocket, RI**

Solution Architect - Data Engineering Aug 2019 – April 2023

Retail Pharmacy Artificial Intelligence (RPhAI) is a major enterprise level business optimization initiative by CVS Health to leverage Machine Learning and Artificial Intelligence Techniques to Resolve Pharmacy Claim Rejections by the PBM’s. RPHAI uses CDP, EDW, Rx Claims data, EPIC EHR data and Clinical data to support intelligent ML models for business optimization.

Responsibilities:

* Lead Big Data Solution Engineer on Retail Pharmacy Artificial Intelligence (RPHAI) Applications.
* Worked on Enterprise Data Architecture to consolidate, curate and build use cases from Enterprise data.
* Responsible for Architectural Review Board (ARB) approvals for data engineering platform to consolidate and host the data from payers, providers, members, patients, PBM’s and retail.
* Design and implement end-to-end solutions using Machine Learning (ML), Optimization, and other advanced computer science technologies, and lead production deployments.
* Worked on data consolidation from Adobe Customer Data Platform (CDP) and Enterprise Data (EDW)
* Responsible for migrating Data sources and Data processing applications from Aetna’s Hadoop 3.X to CVS Cloud - Azure Data Factory, Azure Synapse, Azure/Snowflake and Azure Databricks.
* Provide solution architecture and data lake architecture to migrate existing data applications and data assets with minimal code changes and resources from Azure Cloud to Google Cloud (GCP)
* Worked on Google Cloud Composer, Data Flow, Data Proc clusters, Apache Beam and Big Query.
* Work with data modeling, data architecture and governance concepts and aggregate large volumes of data from various sources to discover patterns and features necessary to build machine learning models.
* Provided data architecture for Customer Data Platform data and CVS Ad Network data for new use cases.
* Leverage Data Robot to train the Machine Learning models and work closely with Data Scientists for hyper parameter tuning, model identification, A/B testing etc.
* Design and build frameworks to orchestrate ML data pipelines including deployments and support.
* Hands-on development with “big data” platforms including Hadoop, Apache Spark apps and as well as traditional RDBMS data sources (e.g., Teradata, Oracle).
* Build continuous integration/continuous delivery, test-driven development, and production deployment frameworks with automation tools:  e.g., Jenkins, Git, Airflow, Control-M.
* Worked with streaming data using Apache Kafka and Apache Spark Streaming applications.
* Worked on Solution Architecture for AZURE to GCP platform migrations

**Environment:** Azure, GCP, Spark 2.3, Python, Scala, Hadoop, Hive, HDFS, Adobe, CDP, SaaS, EDW, Gitlab, Jenkins, DataRobot, Databricks, Synapse, ADF, Conda, Airflow

**Change Healthcare**

**| Chicago, IL**

Big Data Engineer/ AWS Solution Architect *Aug 2017* – Aug 2019

The IHDP puts CHC in a unique position to be an authoritative source for large datasets for the industry by incorporating financial, clinical, and operational data combined with Artificial Intelligence and Machine Learning. That in turn enables serving existing and emerging markets with new opportunities and innovative products.

Responsibilities:

* Responsible for translating business requirements into technical requirements and application code.
* Responsible for developing tactical and strategic solutions for the Intelligent health data platform.
* Responsible for documenting the technical design and solution architecture for enterprise data lake.
* Develop/enhance big data framework, tools, and governance processes for healthcare claims data.
* Responsible for data ingestion from various payers, providers, institutions, clearing houses and middle vendors on SLA basis for large scale processing and egress to support CHC client’s data science activities.
* Responsible for processing large volumes of EDI claims data and storing it in enterprise data lake.
* Worked on ingestion, transformation, and egress of data from Payers, providers, and clearing houses.
* Analyze complex data ingestion pipelines from sources like EDI, EPIC EHR and Rx data to transform raw data sources into easily accessible models hosted in centralized data warehouse.
* Built Spark / Scala data applications to process raw EDI transactions like Claims (837), Remittances (835) and Eligibility & benefits request/response (270/271) into tabular format (CSV) hosted in data lake.
* Extensive work in data mapping EDI 4010 and 5010 transactions to X12 and cXML/iXML formats.
* Created complicated XSLT scripts to convert X12 formatted data into XML data with tabular structure.
* Worked on Data Rights Management and RBAC models to cater the claims data to the data owners in the most efficient, robust, and secure manner.
* Create custom YAML scripts for launching EMR clusters, scheduling pipeline jobs and custom jobs.
* Create custom code using Apache Spark framework to process large datasets.
* Responsible for performance tuning and hyper-parameter tuning of the Spark based big data applications.
* Created AWS Glue based ETL jobs utilizing the Glue Crawlers, Glue Catalogs and AWS Athena.
* Develop training material for ongoing maintenance of applications built using Big Data solutions.
* Research/Development on Machine Learning, Artificial Intelligence and Blockchain technologies.
* Responsible for unit testing, performance testing and stress testing the big data pipelines.

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

**Cars.com**

**| Chicago, IL**

Machine Learning Consultant *Mar 2017* – Aug 2017

Project was to research, experiment and productionize Big Data and Machine Learning techniques for Predictive Analytics and Business Optimization.

Responsibilities:

* Lead role of Big Data Machine Learning Pod comprising of Data Analysts, Data Scientists and Engineers.
* Worked closely with Data Scientists on Machine Learning Model Engineering and Responsible for execution phase of the Big Data Machine Learning pipelines.
* Provided Solution Architecture to integrate Customer Data Platform (CDP) data into ML workflow and ML use cases to study and analyze customer behavior and customer insights.
* Productionize the Machine Learning Models to run on production HDFS cluster (CDH 5.5.1)
* Worked on model engineering of Regression and Classification algorithms using Spark and TensorFlow.
* Used Hue, Hive and Impala for Data Preparation and Data Validation.
* Worked on supervised learning using Regression for Predictions and Classification for Recommendations.
* Worked on Machine Learning ensemble models like Gradient Boosted Regression and Random Forest Regression for predictive analytics as part of Business Optimization.
* Used Spark Scala and Python APIs for production machine learning pipelines.
* Responsible for documenting Architecture and Data Flow across the systems.
* Worked with DevOps Team for Continuous Integration and Delivery (CI/CD) of the Machine Learning pipelines in non-pod and prod environments.

**Environment:** Cloudera, Kerberos, Spark 1.6/2.0, SaaS, Scala, Python, Kafka, Hive, Hue, HDFS, AWS, S3, Sqoop, Couchbase, Confluence, Adobe CDP, Bitbucket, Jenkins, Artifactory

**McDonald’s Corporation | Sapient Razorfish**

**| Chicago, IL**

AWS Big Data Architect *Nov 2016* – Mar 2017

Strategic Enablement (SE) and Acceleration for a premium client as part of the Enterprise Cloud Migration and Automation Program to move from Monolithic to Micro-service architecture. Key aspects of SE is to capture the Capability Maturity Model, thought leadership in enabling Big Data Architecture, Global Data Lake Setup, Streaming Data, IoT Data (Internet of Things), Predictive Reporting (Machine Learning) and Global Market Solution for Rapid Roll-Out to over 100 Countries.

Responsibilities:

* Strategic Enabler of Big Data Track in this client facing role for the Cloud Platform Migration Program.
* Responsible for capturing Capability Maturity Model for Core Customer Platform and Global Data Platform.
* Responsible for conducting Strategic Enablement workshops with the client side Chief Data Officer (CDO).
* Worked closely with Technology Senior Managers to create Proof of Concepts and Tool evaluation.
* Developed POC Spark Applications using Scala API and replicated them using Talend/Spark, S3 and EMR.
* Worked with DevOps (CI/CD) team for automation plan of Talend/Spark ETL pipelines on AWS EMR.
* Worked on Cloud infrastructure provisioning for Client Databases, ETL tools and other server instances.
* Waterline Data tool assessment for reading the Data in rest and identify Data Quality Scores and metrics.
* Technology tool assessment for Data Science using Paxata and Global Data Lake setup in AWS S3.
* Research and Analysis on Machine Learning algorithms as part of Data Science Enablement strategy.
* Design framework of Data Integration adaptors for pub/sub MQ in Micro-services/API Environment.
* Active role in Internal Capability Prioritization, Cross-functional team reviews and Client Review Meetings.
* Multiple Client presentations to evaluate Maturity Assessment and to estimate story sizing and costs.
* Collaborated with Agile PM’s for creating Pod structures, Agile Stories, Staffing Plan and Cost estimation.
* Research and Analysis of Collibra Data Governance tool and Collibra Connect to integrate Data in motion and apply Data Governance rules and monitoring procedures for Data Stewards.
* Created Agile stories for Data Ingestion (Batch, Stream and IoT), Global Data Lake, ETL Migration to Redshift using Talend/Spark on AWS EMR, Data Marts for 360 views, Prescriptive and Predictive Reporting, Data Science Activities.

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

**Northern Trust**

 **Corporation | Chicago, IL**

Sr. Big Data Consultant *April 2015* – Nov 2016

Derivatives Transformation Program: Northern Trust’s (NT) Derivatives Transformation Program to upgrades the enterprise data exchange format to ISO 20022 (Dodd Frank). Enterprise level upgrades to the data layer and technology layer by adopting and integrating big data technologies with the existing IT environment.

Responsibilities:

* Served as a Sr. Big Data Consultant under Northern Trust - Derivatives Transformation Program.
* Worked as Big Data Lead Consultant for Enterprise Information Management data pipelines.
* Identified appropriate Hadoop technology stack based on NT’s Business Requirements and Use Cases.
* Responsible for integration of Hadoop technologies with the Enterprise IT - Applications and Data Layer.
* Responsible for engineering data pipelines jobs from production cluster (CDH5) to Central Data Lake(S3)
* Worked on EDW and Data Mart architecture using Sqoop, Spark, Redshift, SQL Workbench, and Tableau.
* Solution Architecture for Enterprise Information Management (EIM) using Cloudera(HDFS), AWS & Spark.
* Created Enterprise Security Monitoring (ESM) data pipeline jobs with AWS EMR, Spark, Kafka and AWS S3.
* Responsible for successfully deploying Position Management in PROD using CDH5 Enterprise Edition.
* Developed multiple complex data pipelines for Derivatives Data Services and downstream applications.
* Developed Spark 2.0 applications in Scala using API’s like Spark SQL, Dataframe, Dataset and DStream.
* Developed Proof of Concept for Spark Applications in Scala and Python using MLlib (Machine Learning).
* Developed and executed Hive scripts for ad-hoc Data Analysis and Business Intelligence Reporting.
* Applied Big Data Technologies for Enterprise Information Management and Security Monitoring.

**Environment:** Windows, Linux, Oracle 11g, Mainframe, JDBC/ODBC, RESTful API, AWS, Hortonworks, Cloudera, Spark 1.6/2.0, SQL, NoSQL, HDFS, Talend, HBase, EMR, CSV, XML, JSON, Parquet, Scala IDE, SBT, Oozie.

**Key Bank**

 **| Cleveland, OH**

Senior Consultant – Big Data Architect *September 2014* – April 2015

Enterprise scale initiative by Key Bank to consolidate the Enterprise Data from various sources into the Shared Foundation Data (SFD) and adopt big data technologies for Business Reporting, Management Reporting, Executive Reporting and Predictive Analysis.

Responsibilities:

* Served as a Big Data Consultant on multiple concurrent projects under SFD Program.
* Member of Big Data Architecture Team for Key Bank Data Tier integration with Hadoop technologies.
* Responsible for identifying potential data sources for batch processing and real-time stream processing.
* Configured and Installed Cloudera Manager for cluster setup in Staging and Test Environments.
* Responsible to build data lake using Amazon Web Services (AWS) like EC2, EMR, S3 and Redshift.
* Responsible for launching and configuring production clusters in cloud environments using AWS EMR.
* Responsible for cluster setup with 5 nodes in dev and scale-out to more than 100 nodes in test and prod.
* Responsible for developing data pipeline spark applications in Python using Spark Context and Spark SQL.
* Performed ad hoc Data analytics for Management Reporting using Hive.
* Used Amazon EMR for spot term instances for ad-hoc data analysis on large datasets and S3 for storage.
* Used Sqoop for database imports, overnight jobs, Incremental jobs, file formatting and for Hive exports.
* Responsible for data validation using IBM MDM, Data Quality (DQE) and Test Data (TDE).
* Responsible for loading and dissecting Hive tables, Partitions, Buckets, Joins and Dynamic Partitions.

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

**Wells Fargo**

 **| Des Moines, IA**

Senior Consultant – Web Services Data Migration *April 2013* – September 2014

WFF Data Migration Program: The project is to migrate the Wells Fargo Financial Home Loan Data from Wells Fargo Supreme to Mortgage Servicing Big Data Platform.

Responsibilities:

* Member of Wells Fargo Home Mortgage Web Service Governance (MWS) Group.
* Served as a Senior Business Systems Analyst by supporting over 50 Mortgage Web Services consumed by various Application teams (Java, .NET and mobile apps) grouped into multiple Business Units.
* Created Technical documentation for Application Development teams participating in MWS middleware integrations and to get access to Mortgage Servicing Platform Data (MSP) using Web Services.
* Reviewed and Approved Business Requirements, System Architecture, Data Architecture and Technical Design documents for various Application Development and enhancement projects.
* Responsible for coordinating Data migration with application teams (TOG) in SOA environment.
* Responsible for data integration of Wells Fargo Financial (WFF) Data conversion.
* Responsible for MWS Authorization and Authentication process for Applications to access Mortgage Data.
* Worked on Oracle SQL developer to create, enhance and maintain web service database for Reporting.
* Created weekly and monthly reports on web service consumption by various Business Units.
* Conducted Monthly touch pints with Application Managers and Application Solution Engineers to review web service consumption and leverage existing services provided by Black Knight Financial Services (BKFS).
* Responsible to resolve complex type issues with multiple sources having million record data sets.
* Participated in various Technology Conferences, AWS Summit and Wells Fargo Town hall meetings which promote adoption of cloud and big data technologies.
* Research and analysis on Hadoop Cluster setup on commodity hardware and in the cloud like Amazon EMR and Azure HD Insights.

**Environment:** Mainframe, Windows, Unix, Java, Eclipse, AWS, SQL, Oracle 10g, Business Objects, Hadoop, HDFS, CSV, XML, HTML, SOA, Web services, SOAP UI

**Prudential Financial**

**| Shelton, CT**

Business Intelligence Consultant *March 2011* – April 2013

Pru Inside - New Policy Administration Platform

Pru Inside is implementation of a new Policy Administration Platform for the Pru Inside business using strategic methods and risk compliance rules.

Responsibilities:

* Served as a Business Intelligence Consultant and lead for New Policy Administration Platform Reporting.
* Captured Business Requirements (BRD) and translated them into Functional Specs (FSD)
* Worked with Data Analysts to capture Data Requirements for Business Intelligence Reports.
* Conducted User Requirements session to gather customization requirements for BI Reports.
* Responsible for driving logical data Warehouse and Business Intelligence system solutions.
* Coordinated with the multiple Lines of Business and the Technology groups for Solution identification.
* Created Solution Design for the Data flow from the Repository, source system, source file to Staging tables and pre-staging tables and then to Data Warehouse and from Data Warehouse to the Data Marts.
* Developed SQL procedures data loads from Legacy systems and batch extracts into the Operational Data Stores (ODS) and the Enterprise Data Warehouse (EDW).
* Developed SQL queries and Stored procedures for ETL jobs, Job scheduling, Job triggers and Testing.
* Responsible for setting of error logic for streamlining and automating the data loads for data cleansing
* Conducted preliminary analysis of Big Data Hadoop Transformation Projects and source systems from Data Governance perspective.
* Contributed in development of Business Reports using Informatica Hparsor Big Data in Amazon Cloud

**Environment:** Java, Hive, Hparsor, Big Data, Hadoop, MS Office, ETL, SQL, Oracle, Erwin, QC, BI, Informatica.

**Cognizant Technology Solutions**

 **| Minneapolis, MN**

Business Systems Analyst (BI) *October 2009* – March 2011

Equity Trading System conversion of the equity trading platform of Ameriprise Financial from Charles River Equity Trading System of River Source Investments to the Long View Trading System (LVTS) of Columbia Management Group as part of the acquisition.

Responsibilities:

* Lead BSA on reporting work stream for Report requirements, design, development and delivery.
* Consolidated list of reports from Charles River TS and Long View TS by enhancing existing reports, removing duplicates and also merging some reports for better reporting for Portfolio Managers.
* Created Functional System Design (FSD) and captured Data Requirements (DRD) for over 100 Reports.
* Lead multiple project teams of technical professionals through all phases of the SDLC.
* Implemented changes to the Physical Data Warehouse Model based on gap requirements.
* Created UML diagrams using MS Visio for Business Use Cases during the Design Phase of SDLC.
* Worked closely with PM, BA’s and the developers to ensure no gaps in communication.
* Worked on capturing transformation logic for Extraction, Transformation and Loading (ETL).
* Worked on MS SQL Server 2005 with business intelligence in SQL Server reporting services (SSRS)
* Responsible for coordinating between business teams at Client side and off-shore development teams.
* Responsible for time and cost estimation, Change Requests, defect tracking, monitoring and delivery.
* Prepare test plans, test cases, test scenarios and test strategy by collaborating with the Testing team.

**Environment:** MS Office Suite, Informatica Power Center, ETL, Business Intelligence, Microstrategy, RUP, UAT, SQL, Windows NT/XP, Teradata, Abinitio