



Chiranjeevi Kolli

An adept, technically sound Lead Big Data Engineer proficient in Architecture, Analyzing, Designing, Integrating, re-engineering and developing highly sophisticated systems including Analytics, Big data space, and Data warehousing. Possessing rich and diverse experience in various technologies an in-depth knowledge of disparate platforms, spanned over **8+** years of industry experience.

✉ cj.shans@gmail.com
☎ +60186698370

in [linkedin.com/in/chiranjeevi-kolli](https://www.linkedin.com/in/chiranjeevi-kolli)
S [kolli.chiranjeevi](https://github.com/kolli.chiranjeevi)

TECHNICAL EXPERTISE:

Big Data Technology:	Hadoop Spark Hive HBase Kafka Sqoop MapReduce HDFS Oozie NiFi
Languages/Scripting:	Python Scala MATLAB R Unix Shell Scripting
Database (RDBMS/NoSQL):	Oracle Teradata HBase
System Integration Tools:	Pentaho Data Integration Informatica Teradata GCFR
Visualization:	Tableau
Operating Systems:	Linux AIX Windows
Other tools:	SBT Git ELK Control-M

PROFESSIONAL EXPERIENCE:

Standard Chartered GBS- Client Due Diligence (CDD) Enablement
Lead Big Data Engineer

Feb 2017 - Present
Kuala Lumpur, Malaysia

Technologies: Hadoop, Hive, Spark SQL, Python[NumPy|Pandas|SciPy|Matplotlib], NiFi, Unix Shell Scripting

Description:

- Automating the risk rating generating system to prevent the AML, terrorism and other illegal activities by capturing the updates of customer static information.
- Strategic consumption approach to replacing Teradata GCFR (ELT) with Hadoop Filter Transform Load (HFTL) framework.
- Projected savings of \$1m per annum by replacing the HFTL framework.

Responsibilities:

- Define, design and develop the common Big Data pipeline framework and environment.
- Extraction, preparation and loading the data.
- Enrich and transform the data to perform data analytics.
- Performance analysis, tuning, and recommendation.
- Estimation of complexity, the volume of work and provide the cost estimation.
- Draw the deeper insight into the pipeline for critical discoveries to addressing the client queries.
- Analyzing data lineage of disparate source systems and provide the inputs on the requirements for proposed systems.
- Applying the statistical/mathematical approaches in the area of prediction and optimization.
- Visualizing the data to identifying the hidden data patterns and distribution.
- Capacity estimation for big data distributed applications.

Technologies: Hadoop, Spark, Hive, Sqoop, Kafka, HBase, Oozie, Spark SQL, Python, NiFi and Unix Shell Scripting

Description:

- Building Enterprise Data Management(EDM) Hadoop Data Lake to providing a centralized data platform with deep histories of raw data to enable unknown use cases.
- Data preparation for downstream usage and direct access to each dataset.
- Aimed to source the customer static and transaction data from various originating source systems.

Responsibilities:

- Design and develop the distributed big data applications using Hadoop and Spark.
- Work with business users to gather the requirements for delivering the technical solution documents.
- Establish connection data between Hadoop and RDBMS to servers for building ELT batch processing.
- Automate the Hadoop data processing jobs and reconciliation reports.
- Performance benchmarking and tuning of the modules to meet business SLAs.
- Guide and coordinate with release management for the continuous transition of various source/target systems.
- Take up the technology workshops and training sessions across teams.

Technologies: Hadoop, Hive, Sqoop, Oozie, Python, DB2, Oracle, DataStage, Unix Shell Scripting

Description:

- An Initiative is to migrate the existing EDW data into distributed and scalable environment for benefiting the storage features and reducing the maintenance cost.
- Decommission the traditional storage systems by HDFS.
- Functional reproduction of traditional COBOL/AS400 jobs into Hive ETL batch jobs.

Responsibilities:

- Design and develop the batch data processing jobs using Hadoop ecosystem stack.
- Collaborative development of ETL and DB2 data mapping documents along with data modellers and business analysts.
- Extract, prepare, integrate and load data into target systems.
- Preparing technical solution documents and code review.
- Performance tuning of Hive queries for low latency.

Technologies: Pentaho Data Integration (PDI), Informatica, Oracle and Unix Shell Scripting

Description:

- ETL Re-platform initiative is to replace Informatica by Pentaho Data Integration(PDI) as a strategic integration tool.
- Projected savings of \$3m per annum by replacing four production Informatica servers by PDI.
- Like to like functionality reproduction with equal or better performance.
- Possible reduction of jobs by reusable components, re-engineering.

Responsibilities:

- Design, develop and review the Pentaho jobs/sequencers.
- Understand the Informatica job functionality and develop same with Pentaho.
- Unit testing and test documentation.
- Build reusable components.
- Performance tuning of PDI jobs and SQL's.
- Provide training on Performance tuning of Pentaho jobs.

CutchSoft – Purchase and Manufacturing Module
Software Engineer

Jun 2010 - Aug 2011
Chennai, India

Technologies: C# and MS SQL Server

Description:

- CutchSoft supported and maintained the in-house applications to manage the Manufacturing, Sales and HR activities.
- To help the changing needs, new applications were built on both C# and ASP.net.

Responsibilities:

- Analyze, Design and build enhancements and bug fixes.
- Support the live manufacturing and purchase application.
- Build new reports and provide estimates for enhancements.
- Develop and unit test the code.

EDUCATIONAL QUALIFICATIONS:

Master of Computer Applications

2007-2010

- Jawaharlal Nehru Technological University – Anantapur (Andhra Pradesh, India)

ADDITIONAL DETAILS:**Honors & Awards:**

- SPOT Award at Standard Chartered Bank, Malaysia in 2016, Received from Global Head of Technology, Information Management.
- Best Employee at UST Global, India in 2013, Received from Country Head of UST Global, India.

Certifications:

- Python Programming– QT & T Consulting (Asia) Pte. Ltd.
- Apache Spark – QT & T Consulting (Asia) Pte. Ltd.
- MATLAB Onramp self-paced.
- Hadoop – QT & T Consulting (Asia) Pte. Ltd

Training/Workshops:

- Published Pentaho Performance Tuning document within UST and client NetApp.
- Undertaken new joined onboarding training in DataStage, Hadoop, Spark, Pentaho, DWH, and Unix.