PATHURI SAI HARISH

SOFTWARE ENGINEER • BIG DATA ENGINEER

TECHNO-FUNCTIONAL PROFESSIONAL

Learner & Receiver: Flair to continuously upgrade knowledge purview & keep updated on latest technology trends

PROFILE SUMMARY

- A collaborative engineering professional with substantial experience designing and executing solutions for complex business problems involving large scale data warehousing, real-time analytics & reporting solutions.
- Currently working with Samsung R&D Institute India, Bangalore as Software Development Engineer - Big Data Engineering; currently working in Digital Appliance - Internet of Things (IoT) domain.
- Known for using the right tools when and where they make sense and creating an intuitive architecture that helps organizations effectively analyze and process terabytes of structured and unstructured data.
- Conferred with SRIB – Spot Award, for being one of the first ones in the IoT Analytics team to learn and use Apache Kylin: Jun'20.
- Able to integrate state-of-the-art Big Data technologies into the overall architecture and coordinated with developers through the construction, testing and implementation phase.
- Experience architecting highly scalable, distributed systems using different open source tools as well as designing and optimizing large, multi-terabyte data warehouses.

CAREER CONTOUR

Samsung R&D Institute India, Bangalore | June 2019 - Present
Software Development Engineer - Big Data Engineering

- Designing, creating, testing and maintaining the complete data management & processing systems - establishing and supporting growing infrastructure to strategically analyze Big Data.
- Working closely with the stakeholders & solution architect.
- Ensuring architecture meets the business requirements; also building highly scalable, robust & fault-tolerant systems.
- Held accountable for identifying and defining key metrics, creates insights and dashboards on Tableau/ Apache Superset to drive business decisions.
- Creating innovative solutions to Big Data issues and challenges within the team.
- Building & orchestrating ETL pipelines based on Hadoop Stack involving Spark, Hive, Hbase, Kylin & Airflow; anchoring software development process in Python, Scala, Java, and Spark.
- Research skills --- emphasizing on improving user experience by providing convenience products as per their requirements.
- Earned with an authority to build, test and deploy the services into PROD; also followed Agile Methodologies in the process.
- Taking care of the complete ETL process; discovering data acquisitions opportunities.
- Improving data quality, reliability & efficiency of the individual components & the complete system.
- Introducing new data management tools & technologies into the existing system to make it more efficient.

Samsung R&D Institute India, Bangalore | Jan 2019 - Jun 2019
Student Trainee - Smartliving

- Efficiently & effectively worked on:
  - Proof of Concept that involved collection, analysis and interpretation of a large dataset.
  - GraphDB for efficiently storing and retrieving the data.
  - User profiling and recommendation service built on Spark.
- Skilfully used shell scripting to automate the processes; handled Linux Administration Work of project servers.

Mobile:
+91 9591969138

E-Mail:
saiharish97@gmail.com

SIGNATURE SKILLS

Technical

- **DevOps:**
  AWS, Docker

- **WebDev:**
  .NET, HTML 5

- **Programming:**
  C, C++, Python, JAVA, Scala, SQL, Shell

- **Apache Tools:**
  Spark, Hadoop, Kylin, Hive, Zeppelin

- **Visualization Tools:**
  Tableau, Apache Superset

- **Version Control:**
  GIT

Tech Mahindra Ltd., Hyderabad | May 2018 - Jun 2018
Summer Intern – Application Development

- Provided assistance to Solutions Team in building an Android Application for one of its clients; got first exposure to Agile methodologies & systems.

PUBLICATIONS

Device Interaction Graph: Directed Decision Graph for Settings Auto-Completion

- As Co-Author: Jan 1, 2021
- Published in: IEEE Access (Volume:9)
- Proposed a Device Interaction Graph Solution that automatically recommended and auto-completed the user's preferred settings in real-time. This settings auto-completion solution improved the user experience by reducing the number of manual or verbal interactions required to enable the operation of the device. We have applied the auto-completion solution on the data obtained from 158,213 connected devices and achieved recommendation accuracy of 89.79% with an average of 3.132 interactions.
Link to the Publication

Artificial Intelligence Based Air Conditioner Energy Saving Using a Novel Preference Map (IEEE Access)

- As Co-Author: Nov 2020
- Air Conditioning (AC) systems have contributed to a high percentage of the residential building energy consumption. Most of the recently released AC models are Internet of Things enabled. The data generated from all these ACs can be analyzed to understand the usage pattern and energy saving. In this paper, we have proposed a Cloud based Artificial Intelligence (AI) solution that uses the data from 37,748 ACs to analyze and generate a novel 2-D Preference Map. We have used the Preference Map in our AI solution to automatically generate Top-K energy saving recommendations.
- Link to the Publication.

Running Kylin on Docker like a Pro! (Medium)

- As Author: Aug2020 - Present
- Contribution back to the community that assisted people planning to work on Apache Kylin.
- It was accepted into Analytics Vidhya's publications on the site.
- Link to the Publication.
- Published in: Medium

ONLINE CERTIFICATIONS

Apache Kylin • Implementing OLAP on the Hadoop Platform • Udemy (Apr 2020)
- Learned how Apache Kylin is able to provide constant query speed regardless of the ever-growing data volume.
- Credential ID: UC-a3f94409-5847-4d6c-a2ba-77b9ee19d1d6.
- Link to Certificate.

Machine Learning • Stanford Online • Coursera (Jan 2019)
- Learning of Machine Learning Algorithms & Concepts that are used globally in software industry.
- Credential ID: G6ZSETYYR4GD.
- Link to Certificate.

Data Structures • Coursera (Aug 2018)
- Credential ID: LSCLZUA65BJ9.
- Link to Certificate.

Algorithmic Toolbox • Coursera (Jul 2018)
- Credential ID: 4GFV7LE36L47.
- Link to Certificate.

REFERENCES AVAILABLE ON REQUEST

Functional

- **Software Development:**

Requirement Gathering

Quality Assurance

Testing

Post Implementation Support

Agile

- **Big Data Engineering:**

Data Architect

Data Collection & Transformation

Data Extraction

ETL Jobs

Installation & Configuration

SCHOLASTICS PORTFOLIO

2019

Bachelors in Technology - Computer Science & Engineering

MIT - Manipal Institute of Technology, 8.69 CGPA

2015

Higher Secondary Certificate – 12th

FIITJEE - Dilsukhnagar (Excel Junior College, Hyderabad), 94.1%

2013

Senior Secondary Certificate – 10th

Sri Chaitanya Techno School, Hyderabad with 9.7 GPA

ACADEMIC PROJECTS

- Library Management System using JAVA
- Alumni Portal using .NET