

Hesham Rafi

Data Engineer - S.No

Singapore

- Email me on Indeed: <http://www.indeed.com/r/Hesham-Rafi/c1055a0b519cc6e2>

IT professional with 11 years of comprehensive experience in the areas of:

Big Data Technologies Architecture Design Data Warehouse Architecture
ETL Architecture Machine Learning Performance Tuning

- Currently working as Data Engineer with MunichRe Management PTE. LTD Singapore.
- Experience in Architecting DW/ETL solutions end to end, especially in ETL architecture and design which includes, Development architecture, ETL Framework, Batch and schedule design, designing restart and recoverability, Common components/patterns design and performance tuning.
- Strong experience in Azure cloud for deploying Machine Learning models, Web servicing, Flash Dashboard and R Shiny dashboards.
- Strong Technical Knowledge on - Big Data Technologies, R, C/C++, Java/Scala, Spark, Unix and Python shell scripting, Teradata, Oracle, Pro*C, XML
- Experience is developing Machine Learning model, writing Algorithms and deploy Machine Learning models for customer.
- Collaborated with different client at different client locations globally, worked with Amdocs USA, Maybank in Malaysia, AXA in Malaysia, DBS in Singapore, SCB in Singapore. Possess Strong client interfacing and relationship building skills.

Work Experience

Data Engineer

S.No

November 2018 to Present

Organization Name From (MM/YYYY) To (MM/YYYY) Location Last Designation Hell
1 MunichRe 11/2018 Till Now Singapore Data Engineer

Regional Data Engineer

MUNICH RE - Singapore

November 2018 to Present

Project: Data Lake and Machine Learning Use Case

- Designing and developing data pipeline for loading data in Azure cloud and on-premise infra.
- Deploying Machine learning model and Dashboards in Azure cloud.
- Loading source data files into an Azure data lake and DSVM and make available for data scientist.
- Deploying developed R Machine learning model for customer using Shiny.
- Developing dashboards in Flask Python for Machine learning models.
- Contribution on developing Data profiling dashboard using Power BI.
- Contribution on Machine learning model development use case using R and Python.
- Monitoring, and administering Hadoop clusters and Jobs.

Big Data Engineer

Encora Technologie

May 2018 to October 2018

Big Data Engineer

AXA AFFIN, KL - MY

May 2017 to October 2018

Project: Data Lake Development

- Design the Data Lake architect for storing multiple source system data into a Hadoop Data Lake.
- Design and Implemented Ingestion framework DIF for AXA MY using Spark and Scala.
- Using DIF for loading historical data and delta file, which is coming from multiple source systems.
- Interaction with MDM development team and Prepared MDM SIT and UAT test cases for user testing.
- Design and developed the data pipeline and deployed the jobs for loading MDM system data into a Data Lake.
- Installing and configured the Pandora Data Quality tools into our Pandora server.
- Defined the data validation rules and implementing on Pandora Data Quality tool.
- Contribution on conversion of Osprey SAS script into a PySpark and scheduled for daily run for reporting.
- Automated the Data extraction form AS400 system for Data Scientist for developing Machine Learning model.
- Automated Telemetric flume data, which is coming into a json format, parsing and loading into a data lake. And created scheduled job for running into a Data lake.

Big Data Engineer

Verinon Sdn Bhd

May 2017 to May 2018

Hadoop Expert

STANDARD CHARTERED BANK - Singapore

August 2016 to May 2017

Project: Tax Data Mart

- Developing and Automate the source files validation using shell script.
- Developed automate script for error table validation from Hive tables.
- Loaded xml files into a Hive table and parsed the xml file using Hive UDF functions.
- Developed Scala code for connecting Hive environment and running the Hive mapping queries.
- Reading binary files (xlsx) into a Unix environment and Validating and loading into a Hive tables.
- Configuring Control-M jobs for running the daily batch.

Sr. Technical Lead

HCL Axon

July 2011 to May 2017

Hadoop Expert

DBS BANK - Singapore

January 2015 to August 2016

Project: Data Lake and Data Ingestion

- Using Sqoop for Import/Export Unicode Data from and to HDFS and Teradata.
- Loading Hive tables to Teradata layer to perform data check and reconciliation.

- Created necessary commands for table schema creation, exiting and creating scripts / jobs and will be triggered at the Edge Node and Hadoop layer.
- Configuring the Flume Agent for reading data from channel and written Interceptor code for flume agent in Java.
- Developed Hive UDF function using XPath for parsing XML file for loading data into a Hive tables.

Senior Technical Lead

MAYBANK, KL - MY

January 2013 to December 2014

Project: REDW Data Warehouse

- Architect design for running the Informatica batches.
- Developed the automation script for developing the Informatica workflow for loading the files into Teradata tables.
- Developing Shell script for validating the source files and running the Informatica batches.
- Setting Informatica, Teradata environment for executing shell script.
- Configuring the Zena Scheduler for running the Daily and Monthly batches.
- Configuring Hadoop, Hive and flume Environment in RHEL 6.
- Monitoring, and administering Hadoop clusters and Jobs.
- Converted HDFS unstructured data into Structured and loading into a Hive tables.

Lead Engineer

WOLTERS KLUWER - Hyderabad, Telangana

July 2011 to December 2012

Project: JD Edward upgrade

- Contribution project Architect changes for UNIX to Windows Server Environment.
- Setting Windows Server environment for executing shell script using Cygwin.
- Converting UNIX and Perl script for windows compatible and running into a Windows server.
- Contribution in Debugging and Unit test case preparation for all Daily, Monthly and Ad-hoc batch runs.
- Configured job for running the Daily and Monthly batches using Maestro scheduler.
- Developed windows BIP service, which takes invoice files for NAS location and send a mail to a customer email id using C# technologies for development.

Sr. Software Engineer

Mahindra Satyam

May 2010 to July 2011

Software Engineer

Mahindra Satyam - Hyderabad, Telangana

May 2010 to July 2011

Project: AMDOC Smart Billing System

- Creating Tuxedo service using Tuxedo in IBM AIX system.
- Developing business logic for price structure using C++ environment.
- Contribution is writing SIT and UAT test case for user testing.
- Coordinating with other team members.

Software Engineer

IGSA Labs

December 2007 to May 2010

Software Engineer

IGSA SOLUTION LABS - Hyderabad, Telangana

December 2007 to May 2010

Project: RITES - Route Manipulation Planner

- Implementation of K-Shortest Path Algorithms in C/C++ using OOP concept.
- Applying Constraints for switching of network and blocking of routes with K-Shortest Path Algorithm development.
- Research and customization of K-Shortest path With Prof. Rangrajan and his team (IIT Bombay, India)
- Finding shortest path between two O-D pair with single commodity and multiple commodity
- Interface designing and development.
- Development of various kind of report using Crystal report engine API.
- Making license system for application.
- Developed Windows COM component and using in application.

Education

Diploma in Computer Application

INDIRA GANDHI NATIONAL OPEN UNIVERSITY - Delhi, Delhi

June 2008

Bachelor of Science in Computer

PURVANCHAL UNIVERSITY - Jaunpur, Uttar Pradesh

July 2003

Master of Computer Application in Computer Application

INDIRA GANDHI NATIONAL OPEN UNIVERSITY - Delhi, Delhi

Society of Operation Management

IIT - Chennai, Tamil Nadu

Skills

Hadoop (6 years), HADOOP (6 years), Hive (4 years), TERADATA (3 years), C++ (3 years)

Additional Information

TECHNICAL SKILLS

- Big Data Ecosystems: Hadoop, Hive, Spark, PySpark, Scala, Pig, Sqoop, HBase, Flume
- Programming Languages: R, C/C++, Java, C#, VBA, Assembly Language (8085/8086)
- Scripting Languages: Shell Script, Python, Perl, JavaScript, XML, HTML
- Databases: Oracle, Teradata, SQL Server
- Web Technologies: R Shiny, Node.js, React.js, Bootstrap, CSS, Java Script, HTML
- Reporting Tools: Power BI, Crystal Report, R Shiny, Python Flask, Plotly Dash
- ETL Tools: Informatica 9.5

- Platforms: Windows, Linux, AIX
- Cloud Platforms: Azure
- Methodologies: UML, Design Patterns
- Enterprise Scheduler: Oozie, BMC Control-M, ASG Zena, IBM Maestro
- Application Server: Windows Server, Tomcat 5.5

(Hesham Rafi)