BERNARD JOSEPH KANISH

DATA ENGINEER - BIG DATA

PROFESSIONAL SUMMARY

- Around 3 years of experience with 2 years in handling Big Data/Open-source projects with Spark, Kafka, Hadoop, HDFS, Yarn, SQOOP and Hive.
- Having sound knowledge in Batch & In-memory streaming, migration, consolidation, lift and shift of on-premises & cloud environment.
- Rich experience using RDBMS, ETL pipelines, SQL, and data wrangling.
- Have worked in different datasets, read data different source files like CSV, Parquet, Avro, JSON, Sequence files and store it in Hive tables using Spark and move it to the Hive Warehouse, HBASE and SSAS CUBES.
- Designed end-to-end data architectural design like ETL, ELT, Data lake, Delta Lake, Data modeling, Data Analytics, Data discovery for data science team
- Worked on slowly changing dimensions SCD type 1, 2 & 4
- Handled files such as Avro, Parquet, ORC, JSON and CSV.
- Extracted data from Data Lake and stored for Audit purpose
- Working in Agile methodology for past 2 years.

PROJECT

June 2021 - Present

DATA ENGINEER- BIG DATA - INATECH - A GLENCORE COMPANY

- Worked on data warehouse using hive via batch processing for data analytics.
- Worked on spark streaming pipeline for real-time data integration.
- EOD snapshot data captured and stored for analyzing daily PnL of trades.
- Created visualization of Data model using Power BI for decision supporting systems.
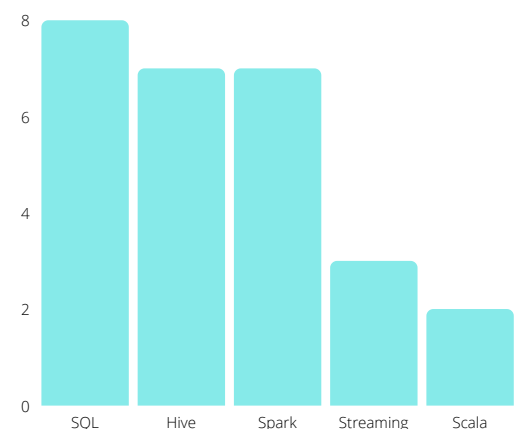
CONTACT

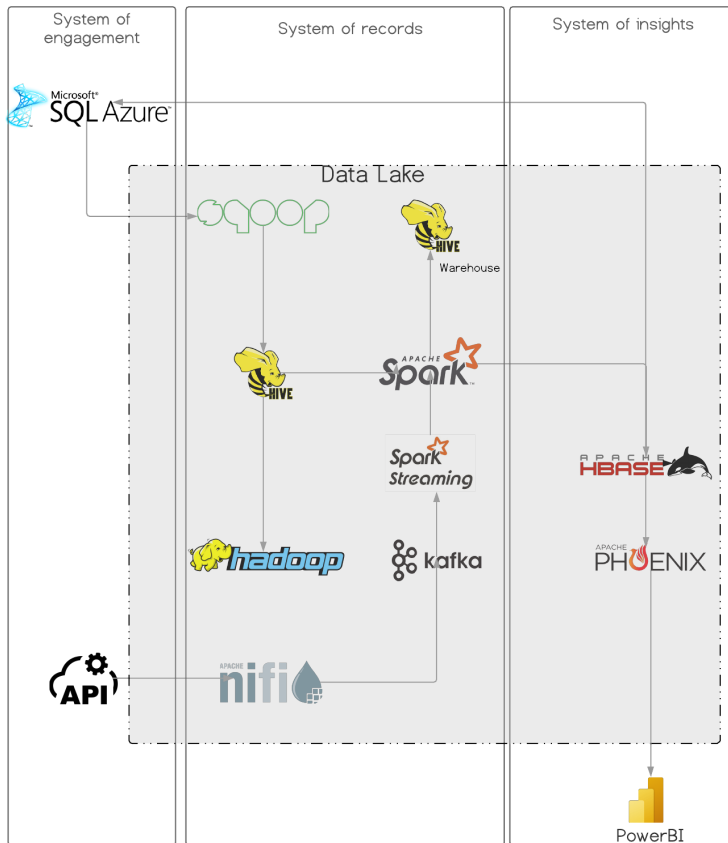☎ +91- 8870722066

✉ joseph.kanish.bernard@gmail.com

in https://www.linkedin.com/in/bernard-joseph-kanish-baa587160

SKILLSET

ETL/Batch	Hive, Spark
Data Ingestion	SQOOP, Nifi
Real time	Spark streaming
Storage	HDFS, AWS-S3
RDBMS	MS-SQL, MySQL
NOSQL	HBase
Scripting	Shell
Programming	Scala
Descriptive Language	MS-SQL, MySQL, DAX
OS/Platform	Linux, Windows
Cloud	AWS S3
Visualization	Kibana, PowerBI
IDE	Eclipse, Jupyter
Build Tools	Maven
Repository	Git, SVN, VSS
Agile	JIRA
Distribution	Cloudera 5.13.0

DEV SKILLS

Skill	Value
SQL	8
Hive	7
Spark	7
Streaming	3
Scala	2

System of engagement | System of records | System of insights

Data Lake

- Microsoft SQL Azure
- SQOOP
- HIVE (Warehouse)
- HIVE
- Apache Spark
- Spark Streaming
- Apache HBASE
- hadoop
- kafka
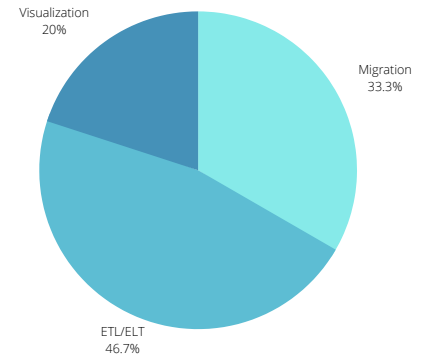- APACHE PHOENIX
- API
- Apache nifi
- PowerBI

- Data ingestion from Microsoft Azure Cloud DB to Hadoop File system using SQOOP.
- Collecting data from HDFS, processing in Hive for Data cleansing and curation to ensure data quality.
- Taking curated data and loading in Spark for further processing like Data wrangling, enrichment and data standardization.
- Loading processed data to HBASE/NOSQL layer on top of Apache Phoenix for interactive data analysis.
- Creating data model in Phoenix for data analytics team. Used as a SQL layer to ensure low latency.
- Maintaining snapshots in Hive as data warehousing layer for batch data processing.
- Furthermore, Analytics team build descriptive analytic dashboards.
- Exporting data from Phoenix to MS-Azure SQL server to juxtapose with OLTP data.
- Using PowerBI as visualization tool to create interactive dashboards and reports.
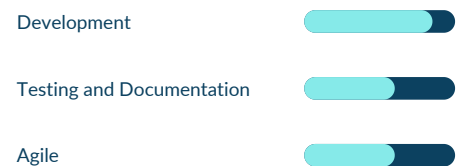
May 2019 - June 2021

SENIOR ASSOCIATE ENGINEER- INATECH - A GLENCORE COMPANY

- Development and testing for complex ETL defects
- Worked on Ad-hoc business requests such as end to end batch execution from front end database to target application
- Worked on Spark DSL and SQL Queries
- Reading CSV files and converting into Parquet followed by processing using Spark

---

Visualization 20%
Migration 33.3%
ETL/ELT 46.7%

Development

Testing and Documentation

Agile

2021
HACKERRANKSQL INTERMEDIATE PROGRAMMING - MASTERS

2020
SOLOLEARN- HTML5 - APPLICATION DEVELOPMENT FUNDAMENTALS
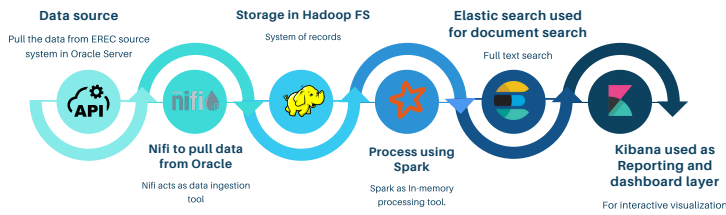
SOLOLEARN- MYSQL CERTIFICATION

EARNED BADGE- HACKKERANK GLOBAL COMPETITION

RANK 1 IN HACKATHON- INCEPTEZ TECHNOLOGIES

SPOTLIGHT EMPLOYEE- INATECH- 2019

BACHELOR OF ENGINEERING (CSE)
SRI VENKATESWARA COLLEGE OF ENGINEERING - ANNA UNIVERSITY
7.2 CGPA

Data source

Pull the data from EREC source system in Oracle Server

Storage in Hadoop FS

System of records

Elastic search used for document search

Full text search

Nifi to pull data from Oracle

Nifi acts as data ingestion tool

Process using Spark

Spark as In-memory processing tool.

Kibana used as Reporting and dashboard layer

For interactive visualization

ENGLISH- FULL PROFICIENCY

TAMIL - NATIVE

June 2018 - May 2019

ASSOCIATE ENGINEER- INATECH - A GLENCORE COMPANY

- Optimizing SQL queries for better performance. A full-fledged BI developer currently working exclusively on Microsoft products which includes SQL 2019, SSIS, SSAS, SSRS and Data Analysis Exchange also known as DAX.
- Develop data models using SQL and Hadoop (Big Data) to process data and deploy cube using Analysis services (SSAS).
- Write optimized DAX queries and generate reports which are product and customer-specific.

CHATBOT REALTIME STREAMING APPLICATION
01/2021

- Chatbots allow businesses to connect with customers in a personal way without the expense of human representatives. For example, many of the questions or issues customers have are common and easily answered.

CUSTOMER CHURN PREDICTION AND CALL VOLUME REDUCTION
12/2020

- A popular telecommunications provider analyzed call logs (CDR) and multiple complex data from multiple sources.
- The aim is to analyze the data to predict the likelihood of customer who could move on based on call type, duration, customer response in digital media, record captured by the activities performed by the call center agent while handling customers
- Participated in hackathon and won the prizes for designed using big data system to achieve the expected results.

I hereby declare that the information furnished above is true to the best of my knowledge

Bernard Joseph Kanish