**Sample Resume**

**PROFESSIONAL SUMMARY:**

Over 4+ years of experience encompassing wide range of skill set in **Big Data, Java, SQL** and **Python** technologies.

* Excellent hands on experience in developing and implementing Big Data solutions and data mining applications on Hadoop using, **HBase, Pig, Hive and HDFS, MapReduce**, **YARN, Sqoop**, **Flume, Kafka, Spark, Oozie, ZooKeeper.**
* Experienced in creating tables using **Hive** and **pig Latin**.
* Extensive Knowledge on developing Spark jobs by using **Pyspark, Spark SQL** and **Scala** accordingly.
* Good Knowledge on developing **Pyspark** jobs using **Spark SQL** and **Dataframes** on **AWS Glue** and schedule them using **step functions**.
* Experience in streaming **Cloud Watch Logs** to **kinesis** and notify through **SNS Topic.**
* Strong knowledge on writing **shell scripts** and scheduling Cron jobs for Postgres Functions.
* Work experience with cloud infrastructure like **Amazon Web Services (AWS).**
* Strong understanding of **NoSQL** databases like **HBase** and **Casandra**
* Experience working with cluster monitoring tools like **Cloudera** manager.
* Extensive experience in data ingestion technologies, such as **Sqoop**.
* Good knowledge on **Kafka** for streaming real-time feeds from external rest applications to Kafka topics.
* Good knowledge on **AWS** infrastructure services **Amazon Simple Storage Service** (**Amazon** **S3**) and **Amazon Elastic Compute Cloud** (Amazon EC2), **AWS Glue, Cloud Watch Rules, Step Functions, Lamda and CI/CD pipeline**.
* Good Knowledge in **Core Java,** **Python**, **Object Oriented Design** (**OOD**), **SQL** databases.
* Developed various **SQL** stored procedures, triggers and views to facilitate consistent data entry into the database.
* Experienced in all facets of Software Development Life Cycle using **waterfall** and **agile methodologies**.
* Worked with various file formats like **XML, JSON, CSV, Avro, Parquet, ORC, Sequence** and other.
* Mastering new technologies with a keen awareness of new industry developments and the evolution of next generation programming solutions.

**TECHNICAL SKILLS**:

|  |  |
| --- | --- |
| Big Data Tools | HBase, HDFS, Hive, MapReduce, Pig, Spark, yarn, Zookeeper, Sqoop, Oozie and Kafka |
| Programing Languages | Java, Python, C#, SQL, R, Scala, X86, Unix bash scripting. |
| Python Libraries | NumPy, SciPy, Scikit-learn, Pandas, Matplotlib, Pytables, Paramiko |
| Database | MySQL, SQL Server, HBase, Casandra. |
| Web Technologies | HTML, CSS, XML, JavaScript, JSON |
| Cloud Infrastructure | Amazon Web Services, Hadoop, Cloudera (CDH 5.6 & CDH 5.11) |
| Tools | Git, Unity 3D, Visual Studio, NetBeans, Eclipse, Lc3, Rational rose, Tableau. |
| Methodologies | Agile(scrum), Waterfall model |
| Operating System | Linux, Unix, Windows and MacOS |
| AWS TECHNOLOGIES | EC2, S3, AWS GLUE, STEP FUNCTIONS, CLOUD WATCH, KINESIS, CI/CD(CODE BUILD, Cloud Formation, Git Hub, Code Pipeline) |

Certification

* Microsoft Azure Fundamentals by MICROSOFT.

[https://www.youracclaim.com/badges/a4928a83-c376-4ac7-bec4- 177be857dc94/public\_url](https://www.youracclaim.com/badges/a4928a83-c376-4ac7-bec4-%20177be857dc94/public_url)

* SNOW PRO CORE CERTIFICATION by SNOWFLAKE University.

<https://www.youracclaim.com/badges/cd5fcaf1-087d-4c17-afd4-a73551de248a/public_url>

Professional Experience:

Company Name Jun 2019 -Present

Cleint: DANNON August 2020 – PRESENT

LEAD DATA QUALITY ANALYST

Location: Denver

* Work as technical lead, where gathering requirements, coordinating and helping in development for both onshore and offshore team members to deliver product to client on time.
* Writes scripts for data quality requirements using **python** and **postgreSQL.**
* Reviews file output and gathers input from Client for revisions
* Preparing business requirement document which show case scope and all steps to follow and develop product based on client requirements.
* Giving power point presentation in standup calls to show the progress/approach in the development and taking inputs from client accordingly for future development.
* Installs and sets up environment and scripts in assigned environment
* Developing **shell scripts** to automate the process of running python scripts.

**Data ENGINEER**

**Client: GE AVIATION JUNE 2019 – June 2020**

**Location: Cincinnati, Ohio**

* Involve in different phases of **Software Development life Cycle** to Analyze, Design, Coding and implement high quality scalable solutions as per the business requirements.
* Optimizing existing data architecture and queries to scale across terabytes of data for visualization and application consumption
* Work collaboratively with technology leads, architects and business partners to define objectives and analytics processing requirements to process large datasets.
* Identify Enterprise **Data Lake** sources that are needed for analytics modeling
* Writing complex **SQL** code to transform, pivot and stitch big data sets like **Oracle, GreenPlum, AWS Aurora, PostgreSQL**
* Migrated big data infrastructure from on-premise **GreenPlum** to the Cloud based architectures like **AWS Aurora, S3, AWS Glue, AWS EMR.**
* Writing **python shell/PySpark, Postgres SQL** code for ETL jobs on daily basis.
* Develop **PYSPARK** scripts using **Data Frames** to automate validation, logging, and alerting for Spark applications running on **AWS GLUE**.
* Partner with technology teams to ensure optimal use of the AWS Cloud Platform Services.
* Developed **Lamda function** to send logs to **kinesis streams.**
* Writing **Bash scripts** for automation processes
* Work with **GitHub Enterprise** for version control and Documenting data flows.
* Optimize and Monitor performance of **Spark applications** running in pre-production and take corrective action and improve in-case of failures before pushing to production.
* Involve in AdHoc stand up and architecture meetings to set up daily priorities and track the  
  status of work as part of highly **Agile work environment**.

**ENVIRONMENTS: AWS GLUE, S3, STEPMACHINES, SPARK DATAFRAMES, PYSPARK, PYTHON, SHELL SCRIPTS, GITHUB, POSTGRES SQL.**

**GRADUATE TEACHING ASSISTANT**

**Wright State University**  **August 2017- December 2018**

**Location: Dayton, Ohio**

* Implemented **Multi node** Hadoop cluster on Cloudera virtual machine.
* Developed Scala scripts using **Data frames, Datasets and RDD** in Spark for data aggregation.
* Worked on processing **batch** and **real time** data using Spark using **Scala**.
* Responsible for creating and writing to hive tables using **HiveQL** also performed partitions, bucketing and joining tables.
* Worked with **Kafka producer and consumer** to stream real time data from websites to spark.
* Worked with **Spark Streaming** Technologies to pull real time data from Kafka to data lakes.
* Developed Spark scripts by using Scala as per the requirement.
* Used **Spark RDD** and **Data Frames** for faster processing of data.
* Experience on **Kafka** and Spark integration for real time data processing.
* Hands-on experience for setting up Kafka mirror maker for data replication across the clusters.
* Loaded data and performed operations using **Spark SQL** and send the results to **TABLEAU** dashboards.

**Teaching Assistant Responsibilities:**

* Assisted professor in teaching duties of **Data Analysis course**.
* Developed simple **MapReduce** programs in java and python to understand its functionality.
* Organized and Developed lab assignments and projects for the course. Handled lab sessions and helped professor in grading.
* Good knowledge in loading data from **UNIX** file system to **HDFS** and performed queries on both HDFS data and local files using **Hive** and **PIG**.
* Used **Pig** as **ETL tool** and preprocessed the data and store it back into HDFS.
* Monitor students to load data from various file system using Unix command line utilities into **HDFS**.
* Load and transform large data sets of structure and semi structured using **Sqoop.**
* Instructed 60 under graduates in their lab sessions.
* Helped students to get good knowledge on subject by clarifying their doubts during lab and help sessions.

**ENVIRONMENTS:** HADOOP, HDFS, SPARK SQL, SQOOP, TABLEAU, MAPREDUCE, Kafka, UNIX, HIVE, PIG, RDBMS, Spark Streaming, ETL.

**Integrating Multimodal Sensory Devices (Thesis Work):**

**Wright State University August 2017-December 2018**

**Location: Dayton, Ohio**

* Augmented **HoloLens** with fully-fledged hand gesture and motion tracking capabilities by interconnecting it with a **Leap Motion Controller.**
* Used **Unity 3D** game engine platform developed server, client applications and established connection using **HoloLens sharing service and holographic remoting player.**
* Data coming from Leap Motion Controller read in **JSON** format on server side and transferred on client side using **C# Scripting** and **JSON parsers**.
* Deployed client application into HoloLens device using **Visual Studio 2017**.
* Developed a Use Case to illustrate this AR model usage and value in real time environment.

**Bigdata/Hadoop Developer**

**Atom Software Solutions, Vishakhapatnam. April 2016 -December 2016**

* Work with the **Teradata analysis** team to gather the business requirements.
* Responsible for building a distributed structured solution using **Big Data/Hadoop** eco system.
* Ingested data to **HDFS** from **RDBMS** using **Sqoop** and performed operations using **Hive or pig.**
* Created hive tables and worked on them using **HiveQL**.
* Developed **MapReduce jobs** using **java** also has good experience in working with **Cloudera Manager (CDH 5.6).**
* Performed performance tuning and troubleshooting of jobs by analyzing and reviewing Hadoop log files
* Generated reports of deployed data from various sources to HDFS using **Tableau.**
* Worked with **No SQL** databases like **HBASE** also responsible for administrating its updates and patches.
* Experience with **Oozie** Workflow Engine in running workflow jobs with actions that run Hadoop Map/Reduce jobs.
* Executed **Oozie** workflow engine to run multiple Hive and pig jobs which run independently with time and data availability.
* Involved in **scrum meetings** and **sprint planning**.
* Worked on issues reported by the **QA team** actively and resolved all the issues with data processing.
* Documented the systems processes and procedures for future references.

**ENVIRONMENTS:** Big Data/Hadoop, HDFS, RDBMS, Sqoop, PIG Latin, Map Reduce, Hbase, Apache Spark, Pyspark, Spark Sql, Scala, Scrum, ETL, Oozie.

**Java & SQL Developer**

**HCL CDC, Hyderabad. May 2014-June 2015**

* Developed a system to track and resolve customer queries which are submitted to organization portal, using **JAVA framework** and **SQL Server** tables.
* Implemented entire system as distributed architecture using centralized data base by establishing connection using **SQL Server**.
* Followed **3NF** normalization to avoid anomalies generated by the users.
* Generated **Data Flow Diagrams (DFD**) and **Unified Modeling Language (UML)** Diagrams which explains system architecture to client.
* Designed user interface using JAVA frame work, connected to database system through **SQL Server**.
* Developed two modules for **client t**o post their queries and for **employee** to post their answers.
* Created Database Objects such as Tables, Indexes, Views, User Defined Functions and Cursors.
* Exposed to all phases of the Software Development Life Cycle from analysis, design, development, testing, implementation and maintenance.

**ENVIRONMENTS:** JAVA 7, SQL, 3NF, Sql Server Tables, Sql Server, Data Flow Diagram (DFD), UML

**Other Projects: January 2018 -April 2018**

* Implemented **Simple Search Engine** on given data using basic term generation, indexing the preprocessed data and implementing query processor using Boolean and vector models.
* Performed **Text Mining** using Machine Learning algorithms like **Feature Extraction**, **Feature Selection** (chi-squared and

mutual information method on dataset), **Classifier Evaluation** (Multinomial Naïve Bayes, Bernoulli Naïve Bayes, kNN, SVM) by cross validating dataset and **Clustering** (k-Means and hierarchical) from feature set

**EDUCATION:**

**Master of Science in Computer Science**, **January 2017 – December 2018**

**Wright State University**, Dayton, OH GPA: **3.66/4**

**Bachelor’s: Information Technology** **July 2012- April 2016**

**GITAM University,** Vishakhapatnam, India GPA:**8.52/10**