samir benzada  
17862962400 samirbenzada@gmail.com

EXPERIENCE

|  |  |
| --- | --- |
| **Google** *Data Engineer* | nan October 2018 - Present |

* Participate in a proof of concept to prove out the effectiveness of using Kafka and DataBricks as a viable data warehouse.
* Experience with distributed stream processing technologies such as Flink, Spark Streaming and/or Kafka Streams
* Spark innovation and experimentation to create new models, techniques, and strategies
* Scripting experience with JavaScript, pySpark, Python, TSQL or other similar languages
* Hands-on day-to-day expert experience in administering a Cloudera cluster with Cloudera Manager, Cloudera Director, Cloudera Navigator
* Experience standing up and using Big Data Hadoop platforms such as Cloudera / HortonWorks and using them to devise and execute analytics as a Data Scientist
* Experience with the following: Python, Hadoop, Hive, and Spark (either Pyspark or Scala)
* Programming experience with the Hadoop ecosystem of applications and functional understanding of distributed data processing systems architecture (Data Lake / Big Data /Hadoop/ Spark / HIVE, etc).
* Conduct MapReduce programming with Apache Hadoop, the Hadoop Distributed File System (HDFS), and processing large data stores.
* Proficiency with Big Data frameworks such as Hadoop, Spark, MapReduce, HDFS.
* Experience or a working knowledge of data modeling in relational and NoSQL data stores (AWS DynamoDB, AWS Neptune)
* Experience with VMware vSphere Ecosystem; experience in VMware Cloud (VMC) on AWS and VMware NSX a plus
* Exposure to creating sustainable, automated processes for data analysis using tools such as Airflow and HDFS
* Handson keyboard role that might cover use of our data lake framework for loading / data ingestion, but also potentially have used Apache Airflow

|  |  |
| --- | --- |
| **Mitchell International** *Data Engineer* | San Diego, CA 92122 October 2016 - October 2018 |

* You have experience or knowledge of Big Data technologies such as Apache Spark, Apache Khafka, Hadoop.
* Experience with ElasticSearch, Apache Kafka. Apache Spark, Apache Hadoop, MongoDB
* Exposure to Hadoop Ecosystem preferably on AWS/EMR, NoSQLbased, SQLlike technologies
* Experience working within host vendor and cloudbased infrastructure; AWS and Rackspace preferred
* Airflow Experience building and monitoring DAGs, developing custom operators and using script templating solutions
* Experience in building, deploying and maintaining data pipelines using open source tools like Python and Airflow
* Experience with AWS platform and technologies including Kinesis, Firehose and Glue.
* Experience with AWS technologies such as EMR, S3, Glue and Redshift (Spectrum)
* Develop software to run on cloud native big data infrastructure built on AWS using Spark, Lambda, S3, and other cloud native services
* Experience working on Amazon Web Services (in particular using S3, SQS and the like)
* Review log files to ensure effectiveness of processes. May collaborate with vendors and IT personnel for problem resolution and interact with vendors, outsourcers, and contractors regarding systemrelated services
* Experience with AWS technologies such as EMR, S3, Glue and Redshift (Spectrum)
* Expertise in designing data warehouses using Google Big Query (Snowflake and Redshift will be considered)
* Write high quality SQL queries to retrieve and analyze data from database tables (ex. Redshift, MySQL)

|  |  |
| --- | --- |
| **Lionbridge Technologies** *Data Engineer* | nan October 2014 - October 2016 |

* Expert knowledge with delivering Big Data Cloudera Solutions in the cloud with AWS or Azure.
* Advanced experience with AD/LDAP security integration with Cloudera, including Sentry and ACL configurations
* Ability to troubleshoot and debug Hive/Impala query and perform query optimization
* Adopt industry best practices in solution engineering and supporting of Pillar 2 big data pipeline data ingestion / aggregation / curation / lineage and serving layer use cases via AWS S3, EMR Hive, Glue Catalog, Redshift Spectrum and StepFunction
* Expertise in developing big data pipelines using technologies like Kafka, Flume, or Storm.
* Familiarity with technologies in the data engineering ecosystem tooling including EMR/Hadoop, Spark/PySpark, Kafka/Kinesis/Flume, and Airflow/luigi/AWS Data Pipeline a big plus.
* Experience with machine learning and data science (AWS Glue, Amazon Athena, data lakes)
* Responsible for leading and implementing the installation and development of key technology initiatives including AI/ML (Data Robot, Sagemaker), Modeling, and migration from an onpremise SAS data warehouse to an AWS based Data Lake architecture

EDUCATION

|  |  |
| --- | --- |
| **universite d'avignon** *master's degree electronic* | avignon |