

Sr. Data Engineer/ Python programmer Sr. Data Engineer/Python programmer Sr. Data Engineer/
Python programmer - BlackRock Dublin, CA Experienced IT professional who has management,
team building and end-to-end solution skills with focus on leveraging information technology seeking
a leadership position using Big Data and data science to maximize business values Developer /
Engineer / Architect with over 8 years of development, administration and programming experience
with primary focus in "Big Data" and emphasis on Hadoop Advanced Hadoop/BigData/Analytic
Experience Looking for hands-on Functional Programming and OO Programming role Looking
for hands-on Spark, Kafka, Scala, Python and Pyspark programming role Big Data Specialties
Spark RDD and Spark DF development Pyspark using Pycharm development Worked with
Hadoop, HBase, Cascading, Zookeeper, Oozie, Hive, HiveQL, MapR, MogoDB, Pentaho & Pig
Worked with AWS Cloud, EC2, EMR, RedShift, S3 etc. Supported Hadoop and Verita as
administrator about 7 to 700 nodes. Supported and administrated users accounts and security
rules with BigData domain. Worked with Application Servers, Tomcat, Oracle and MySQL.
Experienced with distributed systems, large scale non-relational data stores, map-reduce systems,
data modeling, database performance, and multi-terabyte data warehouses. Experienced in a
SaaS environment that has an agile development process. Experienced in Java, Python or other
object oriented programming language. Experienced in Mobile, Linux, Unix, Android, MAC
platform. Experienced with server management, server operating systems(Windows, Linux, Unix),
and VMware. Experienced in testing APIs, Restful APIs. Experienced with full SW Lifecycle and
Agile practice Experienced with scripting language, in Perl, Python, Ruby, QTP for automated and
performance tests. Experienced with testing DNA sequence machines SW for regulatory approval.
Experienced testing networking and storage technologies, protocols and hardware. Experienced
with web services and SOAP UI testing. Experienced with Oracle, TOAD, and SQL. Experienced
in testing UI technologies such as HTML5. Experienced in Big Data/ Hadoop /
Cassandra/Memcached/NoSQL/MapReduce. Lead and coached about 2 to 20 engineers onshore,
offshore and nearshore. Lead the hiring process. Big/data and analytics platform administrator
(Vertica, Greenplum, Hadoop, Hive, Apache Drills, Sqoop, Kafka, AKKA and Docker). Experienced

with Display Advertising, Behavioral Ad Networks Recommended Engines, Personalization.

Experienced with Data Analytics, Data mining, Predictive Modeling. Work Experience Sr. Data Engineer/ Python programmer BlackRock - San Francisco, CA 2018 to Present Hand-on Data Lake Implementation in Big data for huge data warehouses Building data process, ETL transformation, data ingestion and platforms Coding in Pyspark, Python, Pandas, Kafka, Java, Scala and Spark SQL Coding SAP Sybase (ASE) SQL Coding Kafka and Spark streaming and massaging Converted existing spaghetti code to modular code Configured and set up, designed, performance enhancements, scalable systems using with java services exposure. Leading BigData / data warehouse application design, data quality, implementation and testing/deployment strategy. Leading in identification and troubleshooting processing issues impacting timely availability of data in the data warehouse or delivery of critical reporting within established SLAs.

Identified and recommended improvements in production application solutions or operational processes in support of data warehouse applications and business intelligence reporting (ie, data quality, performance, static and dynamic reports, etc.) Researched, managed and coordinated resolution of complex issues through root cause analysis as appropriate Built Spark Engine and BlackRock Sale Business Engine Researched and implemented Blackchain technology Built machine learning ETL pipeline Built deep learning model and data pipeline Utilized multiple large Hadoop (HDP) infrastructure Utilized Git, Jira, Prometheus, Docker, Kubernetes tools Sr. Data Architect/AWS Cloud Realtor.com - Santa Clara, CA 2016 to 2018 DB development and DB performance tuning Built Hadoop development eco platforms Built AirFlow workflows and testing scripts Docker Master and install/deployment Worked on DB persistence with Java Hibernate development in AWS Redshift performance tuning and testing Spark RDD and Airflow workflow designing and testing Report design using MicroStrategy Built lead recommender systems for realtor.com website Scala functional programming with Spark streaming Solr Programming, Hortonworks Programming, VM programming Docker programming Sr. Big data Architect/developer FRB - San Francisco, CA 2015 to 2015 Hand-on Spark RDD and Spark DF development Docker Master and install/deployment NoSQL database design and Spark SQL

development Unix/Hadoop administrator and development Expertise with Splunk UI/GUI development and operations roles. Prepared, arranged and tested Splunk search strings and operational strings. Managed the Splunk licenses and restricted the daily indexing limit. AWS EMR / Redshift administrator/development Cloudera Administrator/development Cloudera Navigator Administrator/development. Kerbero Administrator. Zoomdata administrator. Big data platform design and architecture/development Hive and Hbase development. Security office for big data POC Hadoop, AWS, Cloudera, Streaming, Kafka and big data tools. Work with scripts for Oozie and Hue systems. Cassandra, MongoDB, Hbase, Hive development. Cloudera Hue and Solr development. Data governance policy/management and big data security. Big data security and Kerberos development. Multi-tenant set up and development. Sr. DataBase Architect San Francisco, CA 2015 to 2015 Converting large SAS SQL scripts to HiveQL for Data Science Converting Oracle scripts to HiveQL programming for Data Science Architecture No-SQL databases and coding the MapReduce functions for over 10 years' data about emails campaigning results analytics Working with data science to prototype data model and programming the business logic with Hive, Pig Building database application on Hive, Pig, Hbase, MongoDB, Canssdra, Spark, Solr and shark on a large hadoop nodes cluster in house Backend No-SQL database design, development, architecture and testing Data gaverment Backend Big Data/Hadoop Engineer SS Network - Redwood City, CA 2014 to 2014 Spark RDD and Spark SQL/Hive SQL. No-SQL database modeling. AWS cloud computing and architecture with Hadoop on Big Data. Build cloud application on AWS, S3, EMR, Hive, Pig, Hbase, MongoDB, Canssdra, Spark. Backend No-SQL database design, development, architecture and testing. Big/Data and BI/BW Consultant Cisco - San Jose, CA 2014 to 2014 Created and maintained scripts and programs to Extract, Transform and Load data Automated ETL and aggregation processes using BASH/PHP and/or other scripting languages Created logging and monitoring elements to all Data Warehouse and ETL processes to allow Operations team to monitor Responded to Data Warehouse and ETL process alerts, with support from operations team. Created, maintained and automated processes to distribute data warehouse extracts to various

users Expertise with Splunk UI/GUI development and operations roles. Continually monitored measure and improved all data warehouse and ETL processes for speed, reliability and accuracy. Performed analyses and executed improvements regarding overall data quality, including making recommendations to the business regarding inputs. Sourced and/or created tools to deliver monitoring metrics and dashboard reports to allow the rest of the company to understand the quality and timeliness of data warehouse data. Installed and integrated 3rd-party data maintenance tools such as address cleaning software, etc. Used new technologies to service data warehousing needs such as Hadoop, column databases, etc. Cisco big data analytics platform development, testing and deployment Worked on data integration application design, development, and testing Used Pentaho data integration tool to create data integration jobs Used MapR to run Hadoop / YARN jobs Moved enterprise data from Oracle to Hive for big data platform Moved WebEx, voice, phone, video and email enterprise data to Hive and build data mart. Wrote use story, task, plan and testing cases Wrote scripts for automating testing Wrote design document for enterprise data on board Analyzed deployment log for errors Monitored big data staging environment Monitored hive staging environment Big/Data and BI consultant Cisco - Milpitas, CA 2013 to 2013 Hadoop /MapR BW/BI data warehouse project. Big data move from Informatica and TD. Loaded data from different source to new build Hadoop analyst platform Built Hadoop QA team Used Hadoop /Hive/Sqoop/bash to deployment the data load and query data. Monitored QA environment for Hadoop problems Testing report form MapR report tool Provided ideas/working process to other team Review other reports and code to understand coding logic Managed a technical team or functioning as a team lead. Worked with Hadoop stack (e.g. MapReduce, Sqoop, Pig, Hive, Hbase, Flume). related/complementary open source software platforms and languages (e.g. Java, Linux, Apache, Perl/ Python/PHP, Chef). Worked with ETL (Extract-Transform-Load) tools (e.g. Informatica, Talend, Pentaho). Worked with BI tools and reporting software (e.g. Microstrategy, Cognos, OBIEE, Pentaho). Worked with analytical tools, languages, or libraries (e.g. SAS, SPSS, R, Mahout). Supported business development activities to shape and communicate proposed solutions to client executives - Implemented of ETL

applications: Implemented of reporting applications Application/implementation of custom analytics support Administrated of relational databases Data migration from existing data stores Infrastructure and storage design Developed capacity plans for new and existing systems.

Big/Data and BI/ consultant Santa Clara, CA 2013 to 2013 Spark RDD development and No-SQL database modeling Design/Architecture Big Data Hadoop Testing Frameworks Building Hadoop cluster in Cloud and local Worked with AWS cloud environment: EMR, Redshift, S3 Worked with Scala, Java, Python, Kafka, AKKA, SQL, MongoDB, JSON, Avro, Tableau Worked with Git, SBT, Ant, Maven, Ganglia, Jenkins, Docker Worked with Hadoop, Hbase, Zookeeper, Oozie, Scalding? Spark? Shark Tested automation and coded in Scala Worked with ETL and reporting tools (OBIEE, SAP etc.) Coded on Scala, Kafka, Streaming, Spark and AWS services Coded on Pig, Small workflow, SQL and Hive Built Analytic Spark platform Built Ad revenue feed process

Built Big Data Projects testing platform Big/Data and BI consultant San Jose, CA 2006 to 2012 ATT SAP BW/BI/BigData development support Conducted detailed design applications developed on Hadoop platforms (Feature testing, Regression Testing, Acceptance Testing and Sanity Testing) Implemented Business analysis tools with Hadoop?MapReduce scripts from ETL data to data warehouse for BI and enterprise analysis platform Administrated Log analysis scripts for Business Analyst tool with HDFS Hadoop (file system level) Advised the Hadoop and analytical workloads IO optimizing solutions Running benchmarked Hadoop/HBase clusters Supported development of Hadoop and Vertica Analytics Platform activities. Storage architecture/QA Netapp Hadoop ecosystem administrator and support Tested NetApp Open Solution for Hadoop(NOSH) settings Supported and scaled Hadoop systems Supported cluster failed over test and document the results with various configurations Administrated enterprise-grade storage arrays and eliminated Hadoop network bottlenecks Supported hot-pluggable disk shelves, added storage and administrated services Supported NFS and HDFS file systems Loaded network-free hardware RAID Day-to-day support Hadoop hardware and software issues Used automation scripts to run performance testing of the new storage OS design.

Big Data System Consultant Bank of America TCOE - Concord, MA 2006 to 2009 Supported about over 100 projects and many functional teams

Supported analytic massive amounts of data and BI Communicated and tracked defects to closure Administrated with large data set testing with BigData NoSQL Hadoop and Oracle, DB2 scripts Wrote, reviewed, and executed scripts and tracked defects Expertise with Splunk UI/GUI development and operations roles. Used scripts languages in Perl and Python for Automations testing and test data condition Used scripts languages for ETL load with BW. Middleware Architect Kaiser - Walnut Creek, CA 2004 to 2006 Designed all middleware with 3000 interfaces for BigData.

IBM MQ design/ architecture support. Designed QA environment and production support. Designed Message broker and production support. Education Bachelors in Information Systems in Information Systems University of San Francisco 2004

Name: Joel Brown

Email: robert19@example.org

Phone: (814)596-5734x005