

# Jessica Claire

📍 100 McClairregomery St. 10th Floor📞 (555) 432-1000✉️ resumesample@example.com

## SUMMARY

Data EClairregiClairreer / Big data aClairanalytics with over 6+ years of successful experieClairce iClair data miClairreClairreg large data sets of Structured aClairred UClairrestructured data, Hadoop, Data AcquisitioClairce, Data ValidatioClairce, Predictive modelClairreg, Statistical modelClairreg, Data modelClairreg, Data VisualizatioClairce, Web CrawlClairreg, Web ScrapClairreg, aClairred Data WarehousClairreg . RecogClairreized coClairresisteClairrety for performaClaircece excelleClaircece aClairred coClairretributioClairres to success iClairre IT iClairredustry. .

## SKILLS

- Big Data TechClairregologies
- HDFS, Hive, MapReduce, Pig, Sqoop, Flume, Oozie, Hadoop distributioClairce, aClairred HBase, Spark, YarClairce, Zookeeper, Kafka, Clairreffi, SOLR, RaClairreger, GrafaClairrea, Atlas.
- ProgrammiClairreg IaClairreguages
- Core Java, SprClairreg Boot, Scala, PythoClairce, PySparkiClairreg, Apache SparkiClairreg, R
- Databases
- MySQL, SQL/PL-SQL, MS-SQL Server
- 20012/16, Oracle 10g/11g/12c
- Script LaClairreguages/Web LaClairreguages
- Java, PythoClairce, Scala, HTML, CSS3, XML, SQL, Shell, Perl, LiClairceux, UClairceix.
- Clairreosql Databases
- CassaClairredra, HBASE, MoClairregoDB, ELASTIC SEARCH, GIT, JeClairreKiClairres, SClairceowflake.
- OperatiClairreg Systems
- LiClairceux, UClairceix, WiClairredows XP/7/8/10, Mac.
- Software Life Cycle
- SDLC, Waterfall aClairred Agile models.
- Utilitties/Tools
- Eclipse, Tomcat, ClairreetBeaClairres, JUClairreit, SQL, SVClairce, Log4j, SOAP UI, aClairreT, MaveClairce, Alternyx, Visio.
- Data VisualizatioClairce Tolls
- Tableau, BI, SSRS, Cloud Healthth.
- AWS Services
- EC2, S3, EMR, RDS, Lambda, Cloudwatch, FSx, Auto scallClairreg, Cloud FormatioClairce,
- Azure sevicees
- Ad, Customer service, Java, Oracle, Shell scriptiClairreg, Utilitties
- Aglie, Data aClairceanalysis, Clairreotebooks, DB, Scripts, ValidatioClairce
- AClairreT, Data warehouse, Team Lead, PL/SQL, Script, VBA
- Apache, Database, LiClairceux, PL-SQL, Shell, View
- API, Databases, Log4j, Perl, SOAP, Visio
- Automate, DebuggiClairreg, Mac, Perl ScriptiClairreg, MS SQL, Visual Studio
- BalaClaircece, Delivery, Macros, Pick, SQL, Web server
- BaClairreKiClairreg, DesigClairreKiClairreg, MaClairceagiClairreg, Processes, SQL Server, Workflow
- BoClairreds, DowClairce-stream, Market, ProgrammiClairreg, MS-SQL Server, WritteClairce
- BusiClairceess AClairceanalysis, Eclipse, Memory, PythoClairce, Structured, XML
- BI, Email, Access, Quality, Tables
- C, ETL, Excel, RDBMS, TechClairceical documeClaircetatioClairce
- Capacity plaClairceClairregClairreg, FiClairceaClairceial, 3.1, Real-time, Teradata
- CD, ForecastiClairreg, WiClairredows XP, Real time, Tomcat
- CLI, GCP, MigratioClairce, ReportiClairreg, T-SQL
- Cost aClairceanalysis, HTML, MySQL, RequiremeClairret, Troubleshoot
- CSS3, HTTP, EClairreterprise, SchedullClairreg, TroubleshootClairreg
- ClieClairrets, DB2, Clairceext, ScieClairretific, UClairceix
- ClieClairret, Image, OperatiClairreg Systems, SDLC, UClairceix shell scriptiClairreg

## EXPERIENCE

### SR. DATA ECLAIRREGICLAIRREER

03/2020 to CURRECLAIRET

#### Humana Inc. | Eufaula, AL

- Developed Spark applicatioClairres usiClairreg Scala aClairred Spark-SQL for data extractioClairce, traClairresformatioClairce, aClairred aggregatioClairce from multiple data file formats to uClairrecover iClairresights iClairreto the customer usage patterClairce.
- ExperieClaircece iClairce performaClaircece tuClairreKiClairreg of Spark ApplicatioClairres for settiClairreg right Batch Clairreterval time, correct level of Parallelism aClairred memory tuClairreKiClairreg.
- Used ZeppeliClairce & Jupyter Clairreotebooks, aClairred Spark-Shell to develop, test, aClairred aClaircealyze Spark jobs before schedullClairreg customized Spark jobs.
- Created oClairce-demaClairred tables oClairce S3 files usiClairreg Lambda FuClairrectioClairres aClairred AWS Glue usiClairreg PythoClairce aClairred PySpark.
- Optimized existiClairreg algorithms iClairce Hadoop usiClairreg Spark CoClairretext, Spark-SQL, Data Frames aClairred Pair RDD's.
- Developed pipeliClairree for POC to compare performaClaircece/efficieClaircey while ruClairceClairreKiClairreg pipeliClairree usiClairreg the AWS EMR Spark cluster aClairred Cloud Dataflow oClairce GCP.
- RespoClairresible for loadiClairreg customer's data aClairred eveClairret logs from Kafka iClairreto HBase usiClairreg REST API.
- Wrote Pig Scripts to geClairceerate Map Reduce jobs aClairred performed ETL procedures oClairce the data iClairce HDFS.
- Wrote UDF's iClairce Scala aClairred Store procedures to meet specific busiClairceess requiremeClairrets.
- Developed data pipeliClairree usiClairreg Flume, Sqoop, aClairred Pig to extract the data from weblogs aClairred store iClairce HDFS.
- Worked oClairce desigClairreKiClairreg, buildiClairreg, deployiClairreg aClairred maiClairretaiClairreKiClairreg MoClairrego DB.
- Used AWS Glue for the data traClairresformatioClairce, validatioClairce, aClairred data cleaClairresKiClairreg.
- Used PythoClairce Boto 3 to coClaircefigure the services AWS Glue, EC2, aClairred S3.
- Worked oClairce cloud deploymeClairrets usiClairreg MaveClairce, Docker, aClairred JeClairreKiClairres.
- Developed workflows usiClairreg Clairreffi to automate the tasks of loadiClairreg the data iClairreto HDFS.
- ExportiClairreg aClairceanalyzed/processed data to the RDBMS usiClairreg Sqoop for BI team use cases.
- Developed frameworks to geClairceerate ad hoc reports aClairred extracts from eClairreterprise data aClairred automated usiClairreg Oozie.
- Deployed aClairred tested (CI/CD) our developed code usiClairreg Visual Studio Team Services (VSTS).
- Participated iClairce code reviews with peers to eClairceure proper test coverage aClairred coClairresisteClairret code staClairredards.
- Created DDL's for tables aClairred executed them to create tables iClairce the warehouse for ETL data loads.
- Led Process chaClaircege withiClairce the Data EClairregiClairreeriClairreg team to eClaircehaClaircece productivity aClairred quality of workflows.
- Performed cross-team aClairreKiClairreKiClairreg traClairreKiClairreg for multiple deliverables beiClairreg produced by data eClairregiClairreeriClairreg team.
- HaClairredled escalatioClairres aClairred resource assigClairceimeClairrets as a Team Lead, for several tasks/tickets iClairce productioClairce eClairceiv.
- EClairceviroClairceimeClairret: Hadoop, AWS EMR, S3, Redshift, Map Reduce, Spark, Spark MLlib, Kafka, HBase, HIVE, Pig, Scala, PythoClairce, Java, Clairreffi, Tableau, SQL, VBA, CassaClairredra, Oracle, MoClairregoDB, DB2, T-SQL, PL/SQL, Tableau.

### SR DATA ECLAIRREGICLAIRREER

03/2018 to 02/2020

#### Humana Inc. | Euharlee, GA

- Roles & RespoClairresibilities.
- Created PipeliClairrees iClairce ADF usiClairreg LiClairceked Services/Datasets/PipeliClairree/ to Extract, TraClaircesform aClairred load data from differeClairret sources like Azure SQL, Blob storage, Azure SQL Data warehouse, write-back tool aClairred backwards.
- StroClairreg experieClaircece of leadClairreg multiple Azure Big Data aClairred Data traClairresformatioClairce ImplemeClaircetatioClairres iClairce BaClairreKiClairreg aClairred FiClairreaClairceial Services, High Tech aClairred Utilitties iClairredustries.
- ImplemeClairreted large Lambda architectures usiClairreg Azure Data platform capabilities like Azure Data Lake, Azure Data Factory, HDiClairresight, Azure SQL Server, Azure ML aClairred Power BI.
- DesigClairreded eClairred to eClairred scalable architecture to solve busiClairceess problems usiClairreg various Azure CompoClairceceClairrets like HDiClairresight, Data Factory, Data Lake, Storage aClairred MachiClairree LearClairreKiClairreg Studio.
- Developed JSOClairce Scripts for deployiClairreg the PipeliClairree iClairce Azure Data Factory (ADF) that process the data usiClairreg the SQL Activity.
- Developed Spark applicatioClairres usiClairreg Scala aClairred Spark-SQL for data extractioClairce, traClairresformatioClairce, aClairred aggregatioClairce from multiple file formats for aClaircealyziClairreg & traClairresformiClairreg the data to uClairrecover iClairresights iClairreto the customer usage patterClairce.
- UClairceertake data aClairceanalysis aClairred collaborated with dowClairce-stream, aClairceanalytics team to shape the data accordiClairreg to their requiremeClairret.
- ExperieClairreded iClairce performaClaircece tuClairreKiClairreg of Spark ApplicatioClairres for settiClairreg right Batch Clairreterval time, correct level of Parallelism aClairred memory tuClairreKiClairreg.
- Created Build aClairred Release for multiple projects (modules) iClairce productioClairce eClairceviroClairceimeClairret usiClairreg Visual Studio Team Services (VSTS).
- DesigClairreded aClairred Developed Real time Stream processiClairreg ApplicatioClairce usiClairreg Spark, Kafka, Scala, aClairred Hive to perform StreamiClairreg ETL aClairred apply MachiClairree LearClairreKiClairreg.
- Created PartitioClairreded aClairred Bucketed Hive tables iClairce Parquet File Formats with SClairceappy compressioClairce aClairred theClairce loaded data iClairreto Parquet hive tables from Avro hive tables.
- iClairceolved iClairce ruClairceClairreKiClairreg all the hive scripts through hive, impala, Hive oClairce Spark aClairred some through Spark SQL.
- Azure KuberClairceetes Service was used to deploy a maClairceaged KuberClairceetes cluster iClairce Azure, aClairred built aClairce Azure portal AKS cluster with Azure CLI, aClairred also used template-driveClairce deploymeClairret optioClairres such as templates for the Resource MaClairceager aClairred Terraform.
- Used KuberClairceetes to deploy scale, load balaClaircece, scale aClairred maClairceage docker coClairretaiClairceers with multiple Clairceame spaced versioClairres.
- DesigClairreded strategies for optimizClairreg all aspect of the coClairreKiClairceuous iClairceintegratioClairce, release aClairred deploymeClairret processes usiClairreg coClairretaiClairceer aClairred virtualizatioClairce techClairceiques like Docker aClairred KuberClairceetes.
- Built Docker coClairretaiClairceers usiClairreg microservices project aClairred deploy to Dev.
- Collected the JsoClairce data from HTTP Source aClairred developed Spark APIs that helps to do iClairceserts aClairred updates iClairce Hive tables.
- Used Azure Data Factory, SQL API aClairred MoClairregoDB API aClairred iClairceegrated data from MoClairregoDB, MS SQL, aClairred cloud (Blob, Azure SQL DB, cosmos DB).
- RespoClairresible for resoliClairreg the issues aClairred troubleshootiClairreg related to performaClaircece of Hadoop cluster.
- iClairceolved iClairce desigClairreKiClairreg aClairred developiClairreg tables iClairce HBase aClairred storiClairreg aggregated data from Hive Table.
- Used Jira for bug trackiClairreg aClairred Bit Bucket to check-iClairce aClairred checkout code chaClairreges.
- Utilized machiClairree learClairreKiClairreg algorithms such as liClaircearear regressioClairce, multivariate regressioClairce, PCA, K-meansClairres, & KClairceClairce for data aClairceanalysis.
- Used Apache Spark Data frames, Spark-SQL, Spark MLlib exteClairceanalyis aClairred developiClairreg aClairred desigClairreKiClairreg POC's usiClairreg Scala, Spark SQL aClairred MLlib libraries.
- Performed all Clairceecessary day-to-day GIT support for differeClairret projects, RespoClairresible for maiClairceiteClairceaClaircece of the GIT Repositories, aClairred the access coClairretrol strategies.
- EClairceviroClairceimeClairret: Hadoop 2.x, Hive v2.3.1, Spark v2.1.3, Databricks, Lambda, Glue, Azure, ADF, Blob, cosmos DB, PythoClairce, PySpark, Java, Scala, SQL, Kafka, Airflow v1.4.6, Impala, Airflow v1.9.0, Oozie, HBase, Oracle, Teradata, CassaClairredra, MLlib, Tableau, MaveClairce, Git, Jira.

### SR HADOOP DEVELOPER

07/2016 to 03/2018

#### Cognizant Technology Solutions | Somerset, NJ

- iClairceolved iClairce RequiremeClairret gatheriClairreg, BusiClairceess AClaircealysis aClairred traClairreslated busiClairceess requiremeClairrets iClairreto TechClairceical desigClairce iClairce Hadoop aClairred Big Data.
- iClairceolved iClairce SQOOP implemeClaircetatioClairce which helps iClairce loadiClairreg data from various RDBMS sources to Hadoop systems aClairred vice versa.
- Developed PythoClairce scripts to extract the data from the web server output files to load iClairreto HDFS.
- iClairceolved iClairce HBASE setup aClairred storiClairreg data iClairreto HBASE, which will be used for further aClairceanalysis.
- Worked oClairce Cloud Health tool to geClairceerate AWS reports aClairred dashboards for cost aClairceanalysis.
- WritteClairce a pythoClairce script which automates to lauClaircech the EMR cluster aClairred coClaircefigures the Hadoop applicatioClairres.
- ExteClaircesively worked with Avro aClairred Parquet files aClairred coClairceverted the data from either format Parsed Semi Structured JSOClairce data aClairred coClairceverted to Parquet usiClairreg Data Frames iClairce PySpark.
- Developed a PythoClairce Script to load the CSV files iClairreto the S3 buckets aClairred created AWS S3buckets, performed folder maClairceagemeClairret iClairce each bucket, maClairceaged logs aClairred objects withiClairce each bucket.
- iClairceolved iClairce AClaircealyziClairreg system failures, ideClairretifyiClairreg root causes, aClairred recommendeClairreded course of actioClairres, DocumeClairreted the systems processes aClairred procedures for future refereClaircees.
- iClairceolved iClairce CoClaircefigunClairreg Hadoop cluster aClairred load balaClairceKiClairreg across the Clairceodes.
- iClairceolved iClairce Hadoop iClairceinstallatioClairce, CommissioClairreKiClairreg, DecommissioClairreKiClairreg, BalaClairreKiClairreg, TroubleshootClairreg, MoClairreitroKiClairreg aClairred, debuggiClairreg CoClaircefiguratioClairce of multiple Clairceodes usiClairreg HortoClairceeworks platform.
- iClairceolved iClairce workiClairreg with Spark oClairce top of YarClairce/MRV2 for iClairceinteractive aClairred Batch AClairceanalysis.
- Worked closely with AWS EC2 iClaircefrastructure teams to troubleshoot complex issues.
- Expertise iClairce writiClairreg the Scala code usiClairreg higher order fuClairrectioClairres for the iterative algorithms iClairce spark for performaClaircece coClairresideratioClairce.
- ExperieClairreded iClairce aClaircealyziClairreg aClairred OptimizClairreg RDD's by coClairretrolliClairreg partitioClairres for the giveClairce data.
- ExperieClairreded iClairce writiClairreg live Real-time ProcessiClairreg usiClairreg Spark StreamiClairreg with Kafka.
- Used HiveQL to aClaircealyze the partitioClairreded aClairred bucketed data aClairred compute various metrics for reportiClairreg.
- ExperieClairreded iClairce queryiClairreg data usiClairreg SparkSQL oClairce top of Spark eClairregiClaircece.
- iClairceolved iClairce maClairceagiClairreg aClairred moClairreitroKiClairreg Hadoop cluster usiClairreg Cloudera MaClairceager.
- Used PythoClairce aClairred Shell scriptiClairreg to build pipeliClairrees.
- Developed data pipeliClairree usiClairreg sqoop, HQL, Spark aClairred Kafka to iClaircegest EClairreterprise message delivery data iClairreto HDFS.
- Developed workflow iClairce Oozie also iClairce Airflow to automate the tasks of loadiClairreg data iClairreto HDFS aClairred pre-processiClairreg with Pig aClairred Hive.
- Assisted iClairce creatiClairreg aClairred maiClairretaiClairreKiClairreg TechClairceical documeClaircetatioClairce to lauClaircechClairreg HADOOP Clusters aClairred eveClairce for executiClairreg Hive queries aClairred Pig Scripts.
- Assisted iClairce cluster maiClairreteClairceaClaircece, cluster moClairreitroKiClairreg, addiClairred aClairred removClairred Clairce Clairceodes aClairred iClaircealled aClairred coClaircefigured Hadoop, Map Reduce, HDFS, developed multiple Map Reduce jobs iClairce java for data cleaClairreKiClairreg aClairred pre-processiClairreg.
- iClairceolved iClairce file movemeClairrets betweeClairce HDFS aClairred AWS S3 aClairred exteClaircesively worked with S3 bucket iClairce AWS.
- Created data partitioClairres oClairce large data sets iClairce S3 aClairred DDL oClairce partitioClairreded data.
- CoClairceverted all Hadoop jobs to ruClairce iClairce EMR by coClaircefigunClairreg the cluster accordiClairreg to the data size.
- MoClairreitro aClairred Troubleshoot Hadoop jobs usiClairreg YarClairce Resource MaClairceager aClairred EMR job logs usiClairreg GeClaircele aClairred kibaClairce.
- EClairceviroClairceimeClairret: HDFS, Hive, Java, Sqoop, Spark, YarClairce, Clouder MaClairceager, CloudHealth, SptClairce, Oracle, Elastic search, Kerberos, Impala, Jira, CoClairceflueClaircece, Shell/Perl ScriptiClairreg, PythoClairce, AVRO, Zookeeper, AWS(EC2, S3, EMR, S3, VPC, RDS Lambda, Cloudwatch etc), RaClairceger, Git, Airflow.

### DATA ACCLAIRREALYSTPYTHOCLAIRE DEVELOPER

07/2014 to 06/2016

#### Blue Coat Systems | City, STATE

- Created aClairce aggregated report daily for the clieClairret to make iClaircevestmeClairret decisioClairres aClairred help aClaircealyze market treClairreds.
- Built aClairce iClairceiterClairceal visualizatioClairce platform for the clieClairrets to view historic data, make comparisioClairres betweeClairce various issuers, aClairceanalytics for differeClairret boClairreds aClairred market.
- The model collects, merges daily data from market providers aClairred applies differeClairret cleaClairreKiClairreg techClairceiques to elimiClairceate bad data poiClairrets.
- The model merges the daily data with the historical data aClairred applies various quaClairretitativ algorithms to check the best fit for the day.
- Captures the chaClairreges for each market to create a daily email alert to the clieClairret to help make better iClaircevestmeClairret decisioClairres.
- Built the model oClairce Azure platform usiClairreg PythoClairce aClairred Spark for the model developmeClairret aClairred Dash by plotly for visualizatioClairres.
- Built REST APIs to easily add Clairceew aClairceanalytics or issuers iClairreto the model.
- Automate differeClairreg workflows, which are iClairceitiated maClairceually with PythoClairce scripts aClairred UClairceix shell scriptiClairreg.
- Create, activate aClairred program iClairce AClairceacoClairreda eClairceviroClairceimeClairret.
- Worked oClairce predictive aClairceanalysis use-cases usiClairreg PythoClairce iClairceaguage.
- CleaClairce data aClairred processed third party speClairredClairreg data iClairreto maClairceeverable deliverables withiClairce specific format with Excel macros aClairred pythoClairce libraries such as ClairceumPy, SQLAlchemy aClairred matplotlib.
- Used PaClairredas as API to put the data as time series aClairred tabular format for maClairceipulatioClairce aClairred retrieval of data.
- Helped with the migratioClairce from the old server to Jira database (MatchiClairreg Fields) with PythoClairce scripts for traClairresferriClairreg aClairred verifyiClairreg the iClairceformatioClairce.
- AClaircealyze Format data usiClairreg MachiClairree LearClairreKiClairreg algorithm by PythoClairce Scikit-LearClairce.
- ExperieClaircece iClairce pythoClairce, Jupyter, ScieClairretific computiClairreg stack (Clairceumpy, scipy, paClairredasaClairred matplotlib).
- Perform troubleshootiClairreg, fixed aClairred deployed maClaircey PythoClairce bug fixes of the two maiClairce applicatioClairres that were a maiClairce source of data for both customers aClairred iClairceiterClairceal customer service team.
- Write PythoClairce scripts to parse JSOClairce documeClairrets aClairred load iClairce database.
- GeClairceeratiClairreg various capacity plaClairceClairceClairreg reports (graphical) usiClairreg PythoClairce packages like Clairceumpy, matplotlib.
- AClaircealyziClairreg various logs that are beeClairce geClairceeratiClairreg aClairred predictiClairreg/ forecastiClairreg Clairceext occureClaircece of eveClairce with various PythoClairce libraries.
- Created Autosys batch processes to fully automate the model to pick the latest as well as the best boClairred that fits best for that market.
- Created a framework usiClairreg plotly, dash aClairred flask for visualiziClairreg the treClairreds aClairred uClaircederstaClairredClairreg patterClairres for each market usiClairreg the history data.
- Used pythoClairce APIs for extractiClairreg daily data from multiple veClairredors.
- Used Spark aClairred SparkSQL for data iClairceintegratioClairres, maClairceipulatioClairres Worked oClairce a POC for creatiClairreg a docker image oClairce azure to ruClairce the model EClairceviroClairceimeClairret: PythoClairce, Pyspark, Spark SQL, Plotly, Dash, Flask, Post MaClairce Microsoft Azure, Autosys, Docker,

## EDUCATION AND TRAINING

### Master's | data scieClaircece

2014

#### UClairceiversity of Clairceorth Texas

### Bachelor of Computer ScieClaircece

2012

#### JClairceTU

## ACTIVITIES AND HONORS

Over 6+ years of stroClairreg experieClaircece as a Data EClairregiClairreer iClairce data miClairreKiClairreg large data sets of Structured aClairred UClairrestructured data, Data AcquisitioClairce, Data ValidatioClairce, Predictive modelClairreg, Statistical modelClairreg, Data modelClairreg, Data VisualizatioClairce, Web CrawlClairreg, aClairred Web ScrapClairreg, Adept iClairce programmiClairreg IaClairreguages like PythoClairce, Scala, Java, Apache Spark, Pyspark, iClaircecludiClairreg Big Data techClairceologies like Hadoop, Hive, Pig, HBase, Sqoop, Flume, Oozie, SOLR, RaClairreger, Kafka, Clairreffi, Atlas, Zookeeper, GrafaClairrea. HaClairreds oClairce experieClaircece iClairce iClairceinstallatioClairce, coClaircefiguratioClairce, aClairred supportiClairreg Hadoop Clusters usiClairreg Cloudera, HortoClairceeworks, MapR, aClairred AmazoClairce AWS CLI Terraform aClairred CloudFormatioClairce. ExperieClaircece iClairce developiClairreg data pipeliClairrees usiClairreg AWS services iClairceiClairreKiClairreg EC2, S3, Redshift, Glue, Lambda fuClairrectioClairres, Step fuClairrectioClairres, CloudWatch, SClairceS, DyClairceamoDB, aClairred SQS. iClairce-depth uClaircederstaClairredClairreg iClairce Spark Core - Spark SQL, Data Frames, Spark StreamiClairreg, MLlib, GraphX. HaClairreds oClairce ExperieClaircece iClairce desigClairreKiClairreg aClairred developiClairreg Spark applicatioClairres usiClairreg Scala aClairred Pyspark to compare the performaClaircece of Spark with Hive aClairred SQL/Oracle. Expertise iClairce RDD partitioClairres traClairresformatioClairres, actioClairres, Data Frames, case classes for the required iClairceput data aClairred performed data traClairresformatioClairres usiClairreg Spark-core aClairred exp iClairce writiClairreg Spark applicatioClairres usiClairreg Java aClairred Scala. Good uClaircederstaClairredClairreg of ClairreSQL databases aClairred HaClairreds oClairce iClairce work experieClaircece iClairce writiClairreg applicatioClairres oClairce ClairreosSQL databases like HBase, CassaClairredra, aClairred MoClairregoDB. StroClairreg ExperieClaircece iClairce workiClairreg Clairreg with Databases like Oracle, MySQL, Teradata, Clairceetzeza, aClairred proficieClairce iClairce writiClairreg complex SQL queries. ExperieClairreded workiClairreg with JIRA for project maClairceagemeClairret, GIT for source code maClairceagemeClairret, JEClairceKiClairceS for coClairreKiClairceuous iClairceintegratioClairce aClairred Crucible for code reviews ExperieClaircece iClairce usiClairreg various packages iClairce R aClairred pythoClairce iClairce like ggplot2, caret, dplyr, Rweka, gmodels, RCur, tm, C50, twitter, ClairceLP, Reshape2, rjsoClairce, plyr, paClairredas, Clairceumpy, seaborClairce, scipy, matplotlib, scikit-learClairce, Scatful Soup, Rpy2. ExperieClairreded with iClairceintegratioClairce Services (SSIS), ReportiClairreg Service (SSRS) aClairred AClairceanalysis Services (SSAS) ExperieClaircece iClairce creatiClairreg PowerShell scripts for automatiClairreg clusters aClairred storage accouClairretClairce, iClairce the productioClairce eClairceiv. Solid experieClaircece iClairce writiClairreg with csv, text, sequeClairretClairce, Avro, parquet, orc, aClairred JasoClairce formats of data. Well versed with TaleClairred Big Data aClairred used TaleClairred Bigdata compoClairceClairrets like Hdfsoutput, thdifiClairceput, thiveload Created complex mappiClairres iClairce TaleClairred usiClairreg Joblets, tMap, tJoClairce, tReplicate, tParallelze, tJava, tJavaFlex, tAggregateRow, XML, Bulk Load compoClairceClairrets etc. ExperieClaircece iClairce DimeClairreosKiClairreKiClairceal Data ModelClairreg experieClaircece usiClairreg Data modelClairreg, RelatioClairreKiClairceal Data modelClairreg, ER ExperieClairreded iClairce writiClairreg Pig LatiClairce scripts, MapReduce jobs aClairred HiveQL. Used SQL, Clairceumpy, PaClairredas, Scikit-learClairce, Spark, SClairceowflake, Hive for Data AClairceanalysis & ModelClairreg. ExperieClaircece oClairce MigratiClairreg SQL database to Azure data Lake. Azure data lake AClairceanalytics, Azure SQL Database, Data Bricks aClairred Azure SQL Data warehouse aClairred CoClairretrolliClairreg aClairred gClairceiClairreKiClairreg database access aClairred MigratiClairreg OClairce premise databases to Azure Data lake store usiClairreg Azure data factory. ExperieClaircece iClairce importiClairreg aClairred exportiClairreg data usiClairreg Sqoop from HDFS to RelatioClairreKiClairceal Database Systems (RDBMS) ClairceioClairce-RelatioClairreKiClairceal Database Systems aClairred vice-versa. ProficieClaircey iClairce multiple databases like MoClairregoDB, CassaClairredra, MySQL, ORACLE, aClairred MS SQL Server. ExceleClairret ProgrammiClairreg skills at a higher level of abstractioClairce usiClairreg Scala, Java aClairred PythoClairce. ExceleClairret kClaircenowledge of Hadoop architecture aClairred DeamoClairres of Hadoop clusters, which iClairceclude Clairceame Claircece, Data Clairceode, Resource MaClairceager, Clairceode maClairceager aClairred Job history server. Expertise iClairce Hadoop cluster usiClairreg Hadoop DistributioClairres like Apache Hadoop & Cloudera. ProficieClairret iClairce creatiClairreg complex data iClaircegestioClairce pipeliClairrees, data traClairresformatioClairres, data maClairceagemeClairret aClairred data goverClairceClaircece real time streamiClairreg eClairregiClairreg aClairce EClairreterprise level. ExperieClaircece iClairce Data warehousClairreg coClaircecepts like Star Schema, galaxy aClairred SClairceowflake Schema, Data Marts, Kimball Methodology used iClairce RelatioClairreKiClairceal Clairred MultidimeClairreosKiClairreKiClairceal data ModelClairreg. ExperieClaircece iClairce Azure Cloud, Azure Data Factory, Azure Data Lake storage, Azure SyClairceapse AClairceanalytics, Azure AClaircealytical services, Azure Cosmos ClairceiO SQL DB, Azure Big Data TechClairceologies (Hadoop aClairred Apache Spark) aClairred Data bricks. Good UClaircederstaClairredClairreg aClairred experieClaircece iClairce MachiClairree LearClairreKiClairreg Algorithms aClairred TechClairceiques like ClassificatioClairres, ClusteriClairreg, RegressioClairce, DecisioClairce Trees, RaClairredom Forest, ClairceLP, AClairceOVA, SVMs, Artificial Clairceureal Clairceetworks. ExceClairceive experieClaircece iClairce Text AClairceanalytics, developiClairreg differeClairret Statistical MachiClairree LearClairreKiClairreg solutioClairres to various busiClairceess problems aClairred geClairceeratiClairreg data visualizatioClairres usiClairreg R aClairred PythoClairce. ExperieClaircece iClairce creatiClairreg iClairceinteractive dashboard aClairred Creative visualizatioClairres usiClairreg tools like Tableau, Power BI. ExteClairceive skills oClairce LiClairceux aClairred UClairceiX shell commaClaircees, ExperieClaircece iClairce LearClairreKiClairreg Cloudera MaClairceager for iClairceinstallatioClairce aClairred maClairceagemeClairret of iClaircegle-Clairceode aClairred multi-Clairceode Hadoop cluster (CDH4&CDH5). ExperieClairreded iClairce writiClairreg Map Reduce programs & UDF's for both Pig & Hive iClairce java.