BigData/Spark Developer & MapR Platform support BigData/Spark Developer & MapR Platform support BigData/Spark Developer & MapR Platform support - ExxonMobil Chicago, IL   7+ years of IT professional experience with emphasis on Design, Development, Implementation, Testing, Maintenance and Deployment of Software Applications using Java, J2EE and Big Data technologies.   Big Data technologies and Hadoop ecosystem components like Spark, HDFS, Map Reduce, Pig, Hive, YARN, Sqoop, Flume, Kafka and NoSQL systems like HBase, Cassandra. Expertise in architecting real time streaming applications and batch style large scale distributed computing applications using tools like Spark Streaming, Kafka, Hive, Impala etc.   Hands-on experience in installing, configuring and monitoring HDFS clusters (on premise & cloud AWS). Experience in the IT industry as an Ab-Initio Developer.   Developed prototype for parsing high volume XML files using Hadoop and storing output in HDFS for Ab Initio.   Implemented detailed systems and services monitoring using Nagios, Zabbix, & AWS Cloud Watch.   In depth understanding of Map Reduce and AWS cloud concepts and its critical role in data analysis of huge and complex datasets.   Created custom Database Encryption Connectors that could be plugged into Sqoop to be able to encrypt the data while importing to HDFS/Hive.   Integrated Maven with Jenkins for the builds as the Continuous Integration process.   Experience in fine-tuning and troubleshooting Spark Applications, Hive queries.   Extensive hands on experience in writing complex map Reduce jobs, Pig Scripts and HiveQL scripts.   Experience using various Hadoop Distributions (Cloudera, Horton works, MAPR and Amazon AWS) to fully implement and leverage new Hadoop features.   Experience in Continuous Integration and Deployments (CI/CD) using build tools like Jenkins, Team City, MAVEN, and ANT. Wrote scripts to automate Build.   Build, manage, and continuously improved the build infrastructure for global software development engineering teams including implementation of build scripts, continuous integration infrastructure and deployment tools.   Experience in using Amazon Cloud services like S3, EMR etc.   Experience in Apache Flume and Kafka for collection, aggregation and moving huge chunks of data from various sources such as web server, telnet sources etc.   Worked on Java HBase API for ingesting processed data to HBase tables.   Extensive knowledge and experience in using Apache Storm,

Spark Streaming, Apache Spark, Apache NiFi, Kafka and Flume in creating data streaming solutions.    Experienced in working with Machine learning libraries (spark-MLlib) and implementing ML algorithms for clustering, regression filtering and dimensional reduction.    Extensive understanding of Partitions and bucketing concepts in Hive.    Created few Hive UDF's to perform some complex business specific transformations and rules.    Expert knowledge over J2EE Design Patterns like MVC Architecture, Session Facade, Front Controller and Data Access Objects for building J2EE Applications.    Experienced in using agile methodologies including extreme programming, Scrum Process and Test-Driven Development (TDD).    Use of Ansible for environment automation, configuration management and provisioning Setting up playbooks to deploy, manage, test and configure software onto the hosts.    Integrate Data Meer with tools and cloud-based platforms within the big data ecosystem (e.g. Spark, Tez, Azure, HDI, Amazon EMR, Redshift, Google DataProc)    Intensive work experience in developing enterprise solutions using Java, J2EE, Struts, Servlets, Hibernate, JavaBeans, JDBC, JSP, JSF, JSTL, MVC, Spring, Custom Tag Libraries, JNDI, AJAX, SQL, JavaScript, AngularJS and XML.    Conversant with web application Servers like Tomcat, Web Sphere, Web Logic and JBoss servers.    Experience in development of logging standards and mechanism based on Log4j.    Experience in writing ANT and Maven scripts to build and deploy Java applications. Authorized to work in the US for any employer Work Experience BigData/Spark Developer & MapR Platform support ExxonMobil - Houston, TX October 2017 to Present Description: ExxonMobil is the world's largest publicly traded international oil and gas company, providing energy that helps underpin growing economies and improve living standards around the world, our business wants to cut down the cost of Teradata maintenance and client want to migrate the data to Hadoop form Teradata, so we started using Hadoop eco system. We developed a utility called dart which will bring the data from any database to Hadoop and it is fully automated process. After client famine transformation on that data with the help of Map-Reduce programming and implemented data warehouse in Map Reduce, Hive and Pig. Then we integrated Hive and HBase with reporting and generated multiple reports as per the client requirements. Responsibilities:    Worked on analysing Hadoop cluster and different big data analytical and

processing tools including Sqoop, Pig, Hive, Spark, Kafka and Pyspark. Worked on MapR platform team for performance tuning of hive and spark jobs of all users. Using Hive TEZ engine to increase the performance of the applications. Working on incidents created by users for platform team on hive and spark issues by monitoring hive and spark logs and fixing it or else by raising MapR cases. Analysed large amounts of data sets to determine optimal way to aggregate and report on it. Tested the cluster Performance using Cassandra-stress tool to measure and improve the Read/Writes. Worked on Hadoop Data Lake for ingesting data from different sources such as oracle and Teradata through INFOWORKS ingestion tool. Worked on ARCADIA for creating analytical views on top of tables as if the batch is loading also no issue in reporting or table locks as it will point to arcadia view. Worked on Python API for converting assigned group level permissions to table level permission using MapR ace by creating a unique role and assigning through EDNA UI. Queried and analysed data from Cassandra for quick searching, sorting and grouping through CQL. Migrating various Hive UDF's and queries into Spark SQL for faster requests. Configured to receive real time data from the Apache Kafka and store the stream data to HDFS using Kafka connect. Hands on experience in Spark using Scala and python creating RDD's, applying operations -Transformation and Actions. Extensively perform complex data transformations in Spark using Scala language. Involved in converting Hive/SQL queries into Spark transformations using Scala. Used Pyspark and Scala languages to process the data. Used Bitbuket and Git repositories. Used text, AVRO, ORC and Parquet file formats for Hive tables. Experienced Scheduling jobs using Crontab. Used sqoop to import data from Oracle, Teradata to Hadoop. Used TES Scheduler engine to manage interdependent Hadoop jobs and to automate several types of Hadoop jobs such as Java map-reduce Hive, Pig, Spark, Kafka and Sqoop. Experienced in creating recursive and replicated joins in hive. Experienced in developing scripts for doing transformations using Scala. Written Map Reduce code to process and parsing the data from various sources and storing parsed data into HBase and Hive using HBase-Hive Integration. Experienced in creating the shell scripts and made jobs automated. Environment: HDFS, Hadoop 2.x, Pig, Hive, Sqoop, Flume, Spark, Map Reduce, Scala, Oozie, YARN, Tableau,

Spark-SQL, Spark-MLlib, Impala, Nagios, UNIX Shell Scripting, Zookeeper, Kafka, Agile Methodology, MAPR 6.0, SBT. Hadoop Developer Wells Fargo - Charlotte, NC September 2016 to October 2017 Description: Though Wells Fargo Millions of transactions happen every day and each transaction generate many events. We capture over 190 GB of event data every day. The main objective of this Project is to analyze this Transactional Data and use vast data repository to improve the business model. Responsibilities: Worked on importing and exporting data from Teradata, MySQL into HIVE using Sqoop for visualization, analysis and to generate reports. Loaded CSV files containing user event information into Hive External tables on daily basis. Created Spark applications and used Data frames and Spark-SQL API primarily for performing event enrichment and performing lookups with other enterprise data sources. Build a real time streaming pipeline by using Kafka integration with Storm and Spark Streaming. Perform ETL on different formats of data like JSON, CSV files and converted them to parquet while loading to final tables. Ran ad-hoc querying using Hive and Impala. Extensively perform complex data transformations in Spark using Scala language. Worked extensively with importing metadata into Hive using Scala and migrated existing tables and applications to work on Hive and AWS cloud. Involved in converting Hive/SQL queries into Spark transformations using Scala. Connect Tableau and Squirrel SQL clients to Spark-SQL (Spark thrift server) via data source and run the queries. Extensively worked on Informatica IDE/IDQ. Involved in massive data profiling using IDQ (Analyst Tool) prior to data staging. Extensively worked with the Ab Initio Enterprise Meta Environment EME to obtain the initial setup variables and maintaining version control during the development effort. Designed Ab Initio graphs that would harness Teradata capabilities ELT and balance resources between database and Ab Initio. Worked with Machine learning libraries (Spark MLlib) and done clustering, regression, filtering and dimensional reduction using implemented ML algorithms. Configured Spark streaming to receive real time data from the Kafka and store the stream data to HDFS using Scala (Prototype). Used Impala as the primary analytical tool for allowing visualization servers to connect and perform reporting on top Hadoop directly. Used Machine learning libraries as Spark MLlib and implemented ML algorithms for clustering, regression filtering and dimensional reduction

process with data scientists. Installed and configured monitoring tools Nagios for critical applications. Wrote spark jobs for Data clustering and data processing using Spark-MLlib and cluster algorithms as per functional requirements. Working on Oozie workflow engine to run multiple Hive-QL jobs and on schedulers. Done various compressions and file formats like Parquet, Snappy, Gzip, Bzip2, Avro, and Text. Implemented test scripts to support test driven development and continuous integration. Used Zookeeper to provide coordination services to the cluster. Used Impala and Tableau to create various reporting dashboards. Environment: HDFS, Hadoop 2.x, Pig, Hive, Sqoop, Flume, Spark, Map Reduce, Scala, Oozie, YARN, Tableau, Spark-SQL, Spark-MLlib, Impala, Nagios, UNIX Shell Scripting, Zookeeper, Kafka, Agile Methodology, Cloudera 5.9, SBT. Oracle ETL/PLSQL Developer Chamberlain - Chicago, IL August 2015 to September 2016 Description: The Scope of the project is to create an EDW for 4 business units available and provide BI functionality to their sites. Project is divided into multiple interfaces like Location, Product, Product Hierarchy, Store Plans, Merch Plans, and Allocation. Source is called JESTA for all the interfaces is Oracle Database with Front end Forms. Data from JESTA is Moved into Views and Data is Populated into ODS from the Views. Data from ODS is sent to EDW and EP (Enterprise Planning). Responsibilities: Developed PL/SQL Packages, Procedures and Functions accordance with Business Requirements for loading data into database tables. Created Materialized Views and Partition tables for performance reasons. Worked on various backend Procedures and Functions using PL/SQL. Developed UNIX shell scripts to perform a nightly refresh of the test system from Production databases. Coordinate with the front-end design team to provide them with the necessary stored package and procedures and necessary insight into the data. Developed Informatica mappings to move data from stage to target tables Involved in all phases of the SDLC for Designing & giving recommended approach to satisfy the requirements. Used SQL Server SSIS tool to build high performance data integration solutions including extraction, transformation and load packages for data warehousing. Development of C modules for activation, deactivation and modification of these plans in Clarify Front End. Created SQL*Loader scripts to load data into the temporary staging tables. Designing Tables, Constraints, Views, and Indexes

etc. in coordination with the application development team.    Used TOAD, PL/SQL developer tools for faster application design and developments    Developed procedures using Dynamic SQL. Developed database objects including tables, Indexes, views, sequences, packages, triggers and procedures to troubleshoot any database problems    Tuned complex Stored Procedures for faster execution and Developed database structures, according to the requirements    Creating table spaces, tables, views, scripts for automatic operations of the database activities.    Environment: C, C++, UNIX Shell Scripting, Forms, PL/SQL, Oracle 10g, 11g. Informatica-8.6, SQL, PLSQL, Toad.

Java/ J2EE Developer Aeverie - Lincolnshire, IL July 2014 to August 2015 Description: It provides various financial software services which includes treasury, investment management and enterprise handling. The software system is developed to provide solutions to the corporate banking present in the market. It implemented the functionalities of payments, treasury services and cash management.

   Responsibilities:    The application and deployed on Tomcat Server.    Actively involved in code reviews and in bug fixing.    Environment: Core Java Developed the application using agile methodology and Scrum method of project management.    Involved in group meetings and made substantial changes to the design to improve performance of the Application.    Responsible for front-end UI design using HTML/HTML5, CSS/CSS3, JavaScript, jQuery, etc., taking advantage of the MVC pattern of the Spring MVC to produce higher maintainability of the code.    Developed UI screens using JSP, JQuery, JavaScript, XHTML, and CSS.    Created Server Side of application for project management using Node.js and Mongo DB.    Developed and deployed Enterprise Web Services (SOAP and RESTFUL) and generated client using Jersey and Axis Frameworks using Eclipse.    Extensively used Core Java concepts like Multithreading, Collections Framework, File I/o and concurrency.    Worked on design patterns like delegate, service layer and various internal design frameworks -links, notification and audit frameworks.    Developed and executed unit test cases using JUnit, and Mockito as mocking framework for mocking data.    Used GIT for version control and JIRA issue tracker to file the bugs. Used MAVEN for building, J2EE 1.6, Spring Framework, Bootstrap, HTML5, Java script, jQuery, CSS, Node.js, Mockito, Apache Tomcat 7, Eclipse, XML, Maven, Log4j, REST API, Hibernate, Oracle, JUnit, GIT, JIRA, UML and Apache

AXIS. JAVA/J2EE Developer Freedom Bank, VA June 2013 to July 2014 Description: Freedom Bank provides funds or loans to people with small business requirements. Applicant get their loans sanctioned based on their credit history. The applicant information is maintained in a database along with the details of the loan for repayment. This data is filtered into different categories based on parameters like type of account, loan amount, due date. The filtered data is used for statistics for generating reports. Responsibilities: Responsibilities include Use case modeling, Object modeling using Rose, and ER Database design. Model View Controller (MVC) architecture has been adopted to provide framework. Utilized UML & Rational Rose suite for designing of the System. Followed DAO Pattern and J2EE framework facilitated the integration & deployment of DAO, Continuous Integration, Servlets, JSP and XML. Developed Session Beans to encapsulate the business logic and Entity beans as the persistence objects. Developed EJB-Session Beans that implements the business logic. Used IBM DB2 as Database. Implemented the JMS API's with message priority levels and listener timeout's. Specified, prototyped, developed and tested an object-oriented, multiplatform C++ framework containing support to: data structures, common algorithms sockets, threading. Created Web Services using SOAP, WSDL to provide services to other systems within the company. Enhanced the application for multi-threaded and Polymorphism scenarios. Used Maven and Hudson as build tool and deploying the application. Deployed the application under Web Sphere Server. Resolved the production issues during migration onto the production server. Environment: RUP, Rational Rose XDE, Java, J2EE, Struts 1.1, IBM DB2, Multi-threading, Unix, XML, XSLT, ANT, JDBC, JMS, Eclipse, Visual Source Safe, WSAD 5.1/5.0, Selenium, Apache Cloud Stack, Tomcat Application Server, Web Sphere Application Server 5.1/5.0, SOAP, WSDL. Java Developer PMCS Services Inc - Austin, TX May 2010 to June 2013 Description: The Objective of new system is to enable PMCS Services to strengthen their existing client and attract new client. The PBG System will enable the PMCS Services which will reflect each client s unique relationship with PBG. The functional requirements of the PBG Internet site is to deliver account information, news, marketing information PMCS Services. Responsibilities: Implemented Struts framework based on the Model View Controller (MVC) design paradigm.

Designed the application by implementing Struts based on MVC Architecture, simple Java Beans as a Model, JSP UI Components as View and Action Servlet as a Controller    Used JDBC for data access from Oracle tables.    Used Design Patterns like Singleton, Factory, Session Facade, Service Locator, and Data Transfer Object.    Implemented EJB's Container Managed Persistent strategy.    Apache Ant was used for the entire build process.    JUnit was used to implement test cases for beans.    Participated in the production support and maintenance of the application and System Procedures on a UNIX environment.    Environment: HTML, CSS, JavaScript, JSP, Servlets, Struts1.2, JMS, Unix, JavaScript, Eclipse, Web Sphere Application Server, Oracle, EJB, ANT. Education Bachelor's Skills ECLIPSE (5 years), JAVA (6 years), JSP (5 years), ORACLE (6 years), UNIX (7 years) Certifications/Licenses Driver's License Additional Information Technical Skills:  Big Data  Hadoop, HDFS, Spark, Hive, Impala, Kafka, Hue, Map Reduce, YARN, Pig, Sqoop, HBase, Couch base, Cassandra, Oozie, Storm, Flume, Talend, AWS, Horton works and Cloudera clusters Language Scala, Java, C, UNIX Shell Scripting, AngularJS, PL/SQL, Python  Java/J2EE J2EE, JSF, EJB, HTML, XHTML, AngularJS, Servlets, JSP, CSS, XML, Ajax, Java script, SOAP, Restful  Open source framework and web development Struts, Spring, Hibernate, JavaScript, AJAX, Dojo, J Query, Ehcache, Log4j, Ant, JBoss, Web services, SOA, SOAP, REST, WSDL and UDDI Portals/Application servers Web Logic, Web Sphere Application server, Web Sphere Portal server, JBOSS  Operating system Windows, AIX, UNIX, Linux. ETL Tools  Ab Initio GDE Version 3.0.2.2, Co>Operating system 3.1.6.1, EME, Data Profiler, Familiarity with AbInitio ACE Application Configuration Environment and BRE Business Rule Engine.    Configuration Mgmt. CMVC, Clear Case, Clear quest, PVCS, CVS, Nagios, Puppet, Ansible.  Development Tools Eclipse, Visual Studio, Net Beans, Rational Application Developer, WSAD, JUnit.  Databases Couch base, Cassandra, HBase, Oracle 10g, MySQL, Teradata SQL.  Software Engineering UML 2.0, Rational Rose, Design Patterns (MVC, DAO etc.).

Name: Alyssa White

Email: patrickknapp@example.com

Phone: 4363181210