Senior Big Data Engineer Senior Big Data Engineer Senior Big Data Engineer - ERP ANALYSTS INC * Having 12+ years of experience in all phases of Software Development Life Cycle with around 4 years of experience on Biga data Analytics with hands on experience in implementing solutions. * Worked in USA, Japan, Canada, APAC region and India developing micro services, application software, firmware, device drivers and android applications. * Excellent design and programming skills as developer which includes C, C++, Java, Scala and conceptual knowledge on reporting tools. * Proficient working experience in Hadoop Architecture and Cloudera ecosystems such as HDFS, Spark, Spark SQL, Spark Streaming, Kafka, MapReduce, Pig, Hive, HBase, Oozie, Sqoop, Flume, Zookeeper. * Experienced in using Spark and Spark SQL API's in Java and Scala. * AWS services used S3, EC2, Glue, Athena, RDS, Dynamo DB. * Databases used in integration MySQL, DB2, SQL Server, PostgreSQL, Oracle, MongoDB. * Developed services that provide Migration, Upgrade and Data Conversion for Legacy Data Warehouse and transactional systems to new data platforms. * Perform Functional Validations and Quality Assurance on the data in Semantic Data Mart using Big Data tools such as HDFS, Spark SQL, Hive, Impala to ensure the data meets the stated requirements and business transformation rules through out the data pipeline. * In depth understanding/knowledge of Hadoop Architecture, YARN and various components. * Responsibly performed roles such as gathering requirements, analysis, design, coding, review and testing for achieving the functional specifications. * Familiar with Agile, Waterfall and Scrum methodologies. * Experience in application development using Java, J2EE, JSP, HTML, Java Script, PHP and Web services * Designed new initiatives and showcased PoC's for management approval. Developed Android applications for the business needs. Experience in understanding the requirements in all dimensions and providing customized solutions according to the customer unique demands. * Professional experience in Technical Consulting, Software Engineering, Embedded Linux Systems development, Project Management, Defect and Issue Management. * Firmware experience using C language specific to different printer models and emulations. * Experience in developing device drivers and filters in all major flavours of Linux and knowledge on CUPS, OPOS and JavaPOS. knowledge of internal jar architecture and JavaPOS and integration with systems. * Developed

Linux PCI device drivers on worked RTOS Kernel in Red Hat Linux. Developed a front end GUI application in Linux platform using Qt designer and interfaced the application with PCI boards. * Team player with proven track record of technical leadership on development projects. Worked in culturally diverse environments in various countries including India, Singapore, Japan, South Korea, Canada, United States and other South-East APAC countries. * Visited client/customer place for On-Site installation of the product, for testing and deploying. Attended the meetings with the client and was primarily responsible for reaching the project deadlines. * Strong work ethic combined with a commitment to excel in projects undertaken, equipped with excellent analytical & logical skills. * Good verbal and written English communication skills, strong interpersonal skills, outspoken and self- motivated individual. Work Experience Senior Big Data Engineer ERP ANALYSTS INC - Solon, OH February 2019 to Present The purpose of this project is to provide the executives, business unit managers, and department heads with dynamic analytical data to support decision making on a real-time basis. As a start-up company, iDesign has been challenged with limited resources for managing its voluminous data generated as a result of its success from its ecommerce businesses. As a result, the business often struggles to keep up with inventory levels, vendor management, social media analytics, and ad spend effectiveness among other key business and operational challenges. MDP will provide the data from disparate source systems into data lake and give the provisioning to plan, design, build, and implement a new business analytics platform of the future for iDesign. By implementing a real-time BI capability, mDesign can better leverage its limited resources by establishing business priorities, allocating resources in a more efficient manner, and providing proactive management oversight to help in identifying new market opportunities, better vendor management, more accurate inventory management, and launch new products before the competition. Responsibilities: Gathering requirements to analyze high level business needs, system specifications from business partners and subject matter experts.   Construct the requirements into low-level specifications and prepare system design plan, functional documents for the development by understanding the intricacies.   Consults with higher-ups to validate complex design decisions. Provide advanced coding expertise to mitigate high risk failures or technical

challenges. Integrate the Registration framework and Data Curation services by enabling Profiling, Data Quality and Standardization of customer data prior to Ingesting data into Data Lake or reservoirs. Identify the KPI's according to business use cases. Provide the curated data for BI teams to plan, design, build, and implement a set of Key Performance Indicators on a BI analytics platform; using Tableau as the presentation layer and Azure as the data repository. Environment: Azure, Java, Scala, Spark, Hive, Linux Hadoop Developer ERP ANALYSTS INC - Bloomington, IL October 2018 to January 2019 Property Insurance provides protection against most risks to property, such as fire, theft and weather damages. This includes specialized forms of insurance such as fire insurance, flood insurance, earthquake insurance, home insurance, or boiler insurance. Scope of the Property Data migration is to build a system where users can perform better analysis and view of Property Data, so as to understand business and to take Better Decisions on Business related covering Commercial and Personal property. Using the framework data is copied from Legacy Systems to Hadoop. Property Data Migration project replaces the existing DRS reports with Extracts provided on Hadoop. Property Data Extracts are delivered in three phases, that are being brought from disparate source systems. Phase-1 has 65 attributes of Home QA Pivot, Home Retention, Rental Dwelling, Property Code Description. Phase-2 has 82 attributes of Farm Policy Counts, Farm Retention, Home and Farm exposure, Policy Level, Item Level. Phase-3 has 183 attributes of New Business Characteristics of Home and Farm, Days to Issue. Responsibilities: Used Zena Scheduler as work load management tool to SFTP the data from mainframe server to Hadoop cluster. Copied the files from Hadoop to HDFS Landing Zone using the framework that is built on Java, Pig UDF's and shell scripts. Using the framework, perform Data Cleansing by removing the non-printable characters, Standardizing data, Trim data to maintain the uniformity across all tables and moved the data into Core Zone with a partitioning as snapshot_year_month and snapshot_day of the loads. Copied the Reference tables (Liability Code Construction, Code Contents, County Dwelling Decodes, Multi Policy Decodes, Protective Decode, Surcharge Devices, Policy Status Decodes) as one time loads, Property Master tables (Property Basic, Property Description, Foot Note, Lien, Location, Premium Stats) as daily files, Straight Through Process file,

SIEBEL tables and CBR History data into Core Zone.   Load Core Zone data into proposed conceptual data model to enable the business users to view the   Property Data at various granularities in Curated Zone. All the Property Data is stored in ORC format and Extracts are created in daily, monthly, new business and in category (Home and Farm)   Build extracts as the business needs in different phases and provided these extract tables to model on Business Object Universe.   Environment: Hadoop, Sqoop, Hive, Linux, Java, Zena Scheduler, GIT lab, Hortonworks

Senior Big Data Engineer ERP ANALYSTS INC - Kansas City, MO July 2016 to September 2018 Migrate all the data to data-platform / data-lake replacing legacy sources. This involves decommissioning data migration jobs on mainframe. Migrate RCI system and RSI applications as-is to enterprise data platform and keep provisioning data. Migration will include - the sources coming to AAG platform from power select, TA, SA, sales connect and Argus. All these sources come from processes running on mainframe. This migration would result in 30% FSG data availability on the Data Platform which will enable various capabilities to function such as building consolidated metadata repository, improving data quality by providing data quality scorecard to Data stewards for corrections, Data Management activities, archival retrieval and curated data availability for data discovery among others. Core Services are a one stop platform for all of the key needs that are usually performed on humungous data. Enterprise needs a disparate system such as archiving and retrieving data on-demand, hot and cold data management, moving any format of data from anywhere to anywhere. All these features are available as a service at a click of a button. It also addresses the limitations of various enterprise software. These services also provide data provisioning and data ingesting services as and when needed. It has dashboard capabilities for operational reporting and monitoring purposes. Key services include Control Validation, Data Quality, Delta detection, SCD handlers, SK generation, extraction, conversion to different file formats for the entity registered. Responsibilities: Construct the requirements into low-level specifications and prepare system design plan, functional documents for the development by understanding the intricacies.   Designed and Developed services and spark applications that provide Migration, Upgrade and Data Conversion for Legacy Data Warehouse and transactional

systems to new data platforms.   Register technical metadata for RCI and Power select projects, define configuration metadata for executing Delta, Control Validation, data quality and standardization process for ingesting data into Data lake using services built on Java, Scala utilizing spark framework for each data source.   Dynamic allocation of resources on top of YARN and Spark for various projects like RCI and Power Select.   Generate fact data for positions and transactions for RCI's Fund Trade Monitoring (FTM).   Build Semantic processes to load data to Data Marts, build alerts with reference data for RCI team to run compliance tests.   Define preprocessing rules, tasks, jobs, dependencies, build plan, schedule aligning to SLA for Risk Compliance Intelligence project.   Implement SCD of type-2 using Hive, HBase for semantic delta process in Power select.   Outbound Extracts for converting the output files to different formats irrespective of single source or multiple sources, combining sources, splitting files and partitioning.   Build semantic files with all the business rules and send downstream extracts for power select.   Developed Provisioning Framework for Data Discovery to provide a uniform query engine on disparate sources for seamless user experience to provide insight and analytics.   Build Comprehensive Data Life Cycle - Archive cold data as per retention requirements and retrieval of archived data as per SLA.   Use Kafka service for logging the data from different services. Develop a data pipeline using Kafka to store data into HDFS. Created Kafka Topics and distributed to different consumer applications. Design and code data ingestion, audit/logging, security and data provisioning frameworks using Java, Scala, Python and Shell scripting.   Develop high quality code and unit test all microservices in support of Data Platform Buildout.   Test Data management tools - Provide test Data Generation capability using Java for carrying out Unit Testing, QA, Performance and Regression Testing. Implement CI/CD using GIT, Jenkins, Groovy. Code migration and deployment to different platforms. Automatic Scheduled code builds through Jenkins and deployment to respective platforms. Worked on ingestion of data from various database systems by loading tables into HDFS using Sqoop and export the results back to Databases. Various Databases integrated are Oracle, MySQL, and DB2 to HDFS and vice versa.   Registered an instance as an object of database, file, HDFS as a single point of repository. Registration of sources lets anyone to link any source to any target.

Developed shell scripts that run in background to automate the ingestion and deploying the tables for both snapshots and deltas.   Worked with different File Formats like text file, csv, parquet, Json for querying and processing.   Reconciliation of data as a validation process after the data moving, data encryption at field level, data federation even though the source is residing on disparate systems and data purging as and when data is not needed. Environment: Hadoop, Spark, Sqoop, Hive, HBase, Kafka, Eclipse, Linux, Java, Scala, MySQL, Oracle DB, GIT, Jenkins, Spark SQL, Impala, Cloudera   Project: MDM Catalog Gathi's Master Data Management (MDM) catalog is a framework of processes for creating and maintaining a reliable, accurate, and secure data environment, to consolidate and standardize business process. Gathi framework includes Meta Data Registration, Micro services, Curation services, Processing services, Job Orchestrations, Meta data dashboard. Extraction of data from any database and build the ingestion job on registered object. Data quality, Statistical profiling, Data standardization can be executed as a service or during meta data capture. Each step logs job status, warnings, errors, record read/write count into Audit tables. Logging is done on real-time. Process is aborted when there is a breach in threshold. Failure in any step triggers Alerts to the operations team to take action. Dependencies on job/SLA/execution are controlled via Job Orchestration framework. Responsibilities: Developed Registration and Data Curation services enable Profiling, Data Quality and Standardization of customer data prior to Ingesting data into Data Lake or reservoirs.   Key features include: Sandardization to ensure File and Data level consistency, Maximizing usability of data, Optimal Storage Space Utilization, Cold data identification and transfer into archival zone.   Using PostgreSQL database engine on AWS RDS for metadata registration.   Integrated the product with AWS services to demonstrate product with cloud capabilities. Using core services data is moved to S3 from SQL server systems. Data is moved to the bucket as parquet files, which stores the data in compressed columnar format and which can be used by any JDBC enabled service.   Using crawler feature in Glue, configured those data stores to run queries. As all the data is available in S3, exposed these table to users.   Schedulers are configured in crawler, once the data is available in Glue, through Athena provided the capability to business users or data analysts to execute the

queries. Batch optimization service ensures timely completion of batches within the acceptable SLA. Key features include: Optimizing existing batch schedules, Migrate long running jobs from legacy environments to Big Data Built Data Virtualization service that delivers a unified and integrated view of data, as required, from different source systems. Key features include: Federated data access across hetro/homogeneous sources, Single view of Enterprise data, Single API to access data, Easy to Augment Discovery Process Build Provisioning Framework for Data Discovery to provide a uniform query engine on disparate sources for seamless user experience to provide insight and analytics. Created Test Data management services - Provide test Data Generation capability using UI for carrying out Unit Testing, QA, Performance and Regression Testing. Environment: Hadoop, Spark, Sqoop, Hive, HBase, Kafka, Eclipse, Linux, Java, Scala, MySQL, PostgreSQL, SQL Server, GIT, Jenkins, Spark SQL, Impala, Cloudera, AWS (EC2, S3, RDS, Glue, Athena) Java Developer EPSON SINGAPORE PTE LTD - Singapore January 2015 to June 2016 Epson new intelligent printers are in-built with Linux/Windows OS. Application systems where the transactions occur also has transaction data along with operational primary data stored in RDBMS systems. Key objective was to develop ETL processes and move the data for analysis, both by vendor itself and operational businesses. Responsibilities: Gathered the business requirements from the Business partners and subject Matter Experts. Installed and configured Hadoop clusters for application development and Hadoop tools. Responsible for building scalable distributed data solution using Hadoop Developed ETL processes to load data into HDFS using Sqoop and export the results back to RDBMS. Data and transaction histories into HDFS for further analysis. Worked on custom Pig Loaders and storage classes to work with a variety of data formats such as JSON and XML file formats. Developed multiple Map Reduce jobs for data cleaning. Wrote the shell scripts to monitor the health check of Hadoop daemon services and respond accordingly to any warning or failure conditions. Used Flume to collect, aggregate and store the application log data from different sources like web servers, mobile and network devices and pushed to HDFS. Developed Pig Latin scripts to extract the data from the web server output files to load into HDFS Handled importing of data from various data sources, performed

transformations using Hive, MapReduce, and loaded data into HDFS.   Created hive tables, loading data and write hive queries that will run internally in a map reduce way.   Deep understanding of schedulers, workload management, availability, scalability and distributed data platforms.   Installed and configured Hadoop MapReduce, HDFS, developed multiple MapReduce jobs in Java for data cleaning and pre-processing.   Involved in managing and reviewing Hadoop log files.   Involved in running Hadoop streaming jobs to process terabytes of text data.   Developed HIVE queries for analysts.   Used Sqoop to extract and ingest data into RDBMS. Marketing teams used this data for visualization and for analyzing market situation, customer usage trends and needs of various vendors for estimating and predicting user behavior. Environment: Hadoop, Hive, Linux, HDFS, Java, SQLite, MySQL, Oracle DB, Cloudera   Project: Claim Management System Java Developer The CMS System March 2014 to December 2014 constitutes of a module called Purchase Requisition. To create vendors by evaluating vendor details and maintenance information. Vendor evaluation consists of evaluating the quality, cost, delivery and rank details of the vendor and vendor maintenance information consists of the agreement terms, policy terms and the annual maintenance details. Web services are called out to get the evaluation and maintenance details of the selected vendor. As details are evaluated a decision can be made to create the vendor if one does not already exist. Upon approval of the PR, user will send the PO to the vendor. Responsibilities: Involved in gathering system requirements for the application and worked with the business team to review the requirement and prepared SRS   Developed the application using Spring Framework following Model View Controller (MVC) architecture with JSP as the view. Developed presentation layer using JSP, HTML and CSS and JQuery.   Developed JSP custom tags for front end. Written Java script code for Input Validation.   Extensively used Spring IOC for Dependency Injection.   Developed J2EE components on Eclipse IDE. Used Restful web services with JSON.   Developed EJBs to create the new vendor with the details provided   Analysis and reproducing the defects reported and fixed them.   Documented the root cause analysis and proposed the fixes for the defects.   Participated in support and maintenance meetings of the project, providing weekly updates on the defects to the higher management.   Understanding

various functional flows involved in enhancing the application to the needs of the users Environment: Core Java, JSP, Servlets, JSON, Java Script, Web Services, spring, Oracle 11g, Ant, Maven, Web logic, Windows, Linux, and Eclipses IDE Java/Android Developer EPSON SINGAPORE PTE LTD - Singapore January 2013 to February 2014 In retail and F&B, POS systems are significantly important to the owner as well to the end-users. To penetrate Point of Sales (POS) market with Epson new intelligent printers along with wide range of android tablets, Epson Singapore has come up with a pro-active approach to develop in-house android application to showcase the in-built features of intelligent hardware. Epson intelligent printers are the world's first POS hardware that can understand XML through the ePOS services. Responsibilities: Designed wireframes and mockups to showcase the functionality along with look and feel appearance. Involved in developing the UI of application and integrated with Epson Printer. Functions of application is to accommodate the restaurant needs in F&B POS App and retailers needs in Retail POS App End-user as well can operate from his own smart device and do the ordering. This will leverage the users with more comfort. Involved in all phases of the project, from designing the specification to launch the app. Consumed the JSON web services and parsed the data using the JSON parser to save the data in SQLite Used MAVEN, GRADLE build techniques to integrate the third party jars as required Worked with most of the Android UI components like List View, Grid View, View pager, Adapters, etc Strived for elegance and simplicity in code while focusing on scalability, readability and standards complicity Integrated the app to work with HID devices such as barcode scanner, magnetic stripe reader, customer display using Epson ePOS SDK APIs Environment: Android, Core Java, XML, Java Script, JSON, Web services, SQLite, ePOS API SDK, Apache Tomcat server, XAMP, Eclipse, Android Studio, github Senior Engineer EPSON SINGAPORE PTE LTD - Singapore November 2011 to December 2012 Nov 2011 - Dec 2012) Zebra Programming Language (ZPL) is a powerful label-definition and printer control language. Labels are defined in ZPL and generated from a host computer system by using commercial label preparation system or any software package like bartender, code soft or nice label. ZPL support in Epson printers creates an opportunity to penetrate Epson Label printers in new market. ZPL

component is responsible for rasterization of the ZPLII commands, responding as a ZPL device, and the implementation of command additions to expose the full color functionality of Epson Printers. Responsibilities: Lead the team of 5 engineers in Singapore and India, excluding myself.   Involved in enhancing the design of emulator for Epson Hardware along with R&D teams   PIC for updating the project information to all parties and top level management in an organization. Understood the Linear & 2D Barcode Specifications and designed schema. Specifications of barcode specifications that commonly used UPC, EAN, Code 39, Code 128, ITF (2 of 5), Code 93, Coda Bar, GS1 Data Bar, MSI Plessey, QR, Data matrix, PDF417, and Aztec barcodes. Enhancements and Developments are done in Code::Blocks IDE in Linux using C++ and Customized Open Source Zint Library in C.   Developed ESC/Label as a uniform programming language for Epson Label printers to integrate virtually with any OS. Environment: C++, Java, Code::Blocks IDE, Open Source ZINT Library, ZPL Specs, SVN Senior Engineer EPSON SINGAPORE PTE LTD - Singapore August 2008 to October 2011 Aug 2008 - Oct 2011) Seiko Epson handlers are manufactured by the factory automation division the Japan-based Seiko Epson Corp. a leading manufacturer of robotic, electronic and imaging products. The company produces a line of well-regarded IC handlers that incorporate the company's robotics technology. Seiko Epson handlers are used for the inspection process by semiconductor manufacturers worldwide. The company produces a wide range of handlers and come equipped with a variety of options suited to most any IC test requirements. Seiko Epson handlers are known for reliability, high precision and low maintenance. Responsibilities: Responsible for coordinating along with the sales team in gathering the requirement, preparing technical specification and quotation for creating new IC handler software.   Customize an Installer Software with HMI and SPEL to improve the performance tests and providing the enhanced HMI to the users.   The HMI (Human Machine Interface) sends and receives the data between the operator and the Handler system. Operator enters the device data into the HMI. The Handler system feeds back the category data received from tester, Temperature Monitor display, and operation status of the Handler to the operator.   HMI is developed using VB and the code or data that machine need to understand is done by using VC++

and SPEL language. Environment: VC++, Visual Studio, Windows 2000, Windows ME, SVN Senior Firmware Engineer EPSON SINGAPORE PTE LTD - Singapore August 2008 to October 2011 Singapore (Reporting to EPSON Japan, R&D Team) Project: Linux Device Drivers and Firmware Development for Printers Role: Senior Firmware Engineer (Aug 2008 - Oct 2011) These developments are done to support the unique requirements received from customer accounts that Epson has in Asia region. Responsibilities: Pioneer in South East Asia region for supporting Linux filters and CUPS drivers in all major Linux distributions for wide range of dot matrix and POS printers. Initiated Epson-Red Hat meet to increase Linux end-users in emerging countries. Developed Linux filters and CUPS drivers for 9-pin and 24-pin printers. Improved the quality of 9-pin printer driver by implementing half-dot method for dot-matrix printers. Provided solutions and on-site support for integrating Epson hardware in Linux environment Developed Epson's first 6-inch customized printer for Indian market. Supported ESCPOS emulation for Dot Matrix Printer for the first time in Epson. Contributed ISCII support for Indian Retail market and Thai 3-pass & 1-pass language support for Epson POS printers. Worked with Epson Malaysia presales team in developing DEC emulation support for the first time on Epson Passbook Printer. Environment: VIM, Embedded C, Motorola Processor, Greenhills compiler, Linux (RedHat, CentOS, SuSE, Ubuntu, Debian), GCC, Clearcase Software Engineer Park Controls and Communications Ltd - Bengaluru, Karnataka August 2006 to July 2008 Park Controls and Communications Ltd, Bangalore, India Project: Antenna Controller Unit Client: ITR - DRDO (Defense Research Development Organization), Chandipur, India Role: Software Engineer (Aug 2006 - Jul 2008) Antenna Controller unit is a stand alone embedded system having a single board PC with AMD Geode Processor suitable for Real time applications. The ACU board is a PCI bus compatible card that plugs into the PCI slot of Processor board. Antenna Controller Unit provides complete control of a pedestal, antenna and tracking receiver(s) in a single chassis. ACU can drive signals for the tracking modulator in a single channel mono pulse tracking feed, receive Automatic Gain Control (AGC) and tracking video from tracking data receivers, provide controls for manual or automatic mode selection such as automatic target acquisition and polarization or frequency diversity selections, and

provide controls for operation of a two-axis pedestal with position and status readouts. Responsibilities: Designed and developed Device Driver in Linux (on 2.6.16 kernel) for Acquisition card Application Integration with elo Touch Screen Interface Developed a GUI Application for processing of real time data, graphical display of processed data and I/O operations to the card Developed a Client / Server application to broadcast the acquired data in real time through LAN and data logging Involved in Design, Coding, Integrating and Testing Environment: C, C++, Qt Designer, RTOS, Linux, AMD Geode LXDB800 Board Education Bachelor of Technology in Computer Science Engineering in Computer Science Engineering JNTU University - Hyderabad, Telangana 2006 Big Data Engineering BITS Pilani Skills Apache, Linux, Shell scripting, Unix, Eclipse, Java, Visual studio, Android studio, C++, Device driver, Hadoop, Hbase, Hdfs, Hive, Html, Javascript, Mapreduce, Php, Pig, Python Additional Information TECHNICAL SKILLS: Big-Data Ecosystem: Apache Hadoop 2.x, Spark, Spark SQL, Scala, HDFS, MapReduce, Pig, Hive, Sqoop, Flume, Oozie, Zoo Keeper, Apache Kafka AWS Services: S3, EC2, Glue, Athena, RDS, DynamoDB Programming: C, C++, Java 8, SPEL Web Development: HTML, Java Script Web Services: XML Scripting: Shell scripting, Python, JavaScript, PHP Databases: MySQL, DB2, SQLite, Oracle, SQL Server, PostgreSQL, MongoDB No-SQL Databases: Cassandra, HBase IDE's: Android Studio, Eclipse, Code Blocks, Qt Designer4, Source Insight, Visual Studio Servers: Apache Tomcat, XAMP, Glassfish Reporting Tool: Tableau Others: Firmware & Linux Device Driver for various embedded products Version Control: SVN, IBM Clear case (Windows), Github Operating Systems: Linux (RedHat, Ubuntu, SuSE, Debian, CentOS), Windows, RTOS, UNIX Domain Experience: Big-Data, POS Applications, Banking Applications, Embedded Products, Software Development, Networking, Factory Automation

Name: Andrew Harris

Email: cummingscameron@example.com

Phone: 368.358.5498