

Sr. Big Data Architect Sr. Big Data Architect Sr. Big Data Architect Washington, DC Currently working as a Sr. Big Data Solutions Architect, I have over 17 years of experience in the architecture, design, construction, and implementation of Information system. Worked as Information Architect, Data Architect, Integration Architect, Database Architect, Data Services Specialist, Data migration Specialist, Data Modeler, DBA. Extensive cross-industry domain experience, mainly, Financial Services, Retail, Insurance, Manufacturing, Utilities. Extensive knowledge in BI, Big Data tools, Data Modeling, ETL, noSQL, J2EE, EAI, SOA, SOI, BI, Web Services, BPEL, JSON, XML, XSD, WS-*. Authorized to work in the US for any employer Work Experience Sr. Big Data Architect Washington, DC September 2016 to Present Developed Scala scripts, UDF's using both Data frames/SQL and RDD/MapReduce in Spark 2.0.0 for Data Aggregation, queries and writing data back into RDBMS through Sqoop. Developed Spark code using Scala and Spark-SQL/Streaming for faster processing of data Developed Oozie 3.1.0 workflow jobs to execute hive 2.0.0, sqoop 1.4.6 and map-reduce actions. Design big data authentication solution using LDAP/Kerberos and Authorization using UNIX groups and HDFS ACLs. Architect the Hadoop data security using DEZ, Data Encryption Zone, and control data access using UNIX groups and HDFS ACLs. Architect/design the Integration solution using Oozie coordinators and Control M. Worked with the technology manager and business stake holders to demonstrate the strategic value of the data lake platform. Lead the platform and data migration from Big Insight 4.1 to Big Insight 4.2. Designed code generation framework using UNIX shell and python to automate the Hadoop code artifacts (BigSQL, hive, HBase, Oozie coordinator, Oozie workflow). Designed data analysis and visualization using BigSQL DSM and IBM Big Sheet. Worked with the IBM Bluemix support to solve the platform issue and apply required patches. Technology Used: IBM BigInsight 4.1/4.2, IBM InfoSphere Datastage, Python, shell script, Hive, HBase, BigSQL, Spark, Pig, BigSheet, IBM DSM, Control-M, Oozie, DB2, OS/360, Linux Sr. Data Architect Veracity Englewood C November 2015 to August 2016 Developing predictive analytic using Apache Spark Scala APIs. Implement enterprise grade platform (mark logic) for ETL from mainframe to NoSQL (cassandra) Assigned name to each of the columns using case class option in Scala. Experienced with batch

processing of data sources using Apache Spark. Implemented Spark GraphX application to analyze guest behavior for data science segments.. Enhancements to traditional data warehouse based on STAR schema, update data models, perform Data Analytics and Reporting using Tableau Architect and Design and code Python/BeautifulSoup to web-scrap material, equipment, labor rate and construction cost data from public websites(like HomeDepot, yellowpage, wdol.gov etc)

Design and Build python/scoop script to exchange data from Hive warehouse to Elasticsearch to service API layer. Architect ELK (elasticsearch, LogStash/FileBeat, Kibana) to aggregate application data and web and application log files for 360-degree view of the customer insight analysis. Technology used: Apache Pig, Apache Hive, Apache SQOOP Microsoft SSIS, SQL Server 2014, Python, Elasticsearch, RabbitMQ, Logstash forwarder, Filebeat, Kibana, Power BI, HDInsight, SSDT, pyodbc

Sr. Enterprise Data and Big Data Architect Belk, Inc October 2013 to October 2015 Develop and design bigdata reference architecture and supporting technology stacks. Provided hardware architectural guidance, planning and estimating cluster capacity, and created roadmap for Hadoop cluster deployment. Provisioned, installed, configured, monitored, and maintained Hadoop components, mainly, HDFS, Pig, Hive, Sqoop, Solr, Hue and Oozie. Done recovery from node failures and troubleshooting common Hadoop cluster issues. Patched and upgraded Cloudera cluster from CDH4 to CDH 5. Provided design and technical assistance with challenging issues that Big Data Integration spans multiple systems beyond Hadoop. Supported Hadoop developers and assisting in optimization of map reduce jobs, Hive Scripts, and other data ingest and extract activities. Defined and managed continual service improvement to Hadoop platform and surrounds Installed rHadoop packages in the Hadoop cluster to build the environment for the data science group to do prescriptive and descriptive modelling. Worked with implementation vendor team to Architect, design and lead the following Innovation projects web/akamai/app server logs analysis and operational dashboard using solrCloud/Hunk/Tableau. Offload batch jobs from mainframe to Hadoop environment. Build the hive/impala schema for the hourly sales transaction to support the EDW reports. data wrangle and data prep from the various data sources for tableau extract. build system to efficiently archive and retrieve large image files.

migrate the oracle data to MongoDB Architect, design, and implement db2 9.7 and oracle11g RAC database system involving SAS, Oracle Applications, MDM software. Build the architecture and technology stack for the Enterprise Information data governance. Architect, design and Integrate the Enterprise Business Reporting tool and environment. Design and implement Integration/Framework data hub layer Data model. Support and augment Oracle Apps/SAS database system. Application/Infrastructure planning and job design and performance tuning for the Data Integration using IBM InfoSphere suite of tools. Design and implement product MDM using Info sphere Collaboration server 9.x Design and model data integration and replication using IBM ISS CDC and Oracle GoldenGate. Technology Used: Cloudera CDH 5.1.3, Hortonworks HDP 2.0, Syncsort, MongoDB, Python, shell script, Java Map reduce, Hive, Impala, pig, scoop, rHadoop, solrCloud, Tableau, Splunk/Hunk, Python, Oracle 11g RAC, DB2 10.5, IIS 11.3, Autosys, Oracle BI, AIX, J2EE, IBM IIS 11.3, IBM InfoSphere Information Governance Catalog, IBM IIS CDC, Oracle GoldenGate Integration/Data Architect BJ's Wholesale Club June 2011 to October 2013 Technology Used: IBM ACE, Erwin 4.5, IBM WebSphere Message Broker 7.0.1, IBM WebSphere 7.0, JDK 1.6, IBM IS DataStage, Oracle 10.x, J2EE, JSF, ARTS, Build POS solution Architecture using IBM ACE and IBM Middleware technologies. Designed Enterprise Integration Layer (IL) development to convert the proprietary data format (TLOG) to the retail industry standard (PosLog) XML format. Build Data model of the Operation Data Store (ODS) using ARTS modeling standards to build a central sales transaction repository. Designed the ETL layer to interface with the legacy back-end accounting and cash settlement systems. Solution and design the web application for using JSF framework to store and manage store attribute data. Information Architect IBM December 2000 to October 2011 Information Architect, Sr. Data Modeler, Team lead WellPoint July 2009 to May 2011 Technology Used: Optim, Erwin 4.5, IBM WebSphere 6.0, JDK 1.6, JIBX, CXF, Hibernate 3.2, Tibco EMS, Qarbon's ViewletBuilder, ClearQuest, DB2 UDB Design and Develop Test Data Management strategy domain areas including processes, approaches, procedures, and tools and environment. Build Test Data Catalogue Solution Architecture. Design the data model for the Test Data Catalogue. Architect, Design, and Lead the Test Data Request application using IBM

ClearQuest and DB2. Develop Test Data Catalog viewlets for the training and knowledge transfer to the test data managers and leads. Sr. Data Modeler, Information Architect, Team Lead, Data Analyst Citi Group August 2008 to June 2009 Technology Used: Oracle RAC 9i, Erwin 4.5, AbInitio 1.15.6, IBM WebSphere 6.0, JDK 1.6, JIBX, CXF, Hibernate 3.2, Tibco EMS. Architect and design the data provisioning hub from CAS to the end system using SOAP Request-Response MEP using Tibco/Hibernate. Architect, design, and implement the complex data archiving and purging solution. Design the Payment Object Model and Data Persistence layer using J2EE/Hibernate/CXF/JIBX. Architecture, design, administration and performance tuning of Oracle 9.x RAC (three node) Status tracking, reporting and problem resolution, Root cause analysis of issues, issue escalation and resolution Information Architect, Team Lead, Data Analyst, Data Modeler Circuit City August 2007 to August 2008 Technology Used: Db2 LUW, Oracle 10g, WebSphere 5.0, WebSphere Message Broker, EJB 2.0, iBATIS, Oracle 360 Commerce, Sterling Commerce Yantra, WebSphere Portal, WPC, POSLog. Worked as an information architect for POS applications which includes application packages like 360 Commerce (now Oracle retail), Sterling Commerce Yantra, and StreamServe, including J2EE, ESB, Web-Services, DB2, SOI, SOA, BPEL, XSD, XML, and XPATH. Also, responsible for leading a project team in delivering solution to the customer. Also, responsible for managing scope, planning, tracking, change control, aspects of the project. Lead teams of DBAs and Data Analyst; responsible for Oracle Retail, Database and technical architecture, administration and performance tuning of a 7x24 mission critical applications, DB2 Replication. Implement WPC's MDM integration with the POS. Design and implement data Archive and purge Technical architecture for the store and central databases. Design store and central Database physical and logical design using ARTS data framework, Capacity and Scalability Planning, Backup and Recovery. Implemented Operation data reporting from different customer and sale data sources using an ODS layer. Review and provide recommendations to the Enterprise technical architecture strategy. Plan the data and software migrations including phase roll out and cut over. Translate customer requirements into formal requirements and design documents, establish specific solutions, and leading the efforts including programming and testing

that culminate in client acceptance of the results. Data Analyst, Data Modeler, and Data Architect, Data Steward, Data Warehouse Expert Suntrust Bank August 2006 to August 2007 Technology Used: Db2 LUW, AbInitio, Cognos, Erwin, IBM BDW (Banking Data Warehouse), Erwin. Analyzed and data modeled the EDW (Enterprise Data warehouse) and Data Mart for Basel II Accord. Analyzed 30+ Source System data received from clients to ensure accuracy and completeness. Profiled Source system data using AbInitio Data Profiler and SAS DataFlux. Resolved data discrepancies and data cleansing as necessary. Defined Data Quality rules, participated in Data Steward and Data Governance. Build TPR central database from 30 plus disparate source system across LOB, these involved more than 15,000 tables; 65,000 data elements and; terabyte of data. Designed Enterprise Schema for Leasing application and build Logical and Physical Data Model containing 1500+ tables. Designed and developed Data Mapping Application for 30+ disparate source systems (COBOL, MS Sql Server, Oracle, and Mainframe DB2), using MS Access, and UNIX Korn script. Produced source to target mapping documents for 65,000 data elements based on data discovery results. Design the data masking application for 30+ source systems. Tuned Data Mart Reports using optimized query and Materialized Query Tables. Designed ETL Logic for EDW and Credit Risk Data Mart. Data Architect Novartis AG June 2005 to August 2006 Technology Used: Oracle 10g RAC, Linux. Performed TPC-C Oracle 10G RAC performance benchmark for NAS and SAN storage on Red Hat Enterprise Linux OS. Proposed database architecture design based on benchmark. Data Service Architect, Database Designer Midwest ISO February 2004 to June 2005 Database, and Application Servers tuning specialist Technology Used: Oracle 9i RAC, Erwin, J2EE, and Spring Hibernate. As Database Tuning specialist Administrator/ Database Architecture, he is involved on the development/Testing/Integration and production databases for Client Project. As Performance Specialist he performed the application baseline and reviewed the client Oracle 9.x database architecture in respect with ORACLE-SUN-EMC Best practice methodology. As Performance Specialist he helped in designing, installation, and instrumentation Veritas i3 (7.2) and TeamQuest (9.2). Optimized database using different SQL queries and materialized views. Database Architect,

Database administrator, Data Modeler General Motors March 2003 to February 2004 Technology Used: Db2 for Z (OS390), Oracle 9i, Erwin, InfoSphere DataStage, MicroStrategy. Designed and maintained the ODS, CDW, and Data Mart's logical and physical data model to support NHTSA critical reporting needs. Implemented detail and efficient ODS and Data Mart. Administrated Databases Oracle 9.x (Sun Solaris) and DB2 7.x (OS390)) on the development/Testing/Integration and production databases for TREAD Project. Extrapolated, designed, and monitored database storage capacity planning. Database designer, Database administrator, Programmer, Data Analyst Price Waterhouse Coopers LLP December 2001 to March 2003 Designed database architecture for the development/Testing/Integration and production databases for nine client projects Lead a team for planning and Installing Oracle Database and client software across geography. Lead an effort to implemented application and database Security using database program, database roles, profiles, privileges, and views. Sr. Oracle Database Administrator, Data Modeler ADP/Statewide Insurance August 2000 to December 2001 installing, upgrading and patching Oracle RDBMS software (9i/10g/11g) creating, maintaining, monitoring and tuning Oracle databases setting up RAC, ASM and Data Guard configurations developing, maintaining, troubleshooting and tuning SQL queries and scripts created for application development and for DBA tasks creating and maintaining UNIX shell scripts (ksh, bsh) for development and DBA tasks Done Oracle Server installation, Database planning, Database creation, Configuration of middleware in Server & Client machine, Configuration of Client Machine, Server cache optimization, Space management, Monitoring of space usage, Security administration, Database Backup, Database Recovery, Monitored, Tuned database performance. Created Logical Model, Physical Model, Domain Dictionary, Trigger, Indexes, View, Procedure/Function, Pre/Post scripts using Erwin Data Modeling Tool. Education MS in Computer Science in Computer Science Big Data Technologies and smartCloud Enterprise (IBM BigData University) MongoDB University

Name: William Cooper

Email: cynthiaanderson@example.org

Phone: 001-248-218-2973x646