

Technology Lead - Hadoop Technology Lead - Hadoop Technology Lead - Hadoop - Neiman Marcus Dallas, TX Work Experience Technology Lead - Hadoop Neiman Marcus - Dallas, TX May 2018 to Present AA to store and join customer centric data like click stream, sales, email campaigns in generating UCIDs and personalization which are consumed by CXP through APIs. ? Developed data models for personalization and product recommendations using Storm, Kafka, Hive, Pig. ? Processing PLLC (Private Label Credit Card - Capitol One) data to provide insights into percentage of penetration of sales by associates and stores. ? Developed scripts for data migration of enterprise data from in-house infra to AWS cloud. ? Environment: Hadoop2.6-chd5.13(40 node), AWS cloud, Hive1.2.1, Storm, Cassandra, Solr, CouchDB, EC2, S3, Airflow. ETL/Hadoop Developer Lutron Inc - Philadelphia, PA April 2016 to April 2018 Hadoop and Informatica based ETL and analytical system to have insights about customer's usage of Lutron products across different product line leading to future enhancements, improvement in business and services. ? Developed data pipeline using Spark, Kafka, Hive, Pig and HBase to ingest customer system usage data and financial histories into Hadoop cluster for analysis. ? Developed Scala scripts, UDF's using both Data frames/SQL and RDD/MapReduce in Spark for data aggregation and writing data back into S3 through Sqoop. ? Extensively used Informatica to create data ingestion jobs into HDFS using complex data file objects such as AVRO and Parquet and to evaluate dynamic mapping capabilities. ? Implement Data Quality Rules using Informatica Data Quality (IDQ) to check correctness of the source files and perform the data cleansing/enrichment. ? Analyze log records data a day and its aggregated hourly, daily reporting using Tableau. ? Environment: Hadoop2.7, Informatica9.x, Hive1.2.1, Spark1.6, Teradata, Oracle, EC2, S3. Hadoop Developer Cisco Systems - San Jose, CA September 2014 to March 2016 Worked with highly unstructured and semi structured data of 100TB+ in size ? Developed Pig and Hive scripts to be used by end user / analyst / product manager's requirements for adhoc analysis. ? Used Informatica to validate and test the business logic implemented in the mappings and fix the bugs. Developed reusable Mapplets and Transformations. ? Managed External tables in Hive for optimized performance using Sqoop jobs. ? Solved performance issues in Hive and Pig scripts with understanding of joins, group and

aggregation and how it translates to MapReduce jobs. ? Exploring with Spark improving the performance and optimization of the existing algorithms in Hadoop using Spark context, Spark-SQL, Data Frame, pair RDD's, Spark on YARN. ? Worked with Hadoop-Kerberos security environment which is supported by the Cloudera team. ? Environment: 32 Node Hadoop 2.6 cluster, Informatica9.x, HDFS, Flume 1.5, Sqoop 1.4.3, Hive 1.0.1, Spark 1.4, HBase, XML, JSON, Teradata, Oracle, MongoDB, Cassandra. Hadoop Developer Bayer Healthcare - Leverkusen, DE November 2013 to August 2014 Migration of 100+ TBs of data from different databases (i.e. Oracle, SQL Server) to Hadoop. ? Wiring code in different applications of Hadoop and Informatica Ecosystem ? Extensively involved in performance tuning of the Informatica ETL mappings by using the caches and overriding the SQL queries and also by using Parameter files. ? Worked on various file formats Avro, SerDe, Parquet, and Text by using snappy compression. ? Used Pig Custom Loaders to load different forms of data files such as XML, JSON and CSV. ? Designed dynamic partition mechanism for optimal query performance of system using HIVE to reduce report time generation under SLA requirements. ? Environment: Hadoop 2.2, Informatica Power Center 9.x, HDFS, HBase, Flume 1.4, Sqoop 1.4.3, Hive 0.13.1, Avro 1.7.4, Parquet 1.4, XML, JSON, Oracle 11g, Amazon EC2, S3. ETL Developer Grattan Plc - Bradford February 2010 to October 2013 Developed mappings/sessions to import, transform and load data into respective target tables and flat files using Informatica Power Center for data loading. ? Automation of the Informatica ETL jobs for different ETL design pattern. ? Extensively used Transformations like Router, Aggregator, Source Qualifier, Joiner, Expression, Aggregator and Sequence generator by using Source Analyzer, Warehouse Designer, Mapping Designer & Mapplet, and Transformation Developer. ? Environment: Informatica Power Center 9.x (Repository Manager, Designer, Workflow Manager, and Workflow Monitor), Oracle 11g, SeaQuest, HPDM, SQL Server, Teradata, Toad, Control-M. ETL Developer Star Health and Allied Insurance Company Ltd - Bengaluru, Karnataka October 2007 to February 2010 Extensively used Slowly Changing Dimensions technique for updating dimensional schema. ? Processed data using various transformations like Aggregator, Router, Expression, Source Qualifier, Filter, Lookup, Joiner, Sorter, XML Source qualifier and web-consumer for WSDL.

? Used Informatica user defined functions to reduce the code dependency. ? Environment: Informatica Power Center 8.x, Informatica Power Connect, Power Exchange, Power Analyzer, Toad, Erwin, Oracle 11g/10g, Teradata V2R5, PL/SQL, ODI, Trillium 11. ETL Developer United Overseas bank - Singapore October 2005 to September 2007 Used SSIS as an Extract Transform Loading (ETL) tool of SQL Server to populate data from various data sources, creating packages for different data loading operations for application. ? Extensive use of Transact-SQL, stored procedures, trigger scripts for creating database objects. ? Generated various reports using features such as group by, drilldowns, drill through, sub-reports, Parameterized Reports. ? Deploying new strategies for checksum calculations, and exception population using mapplets and normalizer transformations. ? Environment: SQL Server 2005, T-SQL, SSIS/DTS Designer and Reporting tools, Control-M. Java Developer ExpertNet CAD - Pune, Maharashtra January 2004 to August 2005 Developed the web applications using Spring MVC Framework including writing actions/ classes/ forms/ custom tag libraries and JSP pages. ? Worked on Integration of Spring and Hibernate Frameworks using Spring ORM Module. ? Implemented caching techniques, wrote POJO classes for storing data and DAO's to retrieve the data and did database configurations. Java Developer Honeywell - Bengaluru, Karnataka January 2002 to January 2004 Implementation of routing and shortest path algorithms along with parsing logic for device discovery using Heart-Beat ? Implementation of Java Native Interface(JNI) API's for Indus Mote to access devices dynamically through C code. Education Master's in Data Analytics in Data Analytics Boston University - Boston, MA

Name: Thomas Crawford

Email: ewatson@example.org

Phone: 671.365.9335x971