**Anmol Raina** *anmol.raina21@gmail.com*

**Data Architect Contact: +919731961166**

# Skills Synopsis

|  |  |
| --- | --- |
| Big data : | Hadoop , Spark , Python ,Scala, Hive ,Hbase ,Cassandra ,Pig, Sqoop, Oozie ,NoSQL ,AWS EC2/S3/Snowball/Redshift, Docker. |
| Others : | Data Modeling, ETL, Obiee, Informatica, Workday, Database, Unix. |

**Experience Summary**

Over eleven years of experience in  building data pipelines/data centric applications using distributed storage platforms like HDFS, S3, NoSQL databases (Hbase, Cassandra, etc) and distributed processing platforms like Hadoop, Spark, Hive, Oozie, Airflow, Kafka etc in a production setting. Also actively indulged in People management hiring, performance management.

* Working as a Big Data Architect/Engineering Manager for last six years with strong background of big data stack with major focus on Spark, Scala, Kafka, Hadoop, HDFS, MapReduce, Hive, Mongodb, Hbase, Python, Sqoop, and PIG.
* Work on creating solution for business problems and demo pilot use cases .Also work on defining technical and business requirements for data related functionality and plan and convert the requirements into functional and logical architecture.
* Experienced in multi-dimensional modeling like star schema, snowflakes, normalized and de normalized models.
* Strong experience in working with and optimizing existing ETL processes and data integration and data preparation flows and helping to move them in production.
* Work with partners and manage engineering team and actively participate in hiring, major responsibilities include team member development, training and performance management/compensation planning.

**Professional Experience**

## Data Knight:

## Freelancer

## Jun 2019-Present

* Provide overall architect responsibilities including roadmaps, leadership, and planning, technical innovation.
* Acting as a senior architect, provide technical and process leadership for projects, defining and documenting information integrations between systems and aligning project goals with reference architecture
* Design multiple POCs using Spark and deployed on the Yarn cluster, compared performance of Spark, with Hive and SQL.
* Research, evaluate, and utilize new technologies/tools/frameworks centered on Hadoop and other elements in the Big Data space.
* Handled importing of data from various data sources like Oracle into HDFS vice-versa using Sqoop. Migrating various Hive UDF's and queries into Spark SQL for faster requests.
* Did setup aws cloud environment through initial scope and sizing, data materiality and risk assessment, platform services and security aspects like IAM and performance improvement leveraging Cloud Watch.
* Participate in hiring and mentoring, team of data engineers and data analysts.
* Implemented cloud services IAAS, PAAS , and SaaS which include Openstack , Docker and Openshift .
* Used Kubernetes to orchestrate the deployment, scaling and management of Docker Containers.

##  Teksystems (Morgan Stanley):

## Data Architect

## Jun 2017 -May 2019

* Involved in converting Map Reduce programs into Spark transformations using Spark RDD's on Scala. Import data from different sources like HDFS/HBase into Spark RDD, developed data pipeline using Kafka to store data into HDFS.
* Performed real time analysis on the incoming data. Involved in creating Hive tables, loading with data and writing hive queries which run internally in MapReduce.
* Handled importing of data from various data sources from Oracle into HDFS vice-versa using Sqoop. Performance optimization when dealing with large datasets using partitions, broadcasts in Spark, effective and efficient joins, transformations during ingestion process
* Used OOZIE engine for creating workflow and coordinator jobs that schedule and execute various Hadoop jobs such as MapReduce Jobs, Hive, Spark and automating Sqoop jobs.
* Managed Docker orchestration and Docker containerization using Kubernetes.

## Teksystems (Cisco):

## Data designer/Architect

## April 2015-May 2017

* Build de normalized warehouse for end reporting needs and design ETL changes to load dimensions and fact tables.
* Provide technical leadership and governance of the big data team and the implementation of the solution architecture in Hadoop ecosystem.
* Customized pentaho jobs to load complex facts, before moving the structure to big data system.
* Deploy scalable data warehouse system on cloud using AWS Redshift. Defining spark/python (PySpark) ETL framework and best practices for development. Did work in an agile setup to estimate, plan and execute projects with other stakeholders.
* Integrate various legacy data sources into CRM connect warehouse system, Extensive use of python, shell scripts and Informatica to do cleansing, massaging and to apply business rules on contact employee relationship data.
* Work with partners and manage engineering team and actively participate in hiring and mentoring team members.

## KBACE Technologies:

##  Solution Architect

##  November 2013 - Jan 2015

* Responsible for architectural design, data/dimensional modeling and ETL development initiatives for client Data Warehouse System.
* Analyze business needs, collaborate with stakeholders and successfully build a Dimensional Modeling (Star Schema) ,on the basis of existing business processes, dimensions and measured facts.
* Successfully led Data Extraction, Data Warehousing and Analytics initiatives that reduced operating costs and created customized programming options.
* Extensively worked on developing Obiee Dashboards using guided analytics, interactive dashboard design, and visual best practices. Contributed diligently in fine tuning, and increase performance of analytical reporting system.

## KPMG Global Services:

## Senior Consultant

## May-2013 – Oct-2013

* Responsible for Oracle Business Intelligence 11.1.1.7 implementation with a team of internal staff and external consultants
* Coordinated initial planning phases for corporate initiative to retire legacy management reporting system.
* Developed high-visibility and high-priority Interactive Reporting dashboard solution for operations reporting.
* Involved in iterative development of OBIEE dashboards, Answers requests, charts, pivot table layouts, dashboard prompts.

## Motorola Mobility:

## Senior Consultant

##  Oct-2011 – Jan-2013

* Responsible for Architectural design and implementation for an in-house OBIEE project.
* Integrated the data into the enterprise data warehouse from various source systems.
* Involved in implementing various hierarchies and security restrictions for the corporate data within

OBIEE metadata.

* Responsible for analysis and design of 40+ reports for OBIEE (Oracle Business Intelligence Enterprise Edition).
* Designed Custom Data model to accommodate the changes from People-soft 8.8 to People-soft 9.1.

## Wipro Technologies: Consultant

##  Nov-2009 – Oct-2011

* Involved as an independent developer to customize the OBI Apps implementation for one of the MNC clients.
* Worked on to translate business rules into PL/SQL objects, extract business rules from analysis of existing PL/SQL.
* Customized DAC console to combine various OOTB mappings with customized ones to load data warehouse.
* Analyzed end-user query performance when implementing and tuning OBIEE.
* Lead the data team to convert, extract, and transform data according to requirements. Involved in

Customizing DAC ETL execution plans.

##  Desai Brothers:

**April-2009 -Oct 2009**

* Worked as a data developer to load data from various sources into Oracle data warehouse
* Worked on creating data models using ERWIN and creating scripts for physical model implementation.
* Wrote PL/SQL procedures to implement business logic before loading into end warehouse.
* Generate reports from warehouse and apply computation and graphical representation in Microsoft excel, before sharing end reports via email to end business users.

# Education:

# Bachelors of Engineering from Mumbai University

# Interests:

# Startups, Finance, Data and Swimming