● Anand Mahajan   ● asqlguru@gmail.com   ● +91-9730660890   ● Bangalore , India

Hands-on expert of analytics with an established track record of quickly extracting and envisioning gems of information buried in vast amounts of facts.

Professional Summary

· **Over 14 years as Data Science Analytics /Lead/Architect**
- Expert in building high-performance database designs and the operational infrastructure to support such designs
- Presently working on building Data Pipeline in **AWS tech** stack with **Pyspark** and **EMR** clusters with data volume of more than 50TB/monthly
- Using various AWS tech stack tools such as Kinesis , Lambda , Glue , EMR
- Managing people from various Verticals with various areas of expertise such as Developers, DBA, QA, and Support team
- Experience in managing teams of 3-10 people for various projects
- Developed strategic plan for capacity planning and implemented the same for maintenance, monitoring, security and disaster recovery, load balancing, and replication
- Designed and developed complete workflow for building state-of-the-art data warehouse, from importation to data feeds to implementation of the BI solution with market/executive-focused dashboards
- Developed data pipeline and automated product metrics using Big Data Platform using Hive, Spark, Impala
- Used Integrated Data Ingestion between Mongodb, Hadoop and Vertica Platform
- Development/administration experience with cross-platform databases, such as Vertica, Greenplum, Postgres, MySQL, SQL server
- Experience with commercial BI tools, such as Tableau, Qlikview, Cognos, Microstrategy), as well as open source (Birt)
- Proficient in SQL (9/10), Python (7/10), PL/SQL (8/10), Bash (8/10), Perl (5/10)
- Acquired extensive experience in business analysis, entity relationship, and dimensional **data modeling, data warehousing** (Kimball Methodology), GUI design, OLTP system development based on relational and hierarchical databases, as well as experience relevant to OOD and structured systems analysis, design, programming, and testing
- Well versed in system analysis, **ER/dimensional modeling**, database design and implementing **RDBMS** specific features
- Excellent in **normalization/denormalization** techniques for effective and optimum performance in OLTP and OLAP environments

Education

Louisiana State University, Baton Rouge, LA, USA
- **MS in Engineering Science** (Information Systems), GPA: 3.70
- Concentration: Database Design Concepts (Worked as TA 1.5 yrs. Database Design course

Regional Engineering College, VNIT, Nagpur, India
- Bachelor of Engineering, Mechanical Engineering (66% Avg GPA 3.7)

Founder/Data Strategist   Frena Analytics/Gajadata (India/US)   Jun 2019 – Present
- Working with US clients on data projects (LVMH – NY) & Georgia Pacific (Koch Industries) - Atlanta
- Main goal is to build data pipeline from IOT data
- Incoming data volume is about 2 TB on daily basis
- We are converting all old architecture into new architecture to process data faster with 99.99% accuracy using Pyspark and EMR clusters .
- Using Pyspark data frame architecture to convert data from unformatted json files to Parquet file format.
- Able to process 100GB of data within 10 min using 20 node EMR clusters.
- Using Redshift in some of our data pipeline to store aggregate and summarized data.

Data Analytics Executive   Samsung (New Delhi, India)   Jun 2018 – Jun 2019

- Worked with team in extracting, Storing and analyzing big data from GSIM/SA(Home Grown Analytics Platform) with the use of high level analytical queries such as session, partition by , cube , rollup

- Well versed with extracting raw data from Hadoop Analytics using Hive Queries.
- Done detailed SWA market analysis mainly in camera segment w.r.t. various features, settings, modes for over a period of 3 months using GSIM/SA and Hadoop Raw data.
- Business strategy and planning with high focus on implementation and execution. Have led multiple strategy and transformation projects in business and process domains mostly in the camera segment.
- Build on automation methods to track new features in India region viz , Chat-over-video, Social camera , Dual Camera, app switching etc.
- Worked on advanced analytics to get 3rd party app usage across various regions in India region across operators extracting GSIM data from Hadoop analytics using Hive queries
- Working from Top Down and Bottom up approaches in meeting end goals for Businesses with data analytics.

Data Consultant   **Equinix**   **Feb 2017 – Apr 2018**

- Equinix is the leading global colocation data center provider by market share, and it operates 175+ data centers in 44 major metropolitan areas in 22 countries on five continents.
- Worked with a team in Data migration for all their Backend Databases from Oracle and MongoDB to Postgres.
- Architected whole solution after understanding business process and release dates
- Build complete end to end solution by using python and sql programming framework
- Build ETL framework using python and using SQL to carry out all data processing activities.
- Using Spark as our go to tool for data processing in a cluster environment.

Data Architect   **Ringcentral, San Mateo CA**   **Apr 2015 – Nov 2016**

Ringcentral (NYSE: RNG) is a publicly traded provider of cloud-based phone systems for businesses

- Worked with Internal as well as external customers, to best understand and utilize the cloud platform.
- This includes all of the services AWS, Hadoop as our Big Data Platform along with Vertica, Reporting (Tableau, Big Data Discovery, D3) and Datawarehouse (Oracle & Postgres) and ETL (Talend, Python)
- Mongo Hadoop-Vertica Integration:

- RNG purchased social media company Glip. integrated all Data hosted by Glip, on mongo servers to Hadoop Cluster and pushed to Vertica cluster for Reporting.
- I led all efforts to make it happen including Architecture, Design development.
- Extracted Data from JSON format to Hadoop cluster using JSON Udf.
- Used Spark stream and Spark SQL streams to create data Pipeline

- Platform Reporting (Hadoop-Oracle-Web-Vertica):

- Integrated all Platform reporting data (Hadoop) with Customer data (Oracle) and API information (real time live data on web) and push to Vertica platform on a daily basis for reporting purpose.
- Migrated all Talend Jobs to Homegrown Python framework

Technology Stack: Python, Hadoop stack (Hive, Spark,Oozie,Talend,Impala,Kafka),Vertica, Shell, MySQL, Postgres, Oracle ,Tableau

Analytics Lead   **Castlight Health, San Francisco CA**   **May 2013 – Apr 2015**

Castlight Health is a pioneer company in digital Healthcare and went IPO in April 2013.

- Led a team of 5-7 people to work on next gen Data Science warehouse/ BI architecture.
- Created framework for ETL such that all ETL scripts will run under it. This helped in giving a centralized repository for all of ETL scripts, which indeed led to centralized logging monitoring and easier maintenance and reporting.
- Worked on Jaspersoft reporting, handled all aspects of it including purchasing, training employees, getting servers configured, Installation and finally talking to business owners and creating required employer specific dashboard.

- Managed Administration of Qlikview (reporting analytics OLAP tool) and guided developers in creating better reports by helping them in Query optimization, Data Architecture.
- Analyzed Individual ETL scripts and worked with the team to make sure we completely revamp ETL scripts. This helped us in reducing ETL load time from 15 hours to 3 hours.
- Knowledge of Amazon EC2, Amazon S3, Amazon SimpleDB, Amazon RDS, Amazon
- Elastic Load Balancing, Amazon SQS, and other services of the AWS family

Technology Stack: Python, Shell, Reporting Tools (Jaspersoft, Django, Qlikview), MySQL, Postgres, and Greenplum, AWS platform

Sr. Database Engineer/Sr. MTS   VMware, Palo Alto, CA   Feb 2007 – May 2013

Packetmotion was acquired by VMware as part of a security product for virtualization

- Responsible for all data integration aspects of product, including data analysis, migration, replication, and business intelligence
- Created design document for data integration and migration for various data sources to build an integrated application environment
- Worked with Product Managers and Solution Architects to understand the business objectives for data integration projects and architected the OLTP and OLAP model
- Used Hadoop as Big Data Solution for the replacement of traditional RDBMS in combination with Postgres.
- As a part of a team, migrated all existing PacketSentry infrastructure to a virtualized environment
- Installation and configuration experience with ESX server
- Hands-on experience with various VMware products, including Vcenter, Vsphere, Vconvertor, and Workstation
- Worked as Lab Proctor for infrastructure labs during hands-on labs at VMworld 2012. At VMworld, we ran 9,080 labs and deployed over 110,000 VMs.

Technology Stack: Python, Shell, MySQL, Postgres,Pivotal cloud, Oracle, Hadoop stack (Hive,Spark,Oozie,Impala), VMware Stack (Vcenter, Vsphere, Vconvertor, and Workstation)

   Notable projects at Packetmotion include:
- Streamlining the data archiving process
  With an embedded database, we need to make sure that data are purged on their own, based on various requirements (e.g., scheduled, space, corrupt data, ad-hoc).
- Source sync (extraction of native audit data from customers' databases, push and analyze them into our product) for Oracle, SQL Server
  This involved complete SDLC, which includes designing the whole process, developing client side code, creating ETL and fact tables to store data, and generating various analytical reports based on the extracted information.
- Re-architecting reporting system for performance
- Aggregation functionality, analysis, and reporting as a deductive process for discovering hidden patterns in customer data.

Technology Stack: Shell, Oracle, Pl/sql, Oracle, Hadoop stack, Birt for reporting