

SAMEER AHMAD SIDDIQUI

(813) 947-8560

Sameerahmad.dba@gmail.com

<https://www.linkedin.com/in/sameer-ahmad-siddiqui>

SUMMARY:

- Over 10 years of total Information Technology experience.
- Extensive experience in Designing, Creating and Securing Multitenant enterprise Big Data platforms.
- Extensive experience in architecting solutions on public cloud like Amazon Web services, Microsoft Azure and Google Computing Services as well as private cloud using containers and virtualization.
- Designed and delivered Big Data Platform as a service solution using Cloudera Director, Kubernetes and Bluedata for Hadoop cluster.
- Having 5+ years of comprehensive experience as a Bigdata Consultant, architecting diverse use cases.
- In depth knowledge of Security infrastructure for Data security, Data obfuscation, authentication, authorization and Data encryption at rest and encryption over wire.
- Expert in designing and implementing orchestrating pipelines using Apache airflow and apache oozie.
- Expert in writing Ansible playbooks for automation and managing Ansible tower.
- Experienced in Performance optimization of SQLs, spark Job. Hands on experience on performance monitoring tools.
- Experienced in Designing and implementing job orchestration using Apache oozie, Apache airflow, CA7 and other tools.
- Hands on experience on data governance and data lineage solution with Apache atlas and Cloudera Navigator.
- Strong Unix experience with Red hat Linux including Red hat IPA server.
- Experienced with performance debugging for DNS, TCP/IP, Memory, CPU, load balancing and Kernel issues.
- Strong RDBMS knowledge with expert in performance management on Oracle, SQL server, PostgreSQL and MySQL
- Expert in Oracle Exadata implementation, Exadata specific performance tuning, Database consolidation and migration.
- Experienced in working with Agile Methodologies.
- Rich experience in deploying Enterprise Manager 12c Cloud control Expert in Oracle database 8i/9i/10g/11g, Real Application Cluster (RAC), Oracle Data Guard And Goldengate.
- Mentored associate engineers with their work deliveries and managing career goals.
- Excellent communication skills. Ability to think analytically with strong interpersonal, management & decision-making skills.

EDUCATION:

- Pursuing **Master of Science (MS)** in Computer Science, Specialization in Machine Learning from Georgia Institute of Technology - 2020 - 2022
- Bachelor of Engineering (ECE) from Anna University – 2006 – 2010

CERTIFICATIONS & TRAINING:

- Amazon Web Services Solutions Architect Associate
- Cloudera Certified Associate Administrator for Apache Hadoop
- Cloudera Developer Training for Spark and Hadoop
- Oracle Exadata 11g Certified Implementation specialist
- Oracle Database 11g Certified Professional.
- Oracle Database 12c: New features for administrators
- Oracle Database SQL Certified Expert.
- Oracle Enterprise Manager Cloud Control 12c: install, upgrade and Using EM cloud control 12c
- ITIL Foundation certified - version 3.
- EMC certified Information and storage associate version 2.
- M102 - MongoDB for DBAs certified

TECHNICAL SKILLS

Big Data Technologies	Spark, Apache Hadoop, Map-Reduce, HDFS, Pig, Hive, Hbase, ZooKeeper, Sqoop, Flume, OOOIE, Solr, Apache Nifi, Apache Kafka
Languages	Python, Scala, Ansible YAML, Unix Shell Scripting, Java.
Relational Databases	Oracle 12c/11G/10G, SQL Server, Postgres, MySQL, Amazon Redshift, Amazon Aurora

EXPERIENCE:

June 2020 – Till Date, The Nielsen Company, Tampa, FL

Description:

- Build a data comparison platform using spark on Kubernetes to validate data produced from parallel systems.
- Designed and implemented the solution for migration to Centos from Redhat for the Big Data platform.
- Designed and Implemented migration of platform to cloudera express from enterprise to save license cost with huge overall cost savings.
- Built an OCR and ICR solution using pytesseract, and google vision API to automate the process of reading the forms and populating the databases.

Dec 2017 – June 2020 Solutions Architect, Cloudera, Tampa, FL

Description:

Responsible for assisting clients with their Hadoop deployments, which entails planning, configuring, troubleshooting, tuning and maintenance. Also provide assistance to Solutions Architects as needed. Responsibility also involves system administration, development, architecture/design for big data ingest, data wrangling, descriptive/predictive analytics, search, systems and security integration engagements along with consulting responsibilities.

- Architected and implemented a big data solution on Private cloud for the research and development group of GSK with container-based Cloudera Cluster deployment using Bluedata on Docker Swarm.
- Automating the deployment of KMS and KTS and encryption over the wire for transient clusters.
- Designed and built the process to automatically tune the cluster configurations based on usage type.
- Design and build the process for scale-up and scale-down clusters on the private cloud.
- Worked Kubernetes based deployment of Cloudera Data Science Workbench to utilize the GPU for enhanced processing.
- Designing the solution for big data deployment for Jabil circuits on AWS using spot EC2s, RDS, S3, route53 , EFS and AWS Secrets manager.
- Guiding the Jabil's AWS cloud architects on building enterprise-wide cloud strategy and deployment patterns.
- Building the internal REST API to fill the gaps between Cloudera director capabilities and customer requirements.
- Deploying cluster for GSK on Microsoft Azure working with Cloud architects and Automating the cluster deployment using Cloudera Director.
- Analyzing the GWAS (Genome-wide association study) data using Hail on Spark and built graph examples.
- Architecting the solution for enterprise data lake solution on Azure using ADLS V2 and Impala for an energy company.
- Architected enterprise Data Lake using Microsoft PowerBI and Impala on S3 with Active Directory managed Role Based Access Control.
- Designed and built python-based REST API to allow application teams to perform privileged actions automatically which are typically done by the admin team.
- Streamlined the onboarding process and the turnaround time was eliminated.
- Designed API version management for continued development.
- Designed and implemented secured connection to dynamic Hadoop clusters.
- Designed and implemented API authentication and authorization for better management.
- Worked with the platform team of Anthem to secure the Nifi platform
- Architected the Flow lifecycle deployment cycle using the NiFi registry.
- Worked on tuning and implementing best practices for developing the NiFi flows.
- JPMC bank designed a log aggregation tool using Spark streaming and Kafka.
- Extensive experience in tuning Kafka deployments.
- Used Cloudera search(SOLR) to index the entries and create a search box.
- Designed and implemented Apache Spot at Comcast and FirstEnergy. Apache Spot is a community-driven cybersecurity project
- Designed and implemented internal solutions for implementing GDPR.
- Extensive knowledge of implementing CCPA (California Consumer protection Act) for major banks.

Jan 2015 – Dec 2017 Hadoop Admin/DevOps Engineer, TCSL, Tampa, FL

Project Name: AC Nielsen, Tampa, FL

Description: Designing, creating and managing Hadoop Cluster on AWS and On Prem Environment. Writing Ansible playbooks for Infrastructure as Code implementation and automation of application deployment and management on AWS.

- Setting up and managing Hadoop cluster for the Production Environment.

- Setting up RDS, redshift, EC2, Auto scaling groups for application teams.
- Setting up cloud watch monitoring automatically using Ansible playbook.
- Installed, configured and deployed a 100 node CDH Cluster for Development and Production.
- Creating and managing the Best practices documents for team.
- Maintaining and troubleshooting Hadoop core and ecosystem components (HDFS, Map/Reduce, Name node, Data node, Job tracker, Task tracker, Zookeeper, YARN, Oozie, Hive, Hue, Flume, HBase, and Fair Scheduler)
- Release Process Implementation like DevOps and Continuous Delivery methodologies to existing Build & Deployment Strategies.
- Implemented backup setup various Cluster components like Namenode, Hive Metastore, and Cloudera Manager Configuration. Successfully restored the services from backup to address hardware failures.
- Set up Load balancer for Impala daemons.
- Set up High Availability for Name node, Resource manager, Hive, oozie and Hue.
- Worked on Hadoop CDH upgrade from CDH4.x to CDH5.x and patching CDH for bugs.
- Worked on Cloudera manager integration with LDAP for authentication/authorization.
- Successfully enabled Kerberos on all clusters to maintain security compliance.
- Successfully setup a Key Management Server to enable HDFS encryption to meet security requirements to achieve HDFS encryption at rest
- Setup and use of BDR (Distcp) to copy data and perform data backup.
- Worked on managing Hadoop clusters and provisioning Hadoop clusters on AWS using Cloudera Director 2.0.
- Worked on setting up Sentry and managing user permissions through Sentry and HDFS ACLs.
- Setting up Red hat IPA server to manage users and groups on clusters.
- Worked on Hadoop clusters capacity planning and management.
- Monitoring and Debugging Hadoop jobs/Applications running in production.
- Worked on Providing User support and application support on Hadoop Infrastructure.
- Worked on HUE integration with LDAP for authentication/authorization.
- Load data from various data sources into HDFS using Flume
- Experience in setting up and managing the batch scheduler Oozie and also worked on spark.
- Experience in analysing data in HDFS through Map Reduce, Hive and Pig.
- Used Sqoop to efficiently Transfer data from Oracle Exadata to HDFS, Netezza to HDFS.
- Reviewing ETL application use cases before on boarding to Hadoop.
- Worked on Amazon Web Services (AWS) for a multitude of applications utilizing the AWS stack (Including EC2, VPC, Glacier, Route53, S3, RDS, Cloud Watch, Cloud Trail, SNS and IAM), focusing on high-availability, fault tolerance, Load balancing and auto-scaling in designing, Deploying and configuring.
- Involved in Implementation of various AWS cloud Application & deployment using Amazon EC2, S3, RDS(MySQL and Postgres), Elastic load balancing, Auto scaling, Elastic Block Storage, Cloud Watch, Identity Access Management, Implementing Virtual cloud platforms using Amazon Virtual Private Cloud.
- Deploy and monitor scalable infrastructure on Amazon Web Services (AWS) & configuration management using Ansible.
- Implementing and setting up Ansible Tower server
- Implementing and setting up AWS security groups for AWS Instances on AWS Environment.
- Implementing and setting up Route 53 for AWS instances
- Automated AWS infrastructure deployment like VPC creation using Ansible.
- Creation of subnets and Route tables, Internet gateway, virtual gateway customer gateway, Create VPN connections, Create VPC peering between Many VPC's.
- Creating S3 buckets and managing the data lifecycle automatically using Ansible.

- Working with Amazon Snowball to move large scale data from On Prem to S3.
- Store, manage & Monitor applications on S3 storage, including accessibility, Cross Region Replication and create and monitor Events, using Cloudberry Explored and s3cmd CLI.
- Provided security and managed user access and quota using AWS Identity and Access Management (IAM), including creating new Policies for user management in JSON
- Defined AWS Security Groups
- Create Auto-scaling group and setup suspend process resume process for all UCAS application.
- Work with python Boto3 SDK to automate backups to S3.

Environment: Cloudera Distribution of Hadoop, EMR, AWS, Ansible

Sep 2014 – Jan 2015 *Database Architect, Tata Consultancy Services, Chennai, India*

Project: AC Nielsen

- Reduced 20% licensing cost by oracle instance consolidation.
- Developing the Database Standard document for oracle installation, upgrade of database and database Standard Operating Procedure.
- Managed multiples projects simultaneously dealing with Application Developers, Unix Admins, Storage Admins and Project Managers.
- Exadata Consolidation: migrated and upgraded multiple production databases to Exadata full rack with over 45% overall performance improvement.
- Integrated Oracle databases with LDAP using Functions and triggers to automate user management on Production database improving security compliance.
- 20 person month effort saving by automating the using creation/drop process using Shell scripting.
- Tested and prepared standard implementation document for Oracle 12c.
- Secured Personal Identifiable Information by implementing Oracle Data Redaction and Virtual Private Database.
- Migrated Databases which are Not Mission Critical to EnterpriseDB (PostgreSQL), slashing license and renewal expenditure.
- Performed backup review for whole environment
- Performed Disaster Recovery test to develop and document recovery procedure for entire data centre.
- To do database performance review to estimate benefit and ROI from Exadata Migration.
- Exadata specific performance tuning using smart scan, Query offloading and storage indexes.
- Worked with Oracle consulting and sales teams to negotiate and finalize oracle licensing costs and procuring licenses.
- Successfully presented proposed architecture to CTO office and implemented many projects.
- Database design using tools like DBArtisan, SQL developer, Toad etc.
- Internationally culture sensitive with ability to work productively with a multicultural remote team.
- Use of MongoDB for building large data warehouse, Implemented shading and replication to provide high performance and high availability.

Environment: Oracle 9i, 10g, 11g and 12c.

Sep 2010 to Sep 2014

Oracle Database Admin, Tata Consultancy Services, Chennai, India

Project Name: Nielsen

Description:

- Resolved Severity 1 issues within SLA of 4 hours and performed detailed Root cause analysis to provide a permanent fix.
- Automated periodic Exacheck run and resolved all reported issues.
- Managed projects successfully involving new environment builds and upgrades of existing environment.
- Applied Quarterly Oracle Bundle patch on Exadata machines.
- Configured and performed regular backup to ZFS, using RMAN backups, secondary copies.
- Working with Application Team in solving the issues with Batch jobs, Long running queries and related tuning
- Performed user and security administration activities, such as, user profile creation, enrolling users, granting privileges to users, and maintaining system security, handing user connections
- Database maintenance activity-reorg
- Restoring the backups and test the backups (RMAN,COLD and HOT)
- Taking and monitoring RMAN Backups
- Ensuring that storage, archiving, backup and recovery procedures are followed properly
- Completed Refresh different databases using Oracle Export / Import and Datapump utility
- Running Tuning advisor and transporting SQL plan baselines from one DB to production
- Constant Interaction with other support teams – Server, Backup & Storage and Applications team to ensure effective and efficient project delivery.
- On-call support on off-business hours.
- Checking Alert Log errors and to make necessary actions to rectify them.
- Well-versed and comfortable in using OEM and Grid control of the databases.
- Daily Status meeting with the customer
- Handling Service review meetings
- Working proactively with various support teams for the client across the globe
- Liaison with Oracle support for resolving complex technical issues
- Experience in handling tickets and scheduled change Requests and service requests using tools like ServiceNow
- Installing and configuring Enterprise manager 12c cloud control for enterprise monitoring.
- Installation of oracle 8i , 9i ,10g and 11g software on Oracle Solaris, Linux and Windows platforms.
 - Experience in creating and maintaining Physical standby database using command line interface and testing switchover and failover operations
- Experience in creating and maintaining Real Application Cluster Databases and architecting RAC environments.
- Building and managing clusters using VCS, Redhat etc.
- Applying Bundle patches on Exadata Machine and upgrading engineered systems.
- Architecting database servers and database environments.
- Planning the strategy of backup and recovery using user managed, EMC snapshot backups, RMAN, datapump etc.
- Implementation of SAP database, with Data guard and EMC snap view to achieve RTO of 30 minutes.
- Installation of Oracle Exadata, Creation of Disk structure on storage cell, setting up RAC.
- Database monitoring which includes basic health checks, growth of objects, and performance of the database with the help of Enterprise 12c grid and HP monitoring tool.
- Refreshing and cloning of databases using RMAN in RAC and non RAC Environments.
- Handling issues on ASM and RAC environments.
- Worked with Oracle Corporation and opened SR's to resolve the issues.
- Worked with oracle support for procuring and managing licenses.

- Database migration and consolidation to Exadata machines.
- Production DBA, maintaining high level of Data and System availability
- Worked on Oracle 11g binary installation and upgrade
- Responsible for Installations, Migrations and Upgrades. Recently upgraded databases on LINUX & UNIX to 11gR2 version.
- Applying Quarterly Oracle CPU Patches and other bug fixing patches. (OPatch utility)
- Performed user and security administration activities, such as, user profile creation, enrolling users, granting privileges to users, and maintaining system security, handling user connections
- Database maintenance activity-reorg
- Configuration and maintenance of 10g & 11g DataGuard (Physical standby)
- Restoring the backups and test the backups(RMAN,COLD and HOT)
- Taking and monitoring RMAN Backups
- Ensuring that storage, archiving, backup and recovery procedures are followed properly
- Completed Refresh different databases using Oracle Export / Import and Datapump utility.
- Constant Interaction with other support teams across the globe – Server, Backup & Storage and Applications team to ensure effective and efficient project delivery
- On-call support on off-business hours.
- Checking Alert Log errors and to make necessary actions to rectify them.
- Recovery of lost objects and data files
- Analysis of objects or schemas to improve the speed of the database
- Running Tuning advisor and transporting SQL plan baselines from one DB to production
- Well versed and comfortable in using OEM and Grid control of the databases
- Handling all calls in production queue
- Daily Status meeting with the customer
- Handling Service review meetings
- Liaison with Oracle support for resolving complex technical issues

Environment: Oracle 9i, 10g, 11g and 12c.