# Amit Tyagi

**Senior Data Engineer - DBS Singapore**
- Email me on Indeed: http://www.indeed.com/r/Amit-Tyagi/d37cd09e2229ae44

## Work Experience

### Senior Data Engineer

DBS Singapore
November 2018 to Present

Description: Delivering Big Data Platforms as a service to internal customers in a self-service environment. Apart from maintaining and automating the existing platform, the team is supposed to explore other useful technologies and integrate them in the existing platform to support analytics, data science and other projects.
Role:
• Explore and evaluate new technologies and solutions to push our capabilities forward and take on tomorrow's problems not just today's.
• Building software and systems to be responsible for the infrastructure and applications through automation.
• Capacity planning and management.
• Developing frameworks for End to End Data Pipeline.
• Developing and Implementing the use cases by using various latest technologies.
• Resolving customer issues related to Big Data Platform.
• Optimization of the Big Data Platform.

### Senior Data Engineer

DBS - on payroll of Comtel
November 2018 to Present

### Linux, Hadoop, HIVE, Python, Cloudera Manager, Docker, Nifi, Ansible, Graylog, Grafana

Royal Bank of Scotland
November 2017 to November 2018

Tools & Technology:
Linux, Hadoop, HIVE, Python, Cloudera Manager, Docker, Nifi, Ansible, Graylog, Grafana, Alluxio, ELK, Impala, Spark, Presto

Company: Royal Bank of Scotland Project: Data and Analytics
Duration: Nov17 - Nov 18 Tech Lead- Big Data Platform
Description: EDW is the Enterprise level data Warehouse having a business use case for Analytics purposes. EDW is OLAP system used mainly for Analysis and Reporting for Regulatory and as a single ledger across the organization. As a part of Big Data Solution, the team is handling development of four different use cases and actively involved in POCs for upcoming use cases. Cluster size is 150 nodes in Prod.
Role:

• Working with multiple teams in understanding the requirement's and shaping the overall architecture

• Developing and Implementing the use cases
• Managing the Big Data Platform for RBS
• Maintenance and optimization of the cluster
• Automation of manual tasks using Python and shell scripting
• Resolving job failures/performance issues
• Handling user's queries/ issues
• Cluster storage planning
• Implementation of various tools and frameworks in the Hadoop
• Resolving the issues within SLA's.

Tools & Technology:
Linux, Hadoop, HIVE, Python, Cloudera Manager, Sqoop, Kafka, Tableau, HBase, Ansible

### Analyst
Royal Bank of Scotland
November 2017 to November 2018

### Tech Lead-Big Data Platform
Adobe
May 2016 to November 2017

Description: Data is generated through multiple source systems which are fed into hdfs using different data importing tools such as Kafka, Sqoop, scripts etc. Data is processed using hive and pig and then later used by different BI teams for reporting purposes. Kyvos cubes are primarily used to reporting along with Tableau. Cluster size is 180 nodes in Prod.
Role:
• Working directly with the client at client's location
• Managing the Big Data Platform for Adobe
• Maintenance and optimization of the cluster
• Automation of manual tasks using Jenkins, Python and shell scripting
• Resolving job failures/performance issues
• Handling user's queries/ issues
• Cluster storage planning
• Implementation of various tools and frameworks in the Hadoop
• Resolving the issues within SLA's.

### Tech Lead
Tech Mahindra
May 2016 to November 2017

### Linux, Hadoop, HIVE, Python, Cloudera Manager, Sqoop, Kafka, Kyvos, Tableau, HBase
Morgan Stanley
June 2015 to May 2016

Tools & Technology:
Linux, Hadoop, HIVE, Python, Cloudera Manager, Sqoop, Kafka, Kyvos, Tableau, HBase, Docker, Kubernetes

Company: Genpact Client: Morgan Stanley
Duration: June 15 - May 16 Team Lead-Hadoop Admin
Description: Data is generated through calls and web usage. A customer call coming through the PSTN is received first at the IVR (interactive voice response). In the IVR, the customer is asked a series of questions, to determine his service needs. Based on the voice response of different questions, the IVR then transfers the call to appropriate call center through Genesis. The data contains the call transactions, which are used as the source for the populating the data warehouse. The data is also generated through web usage by any user. Hadoop Team is getting daily feeds in form of csv files which are stored on HDFS, and then pig jobs are run on that data for ETL. After which Tableu is used to graphical analysis.
Role:
• Handled a team of 5 members.
• Major role in developing the cluster architecture.
• Optimizing the tuning parameters of cluster.
• Resolving job performance issues.
• Implementing security in cluster using Kerberos.
• Cluster set up for Hadoop and queue monitoring.
• Implementation of various tools and frameworks in the Hadoop.
• Implementing the new Yarn framework in Hadoop.
• Used Sqoop to transfer data between Hadoop and other relational databases

Tools & Technology:
Linux, Hadoop, HUE, PIG, Ambari, Sqoop, Kerberos, Tableau

### Consultant
Genpact Headstrong
June 2015 to May 2016

### Sr. Hadoop Admin
Micron Technologies
February 2014 to April 2015

Description: Micron is a world leader in Semi-Conductors. Data is generated in the process of forming a semi-conductor chip from wafers, which is finally stored in csv files. Hadoop Team is getting daily feeds in form of csv files which are stored on HDFS, and then hive jobs are run on that data for manipulations. After which Neo4j is used to graphical analysis.
Role:
• Major role in developing the cluster architecture.
• Optimizing the tuning parameters of cluster.
• Resolving job performance issues.
• Implementing security in cluster using Kerberos.
• Cluster set up for Hadoop and queue monitoring.
• Implementation of various tools and frameworks in the Hadoop.
• Implementing the new Yarn framework in Hadoop.
• Used Sqoop to transfer data between Hadoop and other relational databases

Tools & Technology: Linux, Hadoop, Hive, Ambari, Sqoop, Kerberos, Neo4j

### Systems Engineer
TCS Ltd

March 2011 to April 2015

## Team member

Deutsche Bank
October 2012 to February 2014

Description: FIM services include the extraction of data, the validation of the data according to data quality measures, translation of the data in accordance with agreed principles. Data feeds were received directly from the sources, CRES and FCL. The received data was injected to desired processing within FIM to ensure it is suitable and available for the processes in the FDW staging area.
Role:
• Hadoop Administrator role includes CDH4 Configuration, Cluster Planning, Installation & Administration
• Hands on experience in Cloudera Manager, Hadoop Installation & monitoring workflow setup using Oozie
• Involved in Service Nodes (NN, SNN, Job tracker) Configuration, Commissioning and Decommissioning of the Nodes.
• Solving problems like under replicated blocks, task tracker failure, job failure and Data Node failure.
• File formats like csv were inputs to HDFS
• Cluster size monitored - 60 nodes in production

Tools & Technology:
Linux, Hadoop, HIVE, Cloudera Manager, Oozie

## Team member

AXA
March 2011 to September 2012

Description: This is a data migration project for development as well as production support, based on onsite/offshore delivery model. Basically, the goal of this data warehouse to track the device and asset related data of the customer. The data is of various types of securities like sales, customer details, accounts, regions, scenarios, loans etc. This Data is loaded into Siebel base tables after applying business logic to the raw data received from ODS using Informatica. This also includes monitoring and fixing production bugs, creating new Informatica mappings, scheduling of workflows for loading new data.
Role:
• Handling of all kind of Development as well as Support activities.
• Conducting status meeting with client using traceability matrix.
• Designed the ETL processes using Informatica tool to load data from File system, and Oracle into the target Oracle 11g database.
• To design Informatica mappings by using various basic transformations like filters, routers, source qualifiers, lookups etc. and advance transformations like aggregators, normalizers, sorters etc.
• Developed 16 mappings/sessions including complex mappings using Informatica power center for data loading.
• Used Informatica Workflow Manager & Monitor to Create, Schedule, and Monitor and Send the Messages in case of process failures.
• Monitoring the daily scheduled jobs running.

Tools & Technology:
Informatica, PL/SQL, Unix

## Education

**B. Tech in UPTU**
KEC College

Central Board of Secondary Education

## Skills

Hadoop (5 years), HADOOP (5 years), APACHE HADOOP SQOOP (4 years), SQOOP (4 years), PYTHON (3 years)

## Additional Information

Technologies: Big Data, Hadoop - Cloudera, Hortonworks, Data Warehousing, Data Enginneering
Programming Languages SQL, Shell scripting, Python
Tools Cloudera Manager, Ambari. Ansible, Docker, Toad, Sqoop, Jenkins, NIFI, Alluxio, ELK Stack, Grafana, Graylog, Presto, Tableau
Domain Experience Banking and Finance, Marketing and Hi-Tech
Databases Oracle, No-SQL Databases