Wameedh Hammadi

Big Data Engineer

Industry Equivalency

4	4	4	3	3
Hadoop	Scala	Spark	Kafka	MongoDB

Equivalency in months

ABOUT ME

I am a Big Data Engineer. I graduated from the University of Surrey, UK with a Master's Degree in Communications Networks and Software. During my studies, I had C++, Java, MATLAB, Web Services and Network Virtualization projects. Since then, I discovered that software development was my main area of interest and decided to pursue that passion at the earliest opportunity.

During my time at Revature, I developed outstanding skills in MySQL, Python and Scala and also had a deep exposure to Apache frameworks such as Hadoop, Spark, Kafka and Airflow. Also worked on multiple projects utilizing AWS EC2 and Cloudera Hortonworks Sandbox virtual machines and collaborated within big teams to perform ETL process, analysis and visualization of Big Data.

Between September 2022 and April 2023, I joined Cognizant as a Software Engineer. Initially, I was trained on Informatica PowerCenter, and upon my completion of two case study assessments pertaining to slowly changing dimensions types 1 and 2, I was selected as a Data Analyst for a project with Novo Nordisk (NNI) in November of 2022. I also completed multiple Udemy courses in that period covering a variety of big data technologies such as Qlik Sense, Tableau, Data Warehousing, Scala, Python and SQL. In January, I was granted access to the NNI commercial data warehouse which enabled me to document and learn about all the NNI databases and the associated schemas and tables stored in Snowflake. Furthermore, I was exposed to the Informatica PowerCenter workflows and locations of source and target files in S3 bucket and became familiar with all the parties involved in the generation and consumption of the data. Additionally, I learned about Veeva CRM system with a focus on its back-end datasets.

Education & Certification

Master's Degree - Networking and Telecommunications

- Nov 2010

QlikSense Visualizations Masterclass

Udemy

Issued On : Dec 22, 2022

Online Transcript :
https://www.udemy.com/certificate/UC-d39cbaa1-ea4b-4036-a5b1-afa80927be27/

Certificate in Qlik Sense Analytics Development

Udemy

Issued On : Dec 23, 2022

Online Transcript :
https://www.udemy.com/certificate/UC-294ac808-ef00-4ec0-8424-1a543996a1bd/

Informatica PowerCenter - Beginner to Expert Level

Udemy

Issued On : Oct 21, 2022

Online Transcript :
https://www.udemy.com/certificate/UC-482f9aab-e34e-4463-a41f-3bb9213cd022/

Snowflake[A-Z] Zero to Hero Masterclass(Core+SQL+API+Visual)

Udemy

Issued On : Dec 17, 2022

Online Transcript :
https://www.udemy.com/certificate/UC-565803c7-da80-414a-ac8b-7d6006cc9d2e/

The Python Bible™ | Everything You Need to Program in Python

Udemy

Issued On : Nov 27, 2022

Online Transcript :
https://www.udemy.com/certificate/UC-03155748-4cb7-4eb6-94d0-99e0d4852ba8/

Microsoft SQL from A to Z

Udemy

Issued On : Dec 20, 2022

Online Transcript :
https://www.udemy.com/certificate/UC-5ef5fbcc-34f7-4a54-8099-0058db2dee33/

The Complete SQL Bootcamp: Go from Zero to Hero

Udemy

Issued On : Dec 21, 2022

Online Transcript :
https://www.udemy.com/certificate/UC-79c737b1-d1b9-4134-bdd6-b302de3449a3/

SQL Programming Basics

Udemy

Issued On : Oct 8, 2022

Online Transcript :
https://www.udemy.com/certificate/UC-4fae82b7-b6de-4620-a687-3aafcb511dcd/

Data Warehouse Fundamentals for Beginners

Udemy

Issued On : May 22, 2023

Online Transcript :
https://www.udemy.com/certificate/UC-e41e012d-89b7-4583-9fa6-f5fa3e59655f/

Scala Applied, Part 1

Udemy

Issued On : Nov 21, 2022

Online Transcript :
https://www.udemy.com/certificate/UC-d7c2a288-c120-4826-ba77-c95eccff6843/

Scala Applied, Part 2

Udemy

Issued On : May 22, 2023

Online Transcript :
https://www.udemy.com/certificate/UC-341ecd5d-5cfd-4713-a68a-6ca799c0d048/

Scala Applied, Part 3

Udemy

Issued On : Dec 1, 2022

Online Transcript :
https://www.udemy.com/certificate/UC-626878b2-594d-46db-9df5-0c70397ecfe3/

Tableau 2022 A-Z: Hands-On Tableau Training for Data Science

Udemy

Issued On : May 22, 2023

Online Transcript :
https://www.udemy.com/certificate/UC-498b7d4d-40e7-4a73-8214-5721369e131f/

Skill Matrix

Data Storage

- RDBMS (4)
- NoSQL (4)
- MySQL (4)
- MongoDB (3)
- HiveQL (4)

Languages

- Scala (4)
- Python (5)
- C++ (3)
- MySQL (4)
- Java (3)

Big Data

- Kafka (3)
- Hadoop (4)
- Hive (4)
- Apache Spark (4)
- HDFS (4)

Data Storage

- RDBMS
- NoSQL
- MySQL
- MongoDB
- HiveQL

Languages

- Scala
- Python
- C++
- MySQL
- Java

Big Data

- Kafka
- Hadoop
- Hive
- Apache Spark
- HDFS
- Qlik Sense
- Informatica
- Snowflake

Apache Spark

- RDD's (4)
- Dataframes (4)
- Datasets (4)
- Spark SQL (4)
- Hive Context (4)

Hadoop

- Hadoop (4)
- YARN (3)
- HDFS (4)
- Hive (4)
- MapReduce (4)

Other Technologies

- AWS EMR (3)
- GitHub (3)
- Linux (4)
- Tableau (3)
- Hortonworks Sandbox (4)

Apache Spark

- RDD's
- Dataframes
- Datasets
- Spark SQL
- Hive Context

Hadoop

- Hadoop
- YARN
- HDFS
- Hive
- MapReduce

Other Technologies

- AWS EMR
- GitHub
- Linux
- Tableau
- Hortonworks Sandbox

Data Warehousing

- **Data Warehousing (3)**
- **Apache Airflow (3)**
- **OLTP and OLAP (4)**
- **Executors (4)**
- **Dimensional Modelling (3)**

Data Warehousing

- Data Warehousing
- Apache Airflow
- OLTP and OLAP
- Executors
- Dimensional Modelling

Work Experiences

Cognizant Technology Solutions — **Sep 2022 - Apr 2023**

Software Engineer

Cognizant AIA / Novo Nordisk — **Sep 2022 - Apr 2023**

Role : Data Analyst

Responsibilities

- After I joined Cognizant AIA in September 2022, I completed multiple Udemy courses covering Python, SQL, Scala, Spark, AWS, Tableau, Qlik Sense, Informatica PowerCenter, Data Warehousing and other Big Data technologies.
- After that, and upon passing two case study assessments pertaining to data warehousing and the implementation of SCD types 1 and 2 in Informatica PowerCenter, I was selected for a

project with Novo Nordisk as a Data Analyst in November 2022.

- Received the work laptop and was granted access to Novo Nordisk Snowflake database in January 2023.
- Documented and got exposed to NNI databases, schemas, tables and the associated Informatica PowerCenter workflows.
- Became familiar with all the parties involved in data production and consumption as well as the storage locations of source and target data files.
- Was exposed to the process of requirements gathering and distribution along with all the associated discussions.
- Learned about Veeva CRM system utilized by NNI sales representatives with a focus on its back-end datasets.

Tools / Technologies

Python, SQL, Scala, Spark, AWS, GitHub, Data Warehousing, Informatica, Snowflake and other Big Data technologies.

US Census Data

US Census Data is a Big Data project that analyzes the United States population from 2000 to 2020 every decade. The data contains over 60 million records, is mined into cloud storage, and is queried to answer questions posed by the project manager. The conclusions drawn from the queries suggest future trends, such as population growth for the next decade. The queries were optimized using Spark and Spark SQL and shown visually with a data visualization tool.

Roles / Responsibilities

- Teamed up with members of the cohort following Agile methodology
- Mined US Census Data using web scraping.
- Utilized Apache Airflow for workflow management.
- Stored the data into AWS S3.
- Further data processing and optimization was carried out on an AWS EMR cluster.
- Coded using Scala Spark and Spark SQL to find trends including future population growth.
- Used Tableau to visualize the data and demonstrate trends.
- Prepared PowerPoint slides and showcased my work.

Environment / Technologies

Scala , Spark SQL, Spark, Data Visualization

Capstone Project - eCommerce and Insurance

Two teams (eCommerce and Insurance) will create Scala programs to publish their data to Kafka and consume the other team's data from Kafka. The eCommerce Team will create a Producer program in Scala to generate eCommerce data (for Orders). The team will have to create records on a continuous basis and publish them to a Kafka Topic (simulate a live stream of data). The Insurance Team will create a Consumer program to read data from the eCommerce Team's topic, process the data and run some analysis on the data, which will them be used for visualizations. Similarly, Insurance Team will create a Producer program in Scala to generate Insurance Claim data. The team will have to create records on a continuous basis and publish them to a Kafka Topic (simulate a live stream of data). The eCommerce Team will create a Consumer program to read data from the Insurance Team's topic, process the data and run some analysis on the data, which will them be used for visualizations.

Roles / Responsibilities

- Collaborated within a team of ten associates and in partnership with another team of 9 associates using agile methods.
- Created a producer program for insurance data in Scala.
- Utilized Kafka and Zookeeper for proper data streaming over the application.
- Utilized Kafka consumer to ingest messages from a Kafka Topic.
- Deployed jar file to Hadoop environment to run the application and connect to Spark and Kafka.
- Executed SQL like queries on eCommerce data using Scala and Spark SQL.
- Output all analyzed results into a Data Visualization Tool.
- Presented the results of the analysis as a group.

Environment / Technologies

Hadoop, Scala , Kafka, Spark SQL, Spark

Technology Project - Insurance Claim

Work in a team and create a Producer program in Scala to generate Insurance Claim data, which will then be loaded into Hive tables. Another team will also create a similar "producer" client application to generate their Insurance Claim data and load it into Hive. Your team creates a "consumer" application in Scala that will read the data generated by the other team and run some analysis queries on that data using Hive, Spark and SparkSQL. Similarly, the other team will create a consumer program in Scala to read the data your team generated and run some analysis on it Hive, Spark and SparkSQL.

Roles / Responsibilities

- Created a program that loads a CSV file from the local machine to HDFS using Scala.
- Created a program that uses the CSV file in HDFS to build a Hive table using Scala.
- Generated an RDD from the CSV file in HDFS and used it to execute a query on insurance data using Scala and Spark on.
- Created a view of a dataframe and used it to execute a query on insurance data in SQL format using Scala, Spark and SparkSQL.
- Executed a HiveQL query on insurance data using Hive Context in Spark Session using Scala and Spark.
- Installed Hortonworks Sandbox virtual machine and utilized it to run jar files generated by the Scala code.
- Took part in daily standup meetings for my team as well as in meetings with other team we partnered with to discuss roles and responsiblities.
- Took part in the project presentation in which I showcased the execustion of 3 queries.

Environment / Technologies

Scala , Hive, Spark, Spark SQL, Hadoop

Foundation Project - Python MongoDB

A Python CLI (Command Line Interface) application. Data should be parsed from a CSV OR JSON file and persisted to MongoDB. The functionality of the application beyond that is up to the associates. Application Requirements: - Application must perform basic CRUD (Create, Read, Update, Delete) functionality with data stored in a database. - Application must be able to read JSON or CSV files, and store the data contained within those files in a database. - Application should have a CLI where users can interact with the application while it is running. - Other than those requirements, the kind of application you have is up to you. Tech Stack: - Python 3.x - File I/O - Collections - MongoDB - PyMongo - git SCM (+ GitHub).

Roles / Responsibilities

- Created a Python Command Line Interface banking application for user interactions.
- Utilized MongoDB to store data collections parsed from JSON files.
- Utilized PyMongo to access the functionalities of MongoDB in Python.
- The application performed a variety of CRUD (Create, Read, Update and Delete) operations on data.
- Application allows users to perform banking operation such as deposit, withdraw, transfer, close account , past transactions, etc.
- Implemented user input validation and error handling.
- Utilized the object-oriented programmimg concept of encapsulation to avoid code repetition.
- Submitted a MVP (Minimal Viable Product) 2 days prior to the final submission due date.
- Presented my project to a member of the Revature Quality Check team, my trainer and members of my Big Data batch.

Environment / Technologies

Python, MongoDB, PyMongo

Email : wameedh_shamil@yahoo.com