MINISTRY OF EDUCATION AND SCIENCE OF THE REPUBLIC OF KAZAKHSTAN

INTERNATIONAL INFORMATION TECHNOLOGY UNIVERSITY JSC

DEPARTMENT OF COMPUTER ENGINEERING AND INFORMATION SECURITY

**INDUSTRIAL PRACTICE REPORT**

**Education program**

**6B06108 – Data Science and Machine Learning**

Done by: Adilbek Spatayev \_\_\_\_\_\_\_\_\_\_\_\_\_\_

«12» July 2022 (signature)

Head of practice: Balgabek A.A. \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

«\_\_\_ » \_\_\_\_\_\_2022 (signature/assignment)

Almaty 2022

CONTENT

1. Introduction------------------------------------------------------------------------1
2. Week 1 report----------------------------------------------------------------------2
3. Week 2 report----------------------------------------------------------------------3
4. Week 3 report----------------------------------------------------------------------4
5. Week 4 report----------------------------------------------------------------------5
6. Week 5 report----------------------------------------------------------------------6
7. Conclusion-------------------------------------------------------------------------7
8. References--------------------------------------------------------------------------8

List of terms and abbreviations

SQL - Structured Query Language

PC – Personal Computer

PL/SQL – Procedure Language / Structured Query Language

ETL - Extract, Transform, Load

INTRODUCTION

1

Week 1 report

Day 1, 06.06.22

It all started with the collection of documents for the start of the internship, namely a trip to the clinic. Then a conversation with the staff and familiarization with the workplace and acquaintance with the project.

Day 2, 07.06.22

I created requests for connecting my PC to databases, to the Informatica PowerCenter program, then after that I studied the documentation on development standards.

Day 3, 08.06.22

First login to the PL/SQL Developer program to start learning SQL. The training began with understanding the database, what tables exist. After that, I was given a small list of tasks to get acquainted with my level of knowledge of SQL

Day 4-5, 09.06.22-10.06.22

I was given an assignment to read and remember the basics of SQL at w3school.com

2

Week 2 report

Day 1-4, 13.06.22-16.06.22

Solving SQL problems on the site https://www.sql-ex.ru/ levels 1-2

Day 5, 17.06.22

Received feedback on the tasks, and worked on fixing bugs and improving my SQL skills, which I also did the rest of the weekend

3

Week 3 report

Day 1-3, 20.06.22-22.06.22

I fell under the training period for employees, where there was also SQL training with lectures and practice, where the intricacies of each operator were discussed, where and by what methods to solve problems, as well as the use of more interesting window functions for a more flexible solution.

Day 4-5, 23.06.22-24.06.22

Upon completion of intensive SQL training, Informatica PowerCenter training began. Lectures were held to familiarize with the program and explain where and how it is applied.

4

Week 4 report

Day 1-3, 27.06.22-29.06.22

Upon completion of the Informatica PowerCenter theory training, we proceeded to the practical tasks of building ETL processes on real (simple) tasks, where it was necessary to apply such simple blocks as Filter, Sorter and Expression.

Day 4-5, 30.06.22-01.07.22

We started to study more serious problems like Source Qualifier, Join Transformation, Aggregator Transformation and Union Transformation. There were some difficulties with these transformations due to the specifics of their work.

5

Week 5 report

Day 1-2, 04.07.22-05.07.22

We were further trained to check the data for accuracy, at first there was a simple reconciliation with the script using operators such as MINUS and FULL JOIN, and bringing all our previously created ETL processes into a common standard form that I read earlier in the first week

Day 3-5, 06.07.22-08.07.22

After validating the data with Excel, we explored functions such as ВПР, СУММА, РАЗНИЦА and using pivot tables for more precise data validation.

6

CONCLUSION

In 5 weeks of practice, I learned a lot of technologies in order to be a Data Engineer, which I can say for sure, I was very lucky that I fell under the personnel training period. Each training went quickly and as productively as possible, which accurately explained to me that specialists in their field work here. Learning technologies such as writing SQL reports, building ETL processes and eventually data validation gives me the opportunity to get a job in other organizations.

7

REFERENCES

<https://www.sql-ex.ru/>

<https://www.w3schools.com/sql>

<https://docs.informatica.com/>

8