Sabai Code

SabaiCode

DATA ENGINEERING BOOTCAMP

align="left" style="color: #FFFFFF; font-size: 13px;
letter-spacing: 1px;" class="title_color">
<!-- ======= section text ====== -->
<div style="line-height: 22px">

ABOUT SABAICODE

SabaiCode is a leading institution specializing in technology education. Led by technology professionals and forethought thinkers, SabaiCode has partnered with the nation's leading telecom company and trained many students who went on to become UX/UI engineers, front-end and back-end engineers at many top tier local and regional companies. Many local and international NGOs rely on SabaiCode to train and direct their beneficiaries to become future builders and innovators. SabaiCode currently offers education programs in web programming, robotics, and data science to students of all ages.

Vision

To develop Cambodian leaders in technology and innovation.

Mission

To equip every Cambodian child with powerful technology skills.

ABOUT THE PROGRAM

OVERVIEW

Data Engineering at **SabaiCode** focuses on helping students to acquire the essential data engineering skills, gain professional experience, obtain solid knowledge in data analysis and machine learning and to become well prepared for careers in data engineering.

WHO IS THIS FOR?

This weekend program is suitable for students who want to become a Data Engineer.

PREREQUISITE: PASSION FOR DATA AND TECHNOLOGY

THIS PROGRAM IS IDEALLY FOR:

Data Scientist / Data Related Field

who want to build more advanced skills in Data Engineering and model deployment.

Recent Graduates

From computer science or engineering field who want to gain advantage in the job market by gaining practical ML Engineering skills

Entrepreneurs

who want to build ML systems and platforms for his/her company.

Career Switchers

who have a full-time job and want to switch their careers to be Data Engineers.

PROGRAM DETAILS

24 WEEKS

Saturday & Sunday full day
8h per day

4+ Instructors

To learn from

8 Modules

To study

16+ hand-on projects

To work together and/or individually

4+ Capstone projects

To get familiar with real-world projects

PROGRAM OUTCOMES

- Gain a solid understanding of the fundamentals of the Python language, its tooling, and the development process
- Understand descriptive statistics and data visualization techniques
- Understand supervised and unsupervised machine learning algorithms
- Deep understanding of the pros and cons of different database systems
- Solid foundation of database concept and fundamental of SQL
- Solid understanding of major big data and cloud platforms such as Hadoop, Spark, Databricks and Kafka
- Hands-on experience with data warehouse modeling and ETL
- Able to build and deploy data pipelines on cloud platforms using workflow orchestration tools such as Apache Airflow

ADMISSION PROCESS

Application Process

The application process consists of three simple steps. An offer of admission will be made to the selected candidates and accepted by the candidates upon payment of the admission fee.

01 Submit Application

Complete the application and include a brief statement of purpose. The latter informs our admissions counselors why you're interested and qualified for the program.

02 Application Review

A panel of admissions counselors will review your application and statement of purpose to determine whether you qualify for acceptance.

03 Admission

An offer of admission will be made to qualified candidates. You can accept this offer by paying the program fee.

Course Highlight

1. **Exploratory Data Analysis**

   Analyze and investigate data sets and summarize their main characteristics, often employing data visualization methods.

2. **Python Programming for Data Engineer**

   Kickstart your learning of Python for Data Preprocessing and Data Engineering.

3. **Data Cleaning and Exploration with ML**

   Core concept, tools, and fundamental for Data Cleaning and Exploration with Machine Learning.

4. **Database and Big Data**

   Introduce database concepts and fundamental knowledge of SQL language.

5. **ETL, And Data Pipeline with shell, Airflow and Kakfa**

   Apply ELT and ETL process, move data through data pipeline and store data in destination systems.

6. **Getting Started with Data Warehouse and BI Analytics**

   Describe how data warehouse serves as a single source of data and analytics with BI applications.

7. **Introduction to Big Data with Spark and Hadoop**

   Characteristics of Big Data, and Literacy with spark and Hadoop

8. **Data Engineer Capstone Project**

   Apply a variety of data engineering skills and techniques you have learned so far with the real-world use case.

✓ Quizzes

✓ Moderated Discussion Boards

Q&A Sessions with Course Leaders

✓ Application Projects

1. Exploratory Data Analysis

Description

Exploratory data analysis (EDA) is used by data scientists to analyze and investigate data sets and summarize their main characteristics, often employing data visualization methods. It helps determine how best to manipulate data sources to get the answers you need, making it easier for data scientists to discover patterns, spot anomalies, test a hypothesis, or check assumptions.

Learning Objectives

- Be able to understand basic statistics for data analysis
- Can demonstrate and discuss various stage in data preprocessing
- Be able to do data cleaning and exploration

Course Curriculum

- Session 1: Examining the distribution of features and targets. Examining bivariate and multivariate relationship between feature and targets
- Session 2: Review on descriptive statistics
- Session 3: Identifying and fixing missing values
- Session 4: Encoding, Transforming and Scaling Features
- Session 5: Feature Selection
- Session 6: Preparing for Model Evaluation

2. Python Programming for Data Engineer

Description

The Python programming course guide you how to write and understand the basic of python programing include: variable, data type, control flow and function of python. In addition, this course provides you to understand and ability to use some library in python, for example, NumPy, Pandas and matplotlib that can help you to analyses and visualize some foundation of statistical data. This course helps you to start and become the proficient in python programing for data analyses, data engineer as well as data science.

Learning Objectives

- Understand the concept of structure of python programming
- Understand the control flow in python
- Identify the python library for basic data analyze

Course Curriculum

- Section 1 – introduction to python environment for data engineer and data science
- Section 2 - introduction to python programing structure
- Section 3 - data type, variable and sequential programing
- Section 4 - data structure includes: list, tuple, dictionary and set
- Section 5 - control flow
- Section 6 - function
- Section 7 - NumPy
- Section 8 - Pandas
- Section 9 - Visualization
- Capstone project: Missing value and outlier detection

3. Exploration with Machine Learning

Description

Machine Learning is one of the most in-demand skills for jobs related to modern AI applications, a field in which hiring has grown rapidly for the last decade. This Bootcamp is intended for anyone interested in related developing skills to Data Engineer and experience to pursue a career in Machine Learning engineer and leverage the data exploration with machine learning. It also complements your learning with special topics.

Learning Objectives

- Understand basis algorithms of machine learning
- Be able to apply machine learning to explore data

Course Curriculum

- Session 1: Preparing for Model Evaluation
- Session 2: Linear Regression Models
- Session 3: Support Vector Regression
- Session 4: K-Nearest Neighbors, Decision Tree, Random Forest, and Gradient Boosted Regression
- Session 5: Logistic Regression
- Session 6: Decision Tree and Random Forest Classification
- Session 7: K-Nearest Neighbors for Classification
- Session 8: Support Vector Machine Classification
- Session 9: Naïve Bayes Classification
- Capstone Project: Case Study

4. Database and Big Data

Description

The course database and big data will guide you how to use excel extract data from database (DB) and performance analysis on that DB incase for relational DB. In addition, you will learn how to use SQL to extract and analyze the data stored in databases, at the same time, NoSQL database will be guided for you in this course as well by applying Mongo DB. Finally, the foundation of data warehouse and big data by using Hadoop to ingest structured and unstructured.

Learning Objectives

- Understand the relational database
- Recognize the SQL and no-SQL database
- Identified the warehouse
- Understand the data ingestion to warehouse as well as big data tool

Course Curriculum

- Section 1 – introduction to relational data base
- Section 2 - identify relationships and retrieving data
- Section 3 - Overview and History of NoSQL Databases
- Section 4 - Comparison of relational databases to new NoSQL stores, MongoDB
- Section 5 - NoSQL Key/Value databases using MongoDB
- Section 6 – Data warehouse vs data base
- Section 7 - Data warehouse architecture (ETL vs ELT)
- Section 8 - Data warehouse testing tutorial (Airflow)
- Section 9 – Introduction to big data with Hadoop
- Capstone project: Data pipeline using Airflow with structure data

5. ETL with Shell, Airflow, and Kafka

Description

Extract, Transform and Load, in other word ETL, processes are applied for cases such as flexibility, speed, and scalability of data are important. You will learn the key different process between ETL and ELT which include: (i) the place of transformation, (ii) flexibility, (iii) Big Data support, and (iv) time-to-insight. In addition, you will learn that there is an increasing demand for access to raw data that drives the evolution from ETL to ELT. Data extraction involves advanced technologies including database querying, web scraping, and APIs. Moreover, students will also learn that data transformation is about formatting data to suit the application and that data is loaded in batches or streamed continuously.

Learning Objectives

- Describe what an ETL process is
- Explain what data loading means
- Describe why ELT is an emergent trend
- Describe the trending shift from ETL to ELT
- Summarize data extraction techniques
- Name data transformation techniques
- List ways information can be “lost in transformation”
- Summarize data loading techniques
- Differentiate batch loading from stream loading
- Contrast ETL and ELT.

Course Curriculum

- Section 1 – introduction to ETL and ETL processes
- Section 2 - introduction to data transformation techniques
- Section 3 - Linux/window command

- Section 4 - ETL Techniques with shell scripting
- Section 5 – Airflow overview
- Section 6 – building data pipeline by using Airflow
- Section 7 – Airflow Kafka
- Section 8 - building data pipeline for streaming data by using Kafka
- Capstone project: Building ETL data pipeline with Airflow and Kafka

6. BI Analysis with Power BI and Tableau

Description

This course will enable you to develop an understanding of the vast amount of data that is available to organizations, and teach you the skills to access, prepare, analyze, and visualize this data to support decision-making, solve business problems, and remain competitive. This course is heavily based on hands-on activities, providing you with practice implementing data analytic techniques and using tools for business intelligence. The focus is on techniques and tools that can be used be used by individuals in an organization to gain insight into complex business problems. The techniques that will be used are extended data analysis and data visualization. These analytics techniques will be supported with applications such as MS Excel, Power BI and Tableau.

Learning Objectives

- Recognize business problems that can be addressed with Business Analytics tools
- Get familiar with overall business analytics concepts, and descriptive analytics techniques
- Develop strong modeling skills in Excel and Power BI
- Learn about data visualization concepts and select appropriate data visualization techniques
- Apply tools to visualize data, including Tableau, Excel, and Power BI

Course Curriculum

- Section 1 – BI introduction
- Section 2 - Role of Data Importance of BI
- Section 3 - Data Manipulation and Analysis with Excel
- Section 4 - Data Modeling using Microsoft Power BI
- Section 5 – Data Analytics Lifecycle
- Section 6 – Data Visualization with Microsoft Power BI
- Section 7 – Data Analysis with Tableau
- Section 8 - Data Visualization with Tableau
- Capstone project: Building Data Visualization with Power BI and Tableau

7. Data Warehouse, Big Data with Hadoop

Description

This course is design to deliver you the knowledge of Data Warehousing principles, Data Warehouse techniques and big with Hadoop application. The course introduces the topics of Data Warehouse design, Extract-Transform-Load (ETL), Data Cubes, Data Marts and big data with Hadoop. You will gain in-depth knowledge of the Big Data framework using Hadoop and Spark. In this hands-on Hadoop course, you will execute real-life, industry-based projects using Integrated Lab.

Learning Objectives

- Describe architecture and methods for storage and provision of enterprise data.
- demonstrate competency in data modeling, including dimensional modeling.
- Learn how to navigate the Hadoop ecosystem and understand how to optimize its use
- Ingest data using Sqoop, Flume, and Kafka.
- Implement partitioning, bucketing, and indexing in Hive
- Work with RDD in Apache Spark
- Perform DataFrame operations in Spark using SQL queries

Course Curriculum

- Section 1 – introduction Data warehousing
- Section 2 - Data Warehousing Design
- Section 3 - Introduction to Bigdata and Hadoop
- Section 4 - Hadoop Architecture Distributed Storage (HDFS) and YARN
- Section 5 – Data Ingestion into Big Data Systems and ETL
- Section 6 – Distributed Processing MapReduce Framework and Pig
- Section 7 – Apache Hive
- Section 8 - Basics of Functional Programming, Scala and Spark
- Section 9 - Spark Core Processing RDD
- Section 10 - Spark SQL Processing DataFrames
- Capstone project: Building big data framework with Hadoop and spark

8. Data Engineering Capstone Project

Description

In this course you will apply a variety of data engineering skills and techniques you have learned as part of the previous courses in this Engineering BootCamp. You will assume the role of a Junior Data Engineer who has recently joined the organization and be presented with a real-world use case that requires a data engineering solution.

Instructor Profiles

Mr. SUOM Sareoun

Mr. SUOM Sareoun graduated with a bachelor's degree in Computer Science and a bachelor's degree in Mathematics from the Royal University of Phnom Penh. He has 17 years of experience in developing software; currently he is a VP of Platform at Mangomap which ranks in the top 5 web mapping platforms. Also, he is an R&D Director at Z1 Data, a data driven company. At Z1 Data he uses big data to build API and AI to support and predict property prices and provide services related to real estate. He is a cofounder and director at SabaiCode, a technology school focused on building the next generation of human capital in Internet of Everything, Computer Software and Data Science.

Mr. CHAN Sophal

He has many years of experience in researching of AI, ML and DL development. Currently, he is working as a lecturer in a government institution and private institutions where he is teaching data science in Bachelor of Engineering and Master of Engineering programs. In addition, he has been involved in AI research and studies in different countries including Japan and India. He was a visiting professor to study and discuss data science curriculum in Paris, France. He is a PhD candidate in medical image analysis with deep learning application. He also has a great interest in IoT, robotic, AI with real estate, AI with medical image and remote sensing research and application. Mr. Sophal graduated with MSIT (Master of Science in information technology) from Prince Songkla University, Phuket, Thailand. He is doing his PhD in data science at Prince Songkla, Hadyai, Thailand.

Dr. PHAUK Sokkhey

Dr. PHAUK Sokkhey graduated with a bachelor's degree in Mathematics from the Royal University of Phnom Penh, a master's degree in Applied Mathematics from Suranaree University of Technology, Thailand. Dr. Sokkhey obtained his Ph.D. in Interdisciplinary Intelligent Systems majoring in Data Science from University of the Ryukyus, Japan. Dr. Sokkhey is a head of the Master's Program in Data Science at Institute of Technology of Cambodia (Techno), and a Data Science coach at SabaiCode. In ITC, he teaches courses in Applied Statistics, Exploratory Data Analysis, and Data Science. Dr. Sokkhey has published 10+ research papers in the fields of Data Science, Machine Learning, and Data Analytics.