Assignment #1, Problem Statement #1
CSCI 5408 - Data Management, Warehousing, Analytics

I, Adesh Nalpet Adimurthy, declare that in assignment 1 of CSCI 5408 course, data scrapping is not done programmatically or using any online or offline tools. However, the webpages within dal.ca domain are visited manually, and some useful information is gathered for education purpose only. Information, such as email, personal contact numbers, or names of people are not extracted. The course instructor or the Faculty of Computer Science cannot be held responsible for any misuse of the extracted data.

1.1 Problem Statement (in-short)**:** Find entities in dal.ca and come up with a scalable ERD/EERD.

2.1 Entities discovered within Dalhousie University website:

Sl no	**Entity**	**Reason for Considering**	**Source**
1.	News	News is a Strong Entity, and News exists across sectors in Dalhousie University. Every item represents a piece of information. Attributes: unique-id (primary-key), date (historical data), title, and description.	https://www.dal.ca/news/news.html
2.	Programs	Program is a Strong Entity, and Dalhousie University offers a variety of programs under different Departments. For example, MACS is a program. Attributes: unique-id (primary-key), department (some programs can be across Departments but has the main Department), type (undergraduate, graduate, professional), faculty, campus, duration, title, and description.	https://www.dal.ca/academics/programs.html

3.	Faculties and Departments	Faculty is a Strong Entity and has a one-to-many relationship with Departments. Dal primarily has 13 faculties with several departments under each faculty (ex: Medicine is a Faculty and Bioethics, Family Medicine, Emergency Medicine, etc. are departments under Medicine)	https://www.dal.ca/academics/faculties.html
4.	Events	Event is a Strong Entity, and Dal has several use cases for Events, because of which sub-types might be necessary to include specific information of the sub-type. Academic Events is one such example, which will have the attributes: department type, program, and level (undergraduate and graduate). Every row in an Event represents an event or information for the given date & time. Attributes: unique-id (primary-key), title, date & time.	https://academiccalendar.dal.ca/Catalog/ViewCatalog.aspx?pageid=viewcatalog
5.	Courses	Course is a Strong Entity, and Dal offers students to opt for a set of courses under each program. Programs have a one-to-many relationship with courses (and across departments). Attributes: course code (primary-key), title, program-id, description.	https://academiccalendar.dal.ca/Catalog/ViewCatalog.aspx?pageid=viewcatalog&catalogid=112&chapterid=7023&topicgroupid=30909&loaduseredits=False
6.	Time Table	A Time Table is a structured schedule of events with the time at which they occur. Time Table is a Strong Entity. While Time Table	https://events.dal.ca/ https://www.dal.ca/

		might seem similar to a Calendar or can be force-fit into a calendar, a time-table is a lot more specific and has other attributes such as repeat. Example: CSCI 5408 is on Tuesday and Friday from September to December at 10:15 AM.	academics/academic_timetable.html
7.	Books & Library	Book is a Strong Entity, and Dal has multiple libraries across campuses. Having an explicit entity for Library (Strong Entity) also makes sense as Dal has several libraries across campuses. Book entity can be a lot more abstract and include books, journals, articles, and publications (sub-types).	https://libraries.dal.ca/ https://www.dal.ca/faculty/law/research/publications.html
8.	Rooms	Room is a weak entity (physical rooms in a library). Dal has study areas and rooms across libraries that can be booked by Users. Note: A user can book n number of rooms, but n users cannot book the same room (they can be in the same room, but only one user can book it). Attributes: room number, location, campus availability, capacity.	https://roombooking.library.dal.ca/
9.	Awards	Award is a Strong Entity, and each row represents a unique award offered at Dal. Awards can be given to teams, hence has a many-to-many relationship with users. To identify users coupled to the award, Award-User (many-to-many).	https://www.dal.ca/news/2021/04/09/2020-21-black---gold-athletic-awards.html

10.	Users	User is a Strong Entity and contains a broad set of sub-types, including students (prospective, current, and alumni), faculty members, etc. Hence, the User is the base entity and has sub-types.	https://www.dal.ca/ Current Students, Faculty, and Staff in the menu.
11.	Applications	Application is a Weak Entity. Every term, a sub-set of the users apply for various programs and certifications. Application has total participation with users as it's a weak entity. Therefore, every row represents an application to a program, job, loan, etc. Attributes: application-id, status, pre-requisites, and other application details for sub-types.	https://www.dal.ca/academics/programs/graduate/applied-computer-science/admissions.html https://dalonline.dal.ca/PROD/bwskalog.P_DispLoginNon
12.	Loans & Accounts	Loans and Accounts are individual entities. Here Account refers to a ledger. Loan is a strong entity and an "Approved loan" to a user is a Weak Entity and has total participation with Users - Relationship is <<borrows/pays>>. Attributes: Loan name/ID, max amount, validity, pre-requisite.	https://www.dal.ca/admissions/money_matters/awards-financial-aid/student_loans.html https://www.dal.ca/admissions/money_matters/tuition_payments/Check_Balance.html
13.	Jobs	Job is a Strong Entity, and Dal has a platform to inform about a variety of jobs on-campus and off-campus. Every row represents a job detail. Attributes: unique-id (primary-key), title, description,	https://mycareer.dal.ca/home.html

		location	
14.	Communities /Clubs	Communities/Culb is a Strong Entity and represents various clubs and societies in Dal. A user can join any number of clubs. Attributes: unique-id (pk), name, description, capacity.	https://www.dal.ca/campus_life/career-and-leadership/career-leadership-programs/DALConnects/community-partners.html
15.	Scholarships	Scholarship is a Strong Entity. Dal offers various scholarships to students and faculty members and is often reserved for a specific set of people. Attributes: scholarship unique name (pk), title, description, validity (start and end date), criteria.	https://www.dal.ca/admissions/money_matters/awards-financial-aid/scholarships.html
16.	FAQs	FAQ is a Strong Entity, and Dal's website has a FAQ section across the application. Every row represents an FAQ, a question, and an answer. Attributes: unique-id (pk), question, answer	https://www.dal.ca/dept/hr/careers_dalhousie/faqs-cbx.html
17.	Donations	Donation is a Strong Entity. Dal receives donations from various sources for the betterment of the community. And the sources are not confined to users within Dal. Every item in Donations represents a donation paid to the University. Attributes: unique-id, from, to (departments or	https://alumni.dal.ca/news-stories/meet-dals-donors/

		groups within Dal), Amount.	
18.	Services	Services is a Strong Entity. Dal offers a variety of services to Students and Faculty members. One such example is Academic Support to students, including Writing Support, Advising, Study Support, and Tutoring. Attributes: unique-id (pk), service type, title, description.	https://www.dal.ca/campus_life/academic-support.html

2.2 Note: There are 20 entities in the above table (without sub-types); some related entities are included and detailed under the same block. The attributes mentioned above are not a comprehensive set of attributes but rather an indication of the top ~4 attributes.

3.1 Initial data modeling - Chen Model:

PDF: https://drive.google.com/file/d/1LbYQjlwWNCwgS4LqfskKHX90K-HttM3u/view?usp=sharing

4.1 Issues in Initial Chen Model (250+ words):

Initially, the Application did not have sub-types and represented applications to various programs, which was incorrect, and the Application is a lot more generic. For example, it can have sub-types, which include job applications. In addition, Loan and Student Account were under the same entity; Introducing a new entity for Loans with the relation "opts for/applies" solved the issue. The Donation was a weak entity and was strongly coupled to the User entity; however, after analyzing various donation pages, the donation source can be anonymous and is corrected to exist independently. Service entity had no relationship with other entities, but most departments and even libraries had a similar structure of information to display a service they offer. Hence, other entities have a one-to-many relationship with services, thereby improving the re-usability. Club/Community entity did not consider the association with members. Each club has a sign-up form and hence has a relationship with the user (one-to-many) as every user can join multiple clubs. The event entity did not have a relationship with other entities; most departments and libraries often conduct events. Therefore, introducing a one-to-many relationship with the department and library solved the issue. Similarly, the news entity is published by other departments and clubs, thereby introducing a relationship is valid. The loan was a weak entity; however, considering loans are pre-defined and approved loans belonging to a user can be a weak entity, but a Loan by itself is corrected to be a strong entity. Finally, several attributes and primary keys have been modified or added to the entities based on the details mentioned on the Univeristy website.

Note:

1. FAQ is a generic entity with a type to determine the category and can have a relationship with most entities. Most pages in the University have a FAQ section and hence, not explicitly indicated in the Chen model.
2. There can be several other candidate keys for the primary key, and often a unique generated identifier can also be used. However, certain assumptions have been made about the uniqueness of the candidates. For example, the Library name is a primary key assuming it's unique.

5.1 Correct Data Modelling (EERD):

PNG: https://drive.google.com/file/d/1P4BiP2WSyLw2ntT0dn9SJax_INYryAWU/view?usp=sharing
Draw.io: https://drive.google.com/file/d/1bj6xNuEvt86hhyYwZathy6bRR5LsG4LW/view?usp=sharing

6.1 References:

The tool used for Chel Model: https://draw.io