A green circle with white text

Description automatically generated

**IT5507 – Fundamentals of Data Science – Lab 4.1**

**Data Flow Diagram DFD Analysis**

**Activity 1)**

In the dynamic field of Data Science, the ability to model and represent data processes is paramount for deriving meaningful insights and making informed decisions. As we embark on this activity exploring Data Flow Diagrams (DFDs), we'll draw parallels between foundational data and process modeling concepts and their applications in Data Science. Consider how these models become invaluable tools in organizing and visualizing intricate data structures and workflows, aligning seamlessly with the data-centric nature of the field. Additionally, we'll discuss the relevance of DFDs in the context of Data Science projects, where understanding data flows and interactions is essential for designing effective data pipelines, implementing machine learning algorithms, and ensuring the overall success of analytical endeavors. This activity serves as a bridge between the foundational principles of DFDs and their practical implications in the dynamic landscape of Data Science, providing a holistic perspective for students aiming to apply these skills in real-world scenarios.

**Data and Process Modelling Concepts and Tools:**

* **Data Modelling:** Involves defining and organizing data requirements and structures. Concepts include entities, attributes, relationships, and normalization. Tools include Entity-Relationship Diagrams (ERD) and Data Flow Diagrams (DFD).
* **Process Modelling:** Focuses on representing processes, activities, and workflows within a system. Concepts include processes, data flow, data stores, and external entities. Tools include Data Flow Diagrams (DFD), Unified Modelling Language (UML) Activity Diagrams, and Business Process Model and Notation (BPMN).

**Context Diagram vs. Diagram 0:**

* **Context Diagram:** Provides an overview of a system, showing external entities and the high-level interactions between them and the system. It uses a single process symbol and does not show internal processes.
* **Diagram 0 (Level 0 DFD):** Represents the main processes within a system and their interactions. It provides more detail than a context diagram, showing internal processes and data flows.
* **Symbol not used in a Context Diagram:** The process decomposition symbol (circle with a cross) is not used in a context diagram.

**Levelling DFDs:**

* **Levelling:** Involves creating a series of DFDs with increasing levels of detail. Start with a context diagram (Level 0) and then decompose each process into lower-level DFDs until a sufficient level of detail is achieved.

**Balancing DFDs:**

* **Balancing:** Ensures that the input and output data flows of a process on one level match the input and output data flows of the corresponding process on the next level. It maintains consistency in data flow throughout the DFD hierarchy.

**Exploded DFD:**

* **Exploded DFD:** Also known as a levelled DFD, it shows a detailed breakdown of a process into its subprocesses at a lower level. It provides a more granular view of the internal workings of a specific process.

**Data Dictionary:**

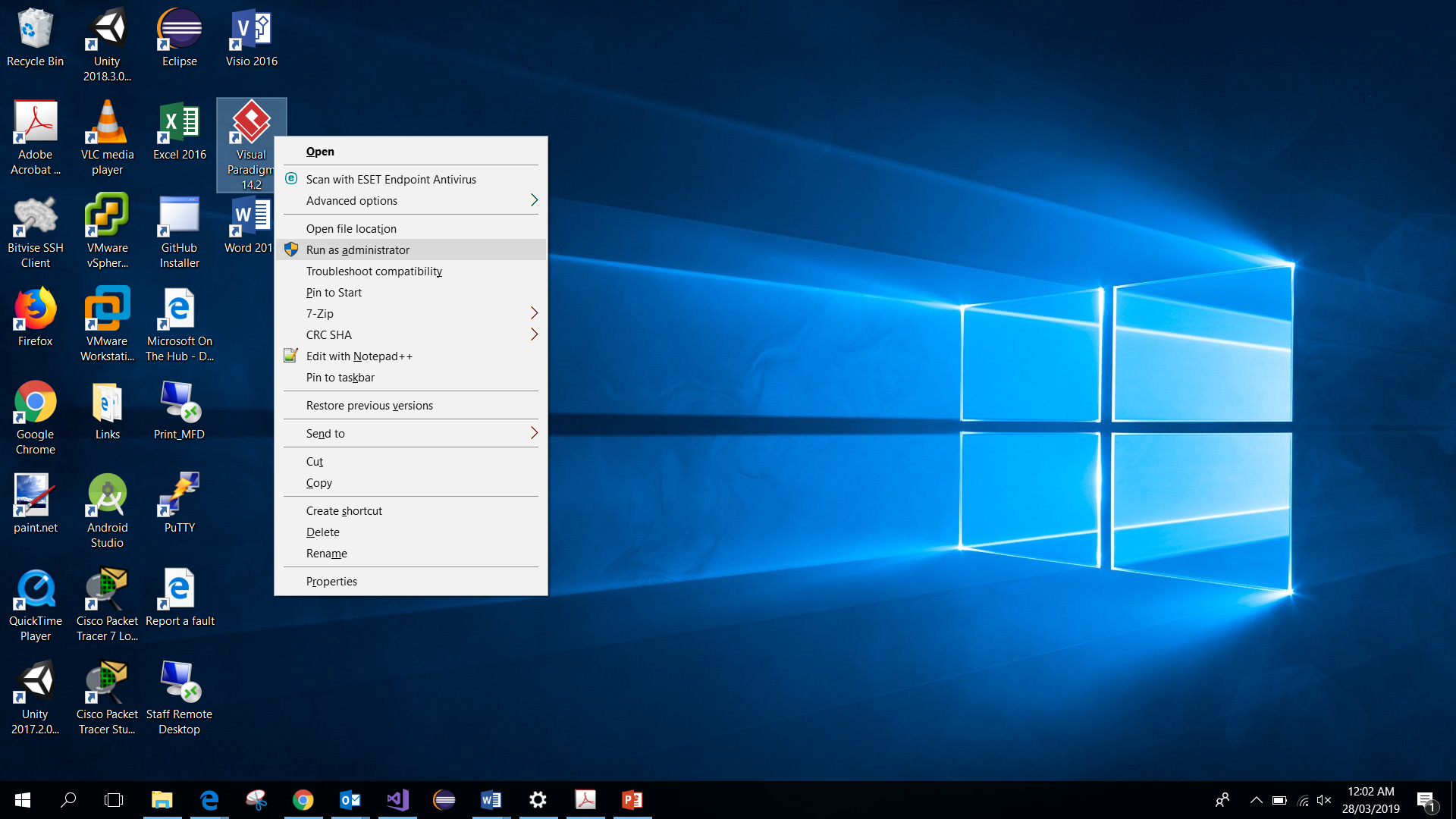
* **Data Dictionary:** A centralized repository that defines and describes data elements and their relationships within a system. It includes attributes, data types, and constraints. Examples of usage include ensuring consistency in data representation across different parts of a system and aiding in database design.

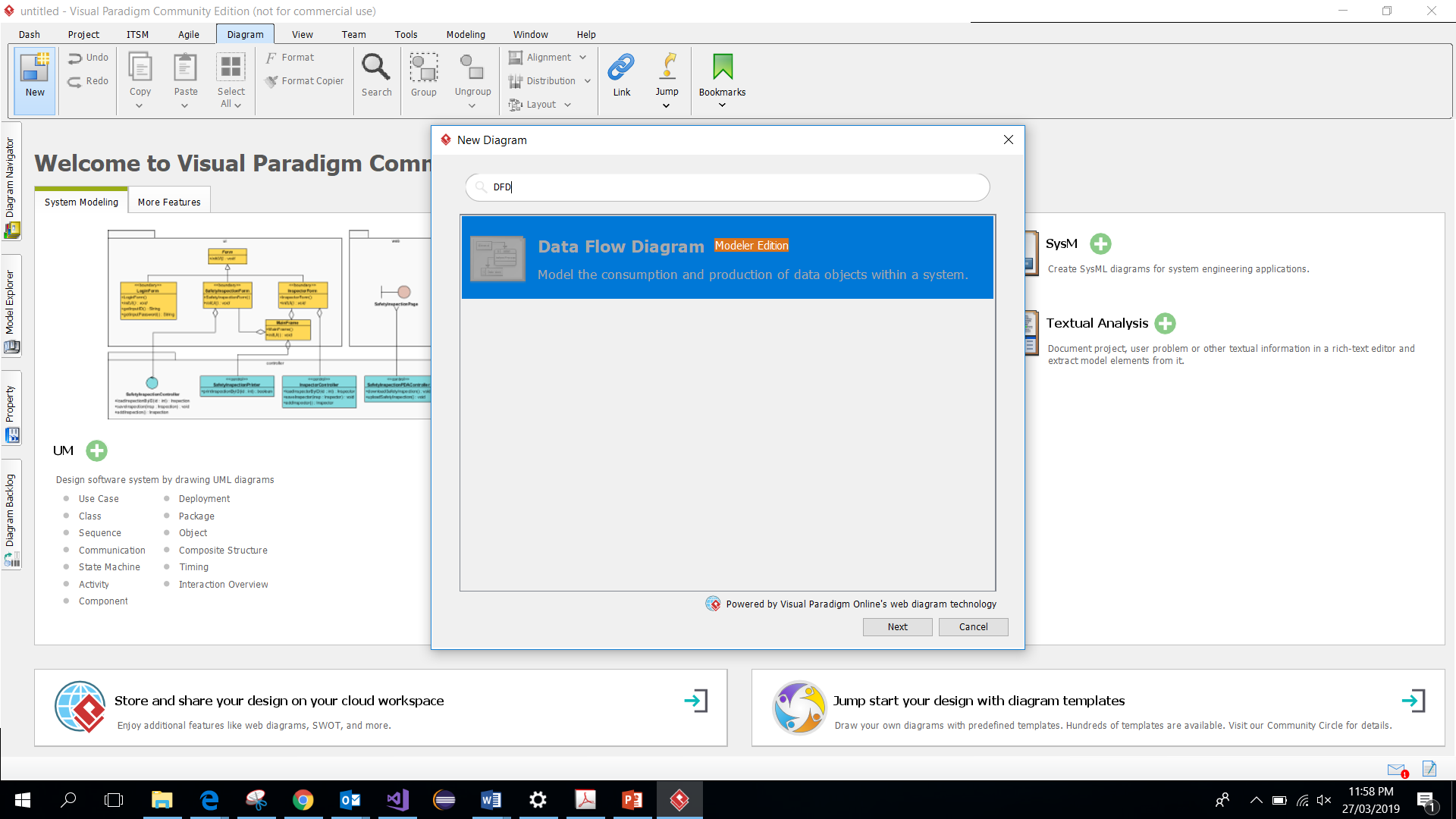
**Activity 2)**

**DFD diagrams with Visual Paradigm**

Visual Paradigm keeps track of the hierarchies of a DFD for you. This makes it much easier to create DFD diagrams.

**Part 1. Set Up Visual Paradigm**

1. Run Visual Paradigm as an Administrator 
2. **Diagram** > **New** > type: **DFD** > Next



1. Choose “Yes (Evaluate Visual Paradigm Enterprise)” > Next.

**Part 2. Draw DFD**

In this section, you will a Payroll System DFD and then add to it. It is important that you use the features of Visual Paradigm to help create the hierarchy.

1. Create a new Visual Paradigm project called *Payroll*. Save the project.
2. Add a new DFD diagram to your project, call it Payroll System Context DFD and create this context level DFD:

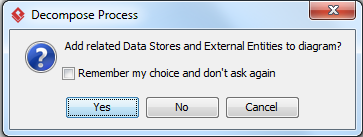
A diagram of a paypal system

Description automatically generated

Note that it is important to number your single process as 0. You do this by right-clicking the process and selecting *Open Specification*. Change the *DF ID* field to 0.

1. To create a level 1 DFD, you right-click the *Payroll System* process and select Decompose.

With a bit of luck, you will get this prompt:



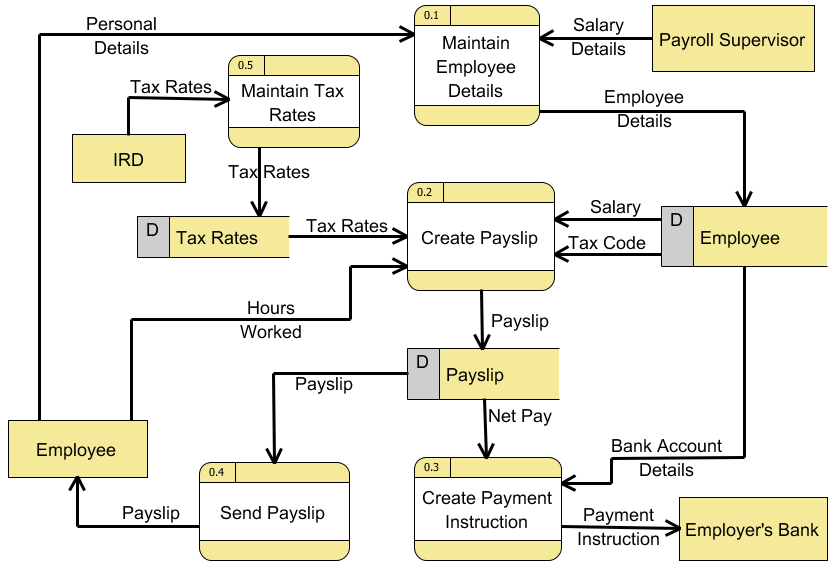
1. Select *Yes* to this prompt. This will create a new DFD page with the external entities already on it.

Note: if you don’t get this prompt, and no entities from the context diagram appeared on your level 1 diagram, go to the *Window Tab -> Project options -> DFD* tab and select the *Prompt* radio tab in the single item and then *Apply* the changes.

1. Double-click the name of the diagram and change it to *Level 1*.



1. Now create this DFD:



Number your processes as shown. Note that Visual Paradigm will create the 0. prefix for you (assuming you labelled the process in the context DFD); you just add the suffix, e.g. 0.1, 0.2, 0.3, and etc. Alternatively, if you labelled the context DFD as 1, Visual Paradigm create the 1. Prefix. Therefore, you can have 1.1, 1.2, 1.3, etc.

**Activity 3)**

Draw a context diagram and a diagram 0 DFD that represent the registration system at a university. Identify the possible external entities, data flows, processes, and data stores.

Diagram

Description automatically generated