



COURSEWORK - REPORT

Academic year 2022/2023 Autumn Term
CS7079 Data Warehousing and Big Data
PART A – Group Work

Abstract

The aim of this coursework is to design, implement, and test a data warehouse based on a given business case scenario, to export data from it and ingest it for further processing on a Big Data platform.

	ADEBAYO TAOFIK ADEKOYA (21053457)	PANKAJ KASHYAP (22012469)	KUNIK SUJIT NAYYAR (22018515)
Prepared By: <u>Group 23</u>	IBRAHIM ADEMOLA OYENEYE (22024597)	SAMRAT RAI (21025221)	SHAMSHER SINGH (22014018)
	SHRUTI ASHOKBHAI PATEL (22021294)	PRUTHVIKKUMAR ATULBHAI VAGHASIYA (ST_ID)	JEFERSON JEFERSON CARGUA BASTIDAS (ST_ID)

Background information

ABC consumer electronics outlet Ltd. (a multi-channel retailer of consumer electronics) has operations in six physical outlets with an online store. The company sells more than ten thousand products within ten categories and 200 brands. The company has already invested in information and communication technologies to run smooth operations.

Currently, the company uses a cloud-based programme Vend to support several of its retail activities. It also uses Linnworks to automate its internet operations and Xerox to help its financial department. Every day, these software programmes generate a massive amount of transactional data. Each software programme stores and manages the generated datasets independently.

Managers at ABC Consumer Electronics Retail Company are struggling to prepare reports and conduct studies that will allow them to deliver information for decision-making.

As a result, a business analyst advised the organisation to build and install a data warehouse that would suit all reporting and analytical needs.

Introduction

The data warehouse solution will exclusively focus on the inventory management business process. As reported by managers, the company is facing difficulties in controlling over-stocks and under-stocks of product items. The three main business activities which are considered under inventory management are:

1. Send purchase orders to suppliers when product items have low inventory levels. The ordered amount sent date, product and supplier data are all logged for each sent purchase order.
2. Receiving purchase orders and storing inventory in proper locations. The ordered amount, received quantity, received date, product and supplier data are all recorded for each received purchase order.
3. Stock control and maintenance entail introducing new items and modifying stock levels of current products. Each product's data, including the current stock level, are recorded.

Where:

Inventory managers are under pressure to satisfy the following reporting and analytical requirements:

- A daily stock level of all products for the last month.
- A weekly report of all products with minimum stock levels.
- Analysing stock levels by brand or product type or supplier
- Daily and weekly sent and received stock orders for the last four weeks.
- Analysing received stock orders by the supplier and by month.

Description of Data Sources

Sample data from the software application “Vend” is provided to understand the overview of suppliers, products, and stock movements. The data from the source includes supplier details, product details and details about stock movements from suppliers to warehouse locations.

1) Analysis & Design of Dimensional Data Model

1.1. To represent the three main activities of the inventory management business area, there are three different central fact tables are considered which described as follows:

1) SENT PURCHASED ORDER

The fact table will record the ordered quantity, sent date, product and supplier details when products are at their minimum level of remaining stock. This fact table seeks to streamline activities related to sending purchase orders. The lowest grain of this fact table is identified as "one sent purchased order number per row" which will store and retrieve all required information from the connected dimensions.

SENT PURCHASED ORDER
Sent purchased order number
Sent purchased order date
Sent purchased order quantity

2) RECEIVED PURCHASED ORDER

The fact table will record the destination of storage, ordered quantity, received purchased order quantity, received date, product and supplier details when products are at their minimum level of remaining stock. The objective of this fact table is to make the process of submitting purchase orders more efficient. This fact table's lowest granularity level is identified as "one received purchased order number per row," which will read and write all necessary data from the linked dimensions.

RECEIVED PURCHASED ORDER
Received purchased order number
Received purchased order sent date
Received purchased order received date
Received purchased order received quantity
Received purchased order ordered quantity

3) STOCK

This fact table will control and maintain stocks which include adding products or adjusting stock levels, as well as generating reports of stocks on demand. The focus of this fact table is to capture the movement of every unit of inventory right from when they are supplied to when they leave the different stores. The grain for this fact table, therefore, is "one product stock level per row" as this will afford stock analysis on a more granular level.

STOCK
Add new product
Add stock to existing product
Current stock level

1.2. To meet the reporting and analysis requirements mentioned by the company, the following dimensions are identified to support three fact tables.

- 1) Supplier: this dimension will contain and store all information about the supplier and it will connect with all three central fact tables using the primary key "supplier_key"
- 2) Date: this dimension will contain and store all date-related information and it will connect with all three central fact tables using the primary key "date_key"
- 3) Order: this dimension will contain and store order details such as purchase order type, order date and destination outlet. Sent purchased orders and received purchased orders will be filtered by purchase order type. It will join two fact tables "sent purchased order" and received "purchased order" by using the primary key "order_key"
- 4) Product: this dimension will store and manage all information about product type, SKU number, name, brand, stock level, price etc. It will join with all three central fact tables using the primary key "product_key"

1.3. Based on given business requirements, and determining dimensions of three central fact tables, different attributes are considered to complete the dimensional data model.

Supplier	Date	Order	Product
Supplier name	Full date	Purchase order Type	Product Type
Supplier phone number	Day name of the week	Purchase order number	SKU Number
Supplier description	Week of month	Ordered quantity	Product name
Supplier Email	Week of year	Ordered unit price	Product Description
Supplier address	Month name	Ordered date	Product condition
Supplier City	Month of year	Destination outlet	Product brand
Supplier postcode	Quarter	Product name	Supplier name
Supplier country	year	⇌ order_key	Product tag
⇌ supplier_key	⇌ date_key		Current stock level
			Product date added
			Retail Price
			⇌ product_key

1.4. Graphical representation of structure: Simple star schema is used to design the Dimensional Data Model.

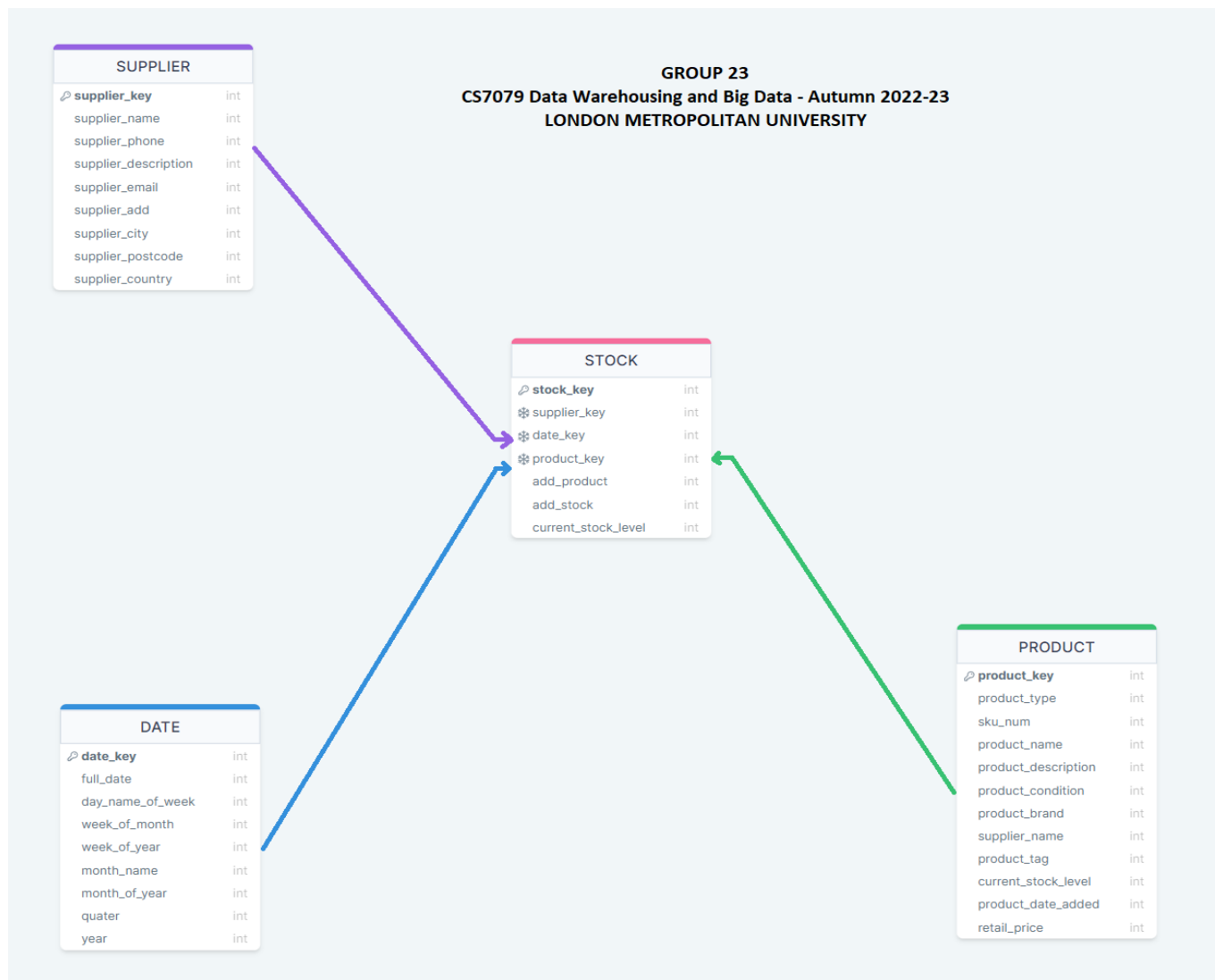


Figure 1: Supplier, Date and Product dimensions are directly connected to the "Stock" fact table.

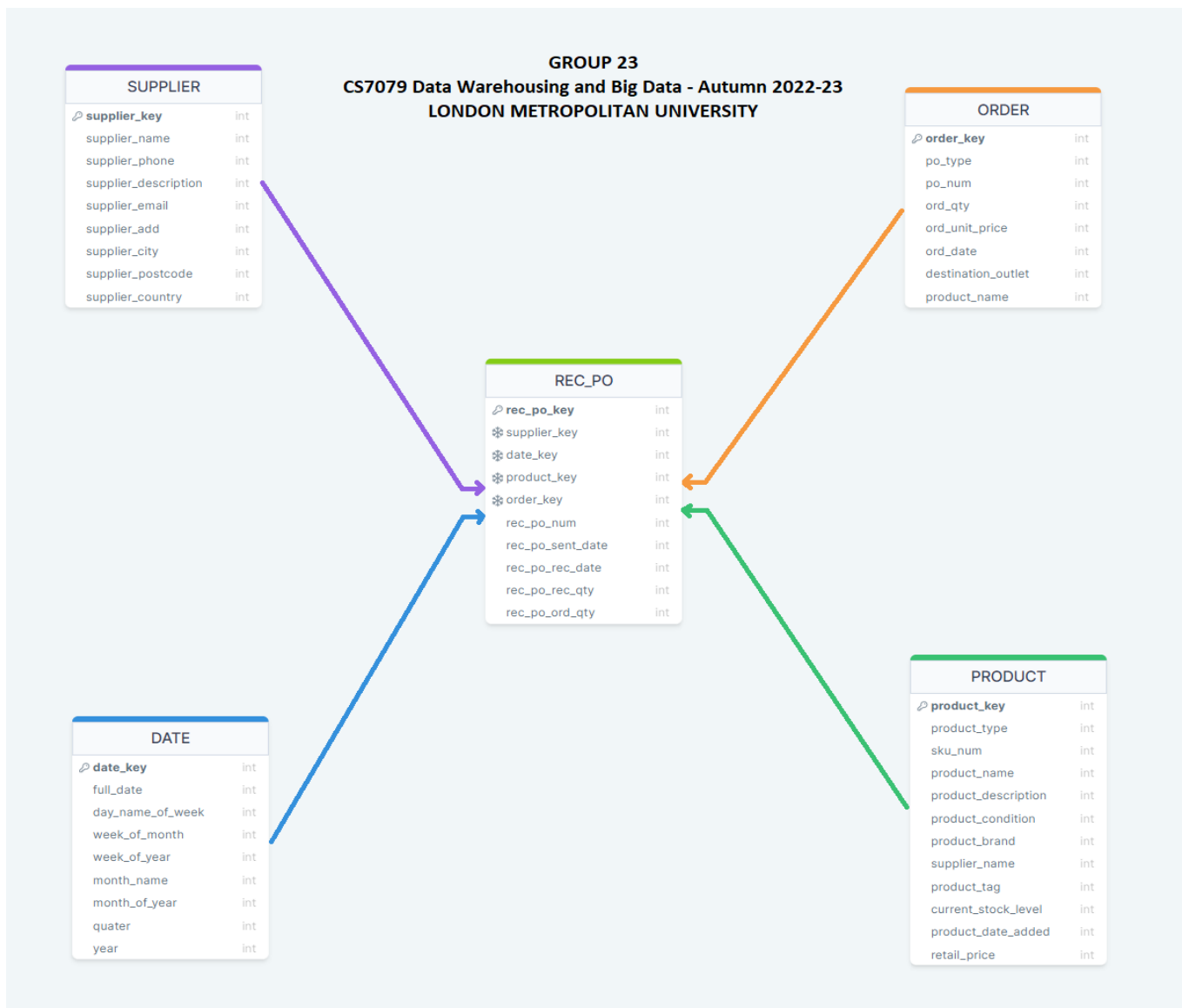


Figure 2: Supplier, Date, Product and order dimensions are directly connected to the "Received Purchased Order" fact table.

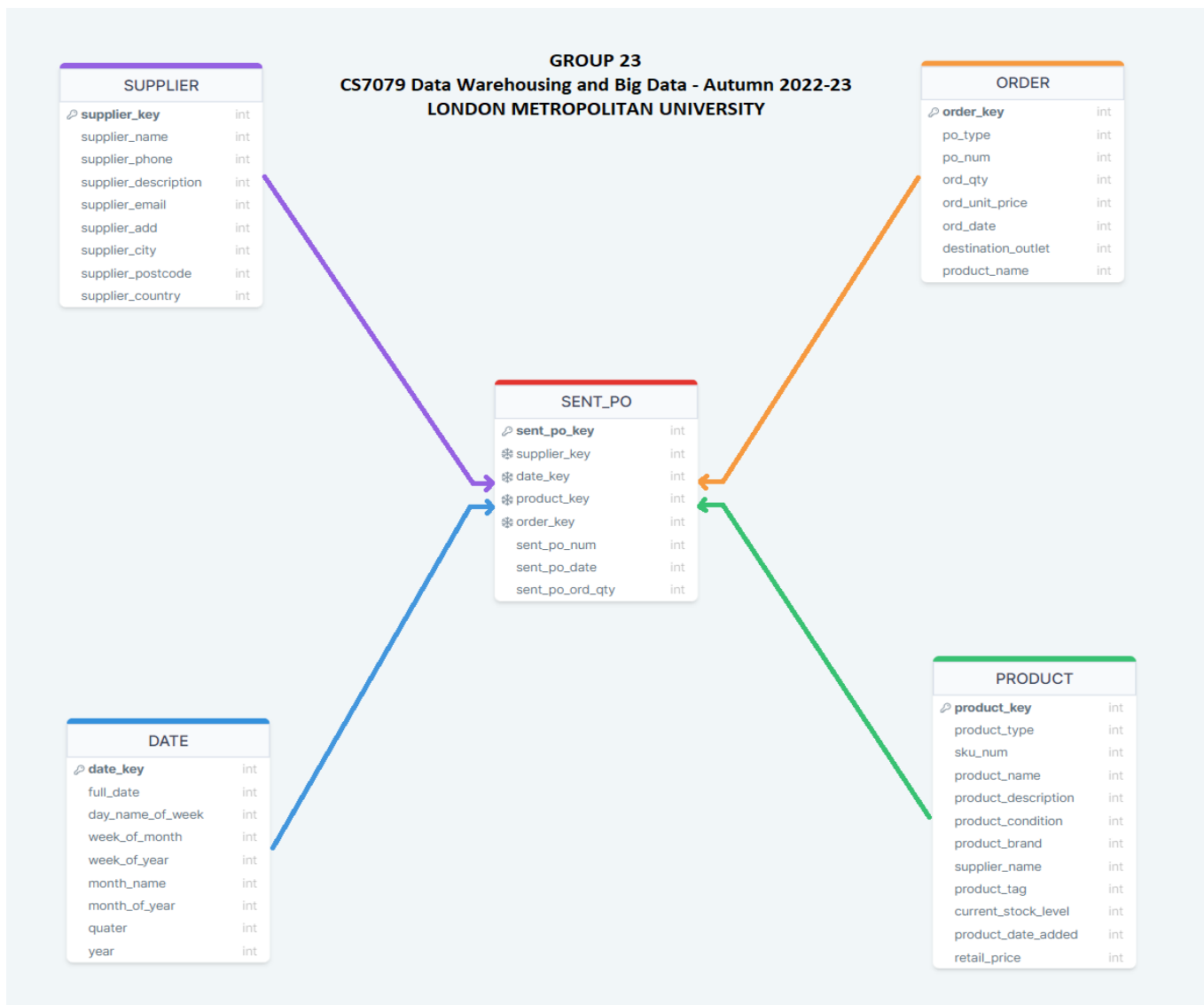


Figure 3: Supplier, Date, Product and order dimensions are directly connected to the "Sent Purchased Order" fact table

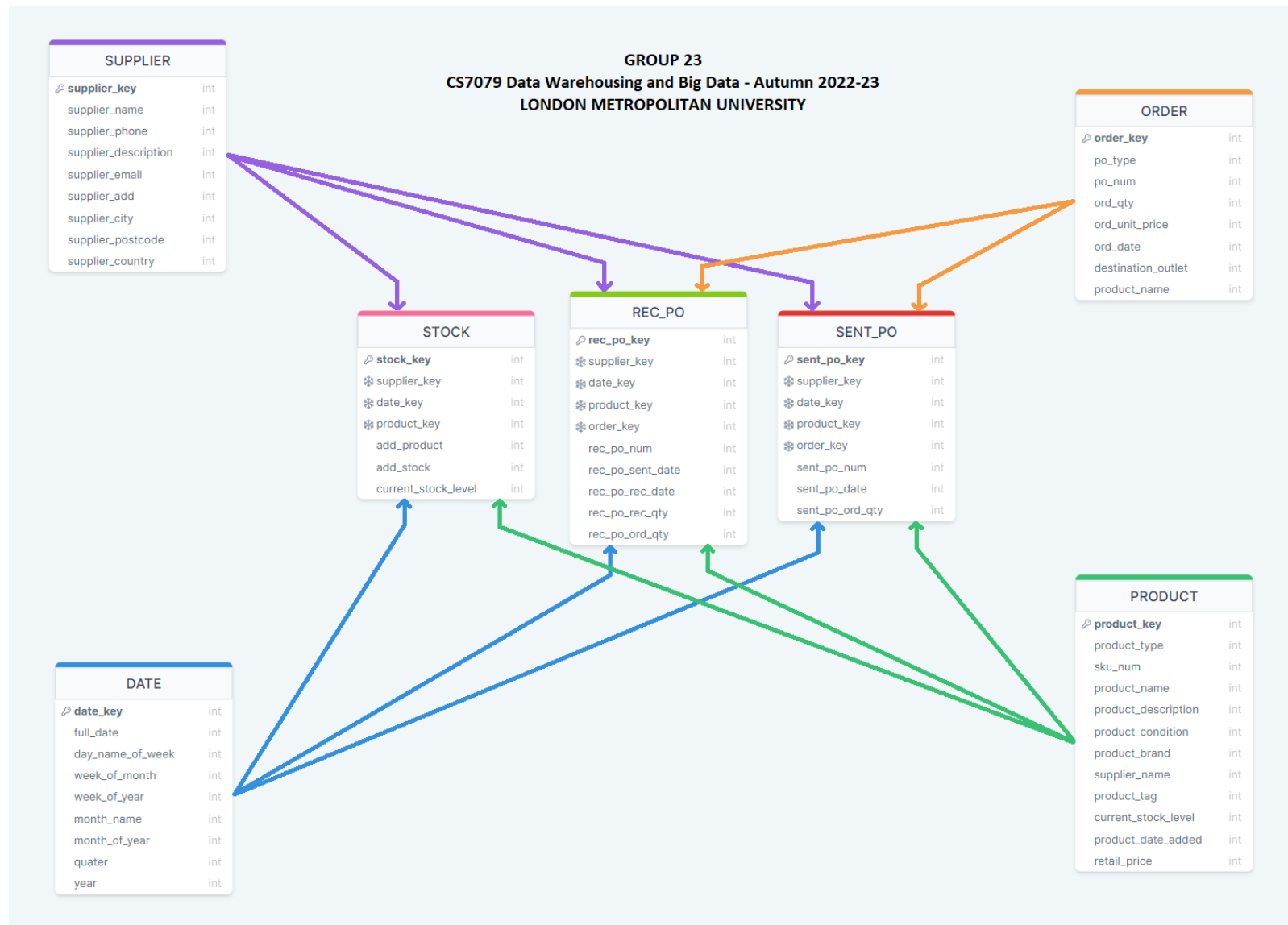


Figure 4: Supplier, Date, Product and order dimensions are connected to central fact tables.