



Data Management Challenge

Data Management Agenda

- ## Use Case Context
- ## Use Case Description
- ## Technical Requirements

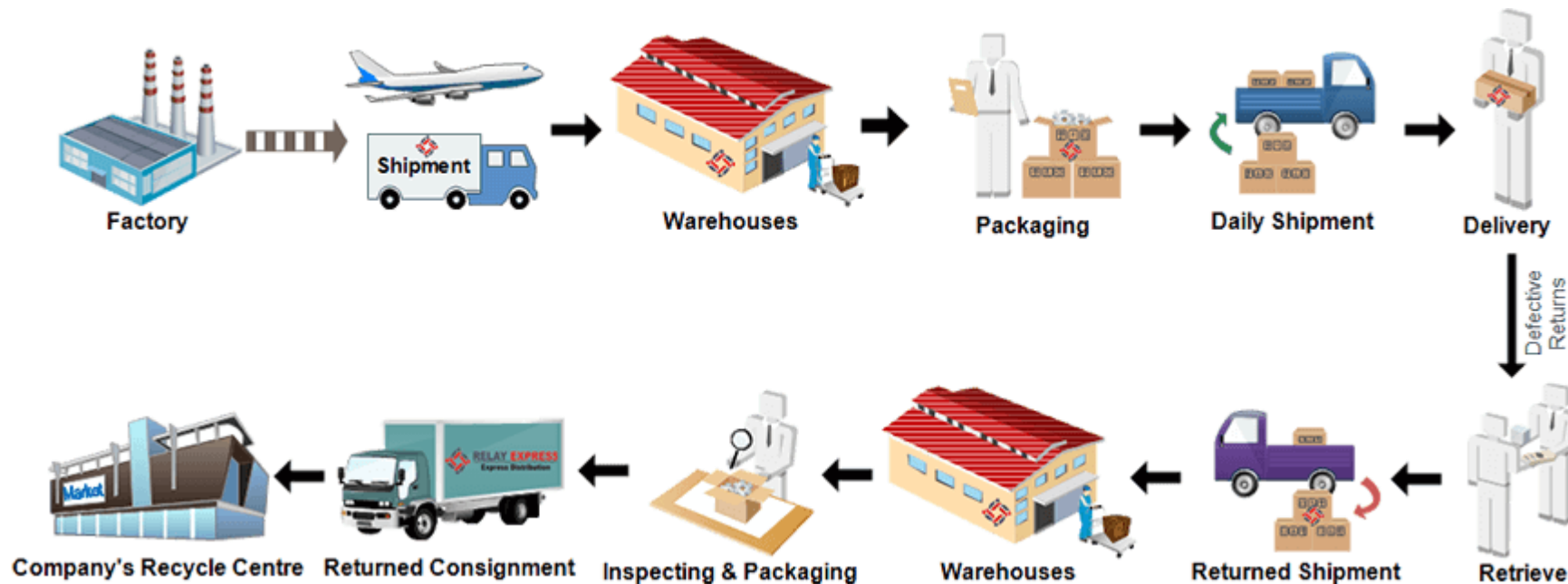
1

Use Case Context

Use Case Context

Supply Chain Introduction

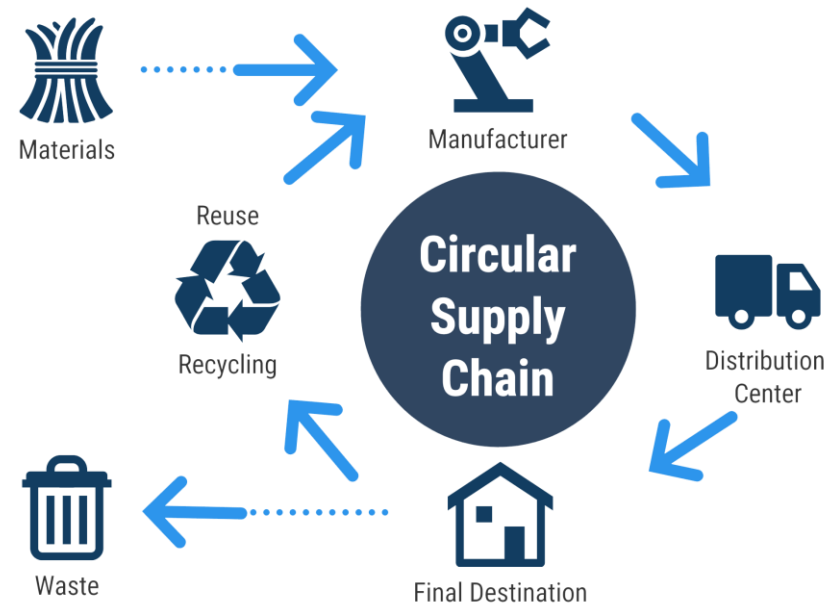
- A supply chain is a network of individuals, organizations, activities, information, and resources involved in creating and delivering a product or service to customers.
- It begins with the raw materials and components needed to manufacture a product and ends with the delivery of the finished product to the end customer.



Use Case Context

Sustainability Challenge

- Due to legal and fiscal requirements, companies have an obligation to identify the nature of each component that makes up a final product, in order to develop auditable reports that contain information regarding plastic consumption, carbon footprint, energy consumption and other indicators.
- In this real use case, we will focus on identifying the spread quantity and nature of each of the components at all stages of the Supply Chain.



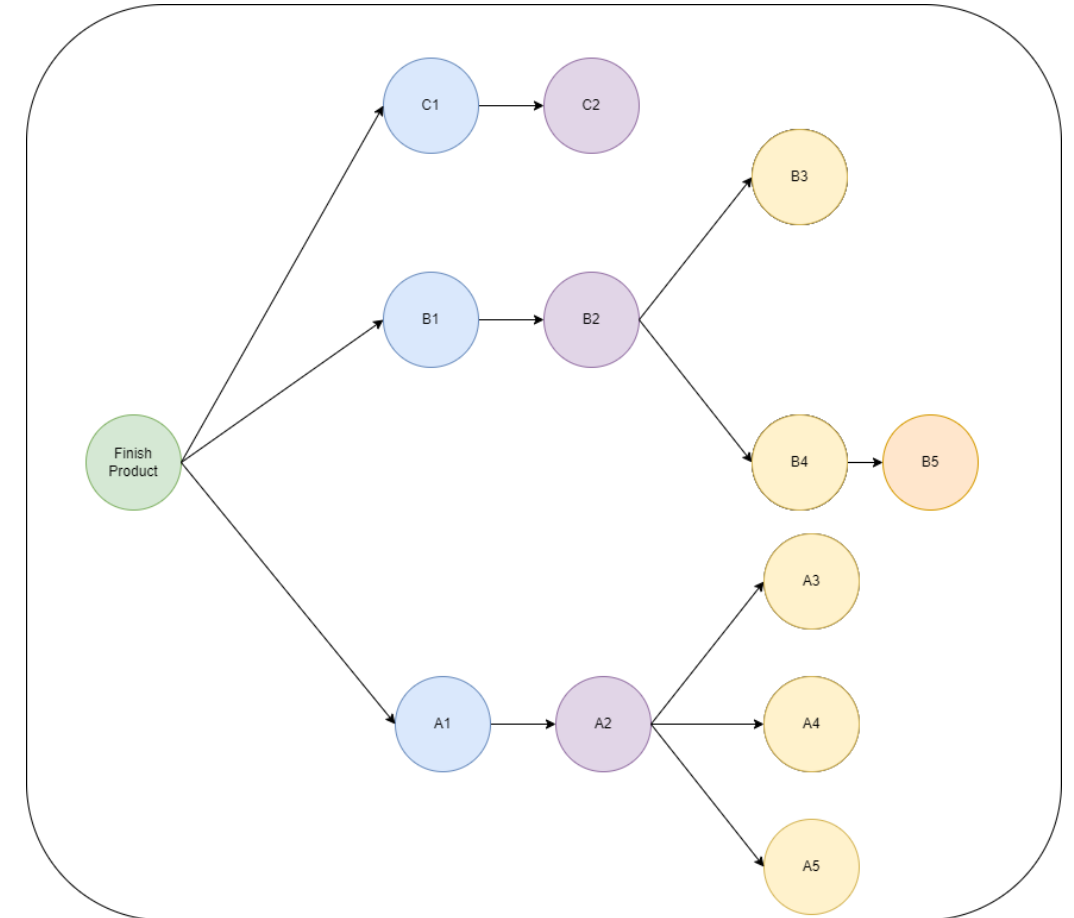
2

Use Case Description

Use Case Description

Use Case Overview

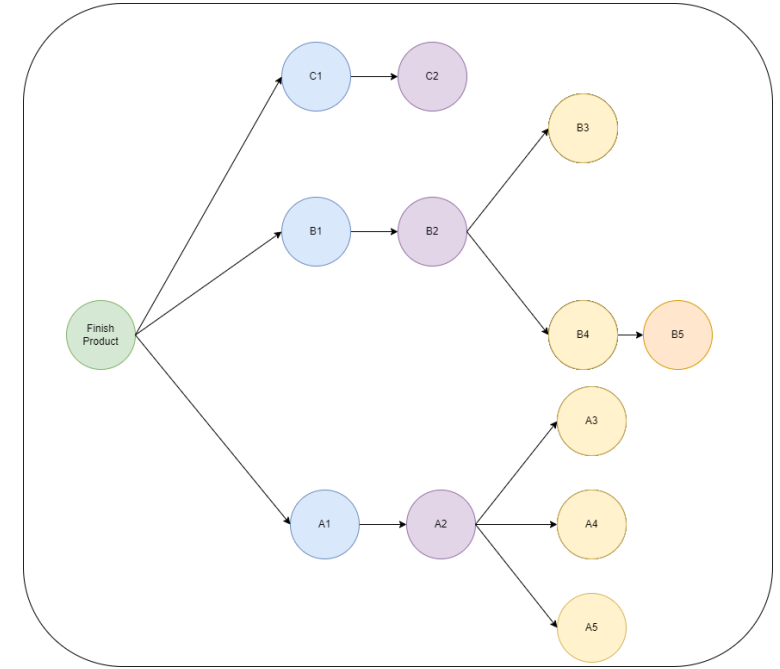
- The main challenge presented by this use case is to **generate** the **supply chain network** using the information stored in SAP's ERP system as a data source.
- A very high percentage of companies use SAP as a transactional system to record all purchase and production orders within the company.
- Once the network has been generated correctly, the **amount of propagation** of the **components** at each **level** of the chain must be **calculated** to generate the final product.



Use Case Description

Use Case Objectives

- As a final objective, a table should be generated that contains:
 - Hierarchy of relationships of the final products with each of the necessary components represented by levels
 - Calculate the propagation amount of a component depending on the propagation amount of its predecessor.



Level	Finish Product	Finish Product Quantity	Batch	Plant	Order	Component	Prior_Component	Prior_Quantity_Component	Batch_Component	Plant_Component	Order_Component	Material Nature	Quantity_Component	Propagated Quantity
0	A	100	XXX	XXX	XXX	A			XXX	XXX	XXX	XXX	100	1,00
1	A	100	XXX	XXX	XXX	A1	A	100	XXX	XXX	XXX	XXX	100	1,00
1	A	100	XXX	XXX	XXX	A2	A	100	XXX	XXX	XXX	XXX	100	1,00
1	A	100	XXX	XXX	XXX	A3	A	100	XXX	XXX	XXX	XXX	200	2,00
1	A	100	XXX	XXX	XXX	A4	A	100	XXX	XXX	XXX	XXX	200	2,00
2	A	100	XXX	XXX	XXX	A5	A1	50	XXX	XXX	XXX	XXX	50	1,00
2	A	100	XXX	XXX	XXX	A6	A2	20	XXX	XXX	XXX	XXX	30	1,50
2	A	100	XXX	XXX	XXX	A7	A3	40	XXX	XXX	XXX	XXX	20	1,00
2	A	100	XXX	XXX	XXX	A8	A4	50	XXX	XXX	XXX	XXX	10	0,40
3	A	100	XXX	XXX	XXX	A9	A8	10	XXX	XXX	XXX	XXX	50	2,00
4	A	100	XXX	XXX	XXX	A10	A9	1000	XXX	XXX	XXX	XXX	100	0,20
4	A	100	XXX	XXX	XXX	A11	A9	1500	XXX	XXX	XXX	XXX	200	0,27
4	A	100	XXX	XXX	XXX	A12	A9	200	XXX	XXX	XXX	XXX	300	3,00
4	A	100	XXX	XXX	XXX	A13	A9	240	XXX	XXX	XXX	XXX	300	2,50

3

Technical Requirements

Technical Requirements

What are we going to provide

- For the correct development of the solution, SAP dummy tables will be provided to the student, containing approximately 10 million records.
- Additionally, the tables will be accompanied by their metadata descriptions to give the students maximum autonomy in developing the algorithm.
- Business logic to achieve all development phases.

SAP Table	Description	Volumetry
CHVW	Standard Inventory Management Transparent Table in SAP, wich stores Batch Where-Used List data	8M
MAKT	Material Management Table that contains the material nature.	400K

Technical Requirements

Development planning

- Phase 1: Build CHVW_AGG based on CHVW.
- Phase 2: Build Connected Hierarchy based on CHVW_AGG.
- Phase 3: Add material consumption information to Connected Hierarchy.
- Phase 4: Build logic for calculating propagated quantity for each component for every finish product.

Technical Requirements

Technical Challenges

- **Recommended programming languages**
 - Python
 - Scala
 - SQL
- **Recommended environment**
 - HDFS
 - Oracle Database
 - AWS, GPC, Microsoft Azure
- **Data Management**
 - Environment resource management
 - Recursive functions in parallel distribution computation
 - Window functions
 - High volume of data management
- **Optimization**
 - Optimal code development

Thank you

