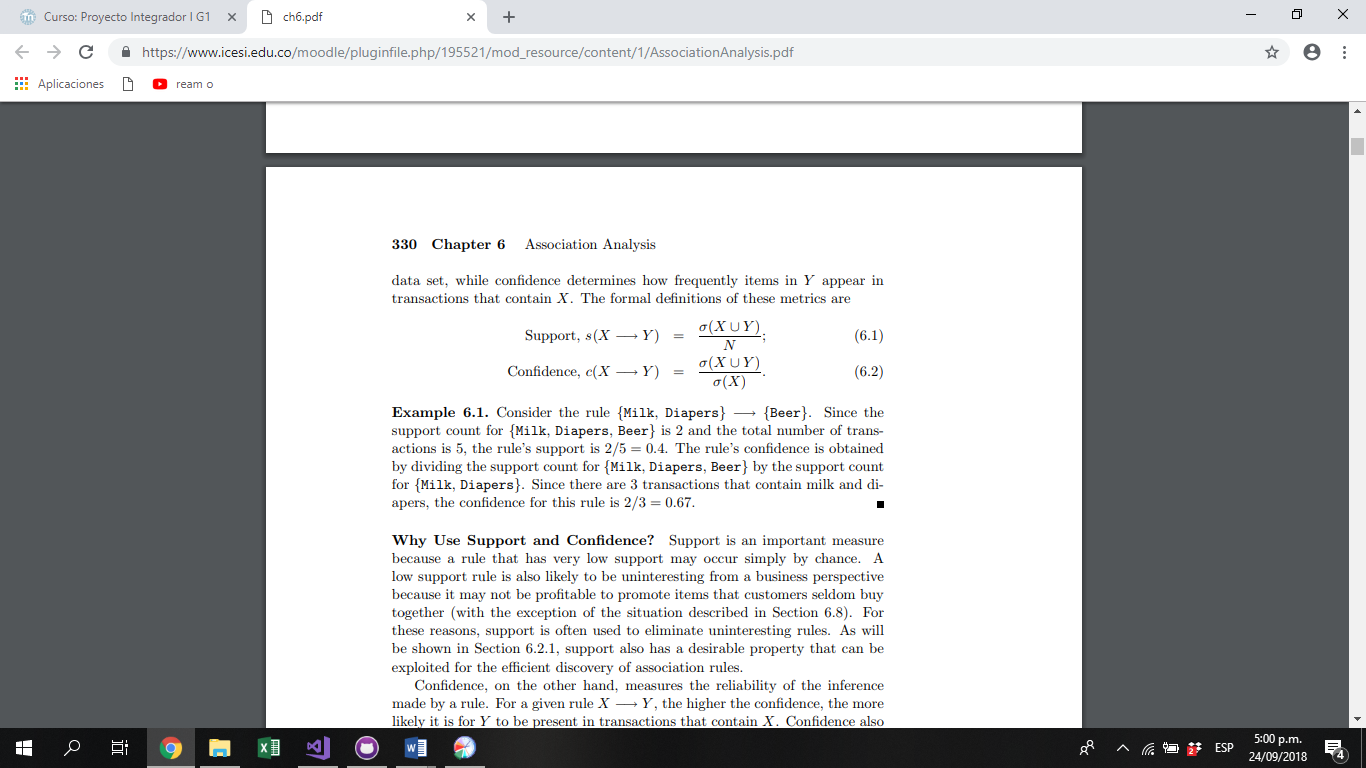
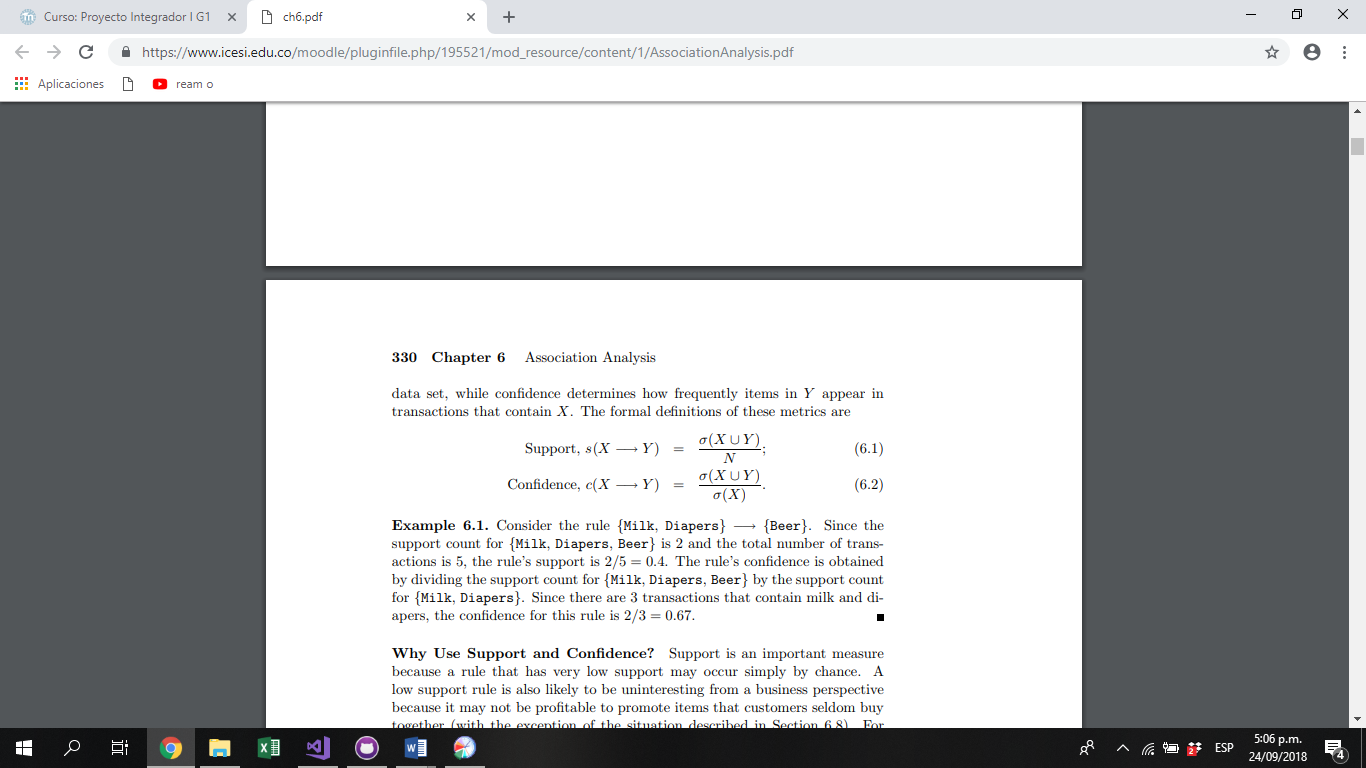
**Fase 2: Recopilación de información:**

*Marco teórico:*

1. Allers Groups: Empresa Colombiana especializada en la venta de equipos médicos, insumos hospitalarios, instrumental quirúrgico y medicamentos, tanto al por mayor como al detal.
2. Enterprise resource planning (ERP): Es un tipo de aplicación que permite la recolección de datos y todo los procesos que conllevan para generar la interpretación de esos datos[[1]](#footnote-1)
3. Data Mining: Es el proceso por el cual se intenta obtener patrones en un conjunto de datos
4. Asociación (data mining): Es una técnica de data mining en donde se utiliza funciones probabilísticas que permiten saber la ocurrencia de un evento B dado un evento A.
5. Representación binaria: Consiste en representar unas transacciones como cadenas binarias, donde 1 indica que el producto aparece en una transacción y 0 que no.
6. Support: Frecuencia en la que un itemset dado (sets de productos) aparece en las transacciones. 
7. Frequent Item Sets: Itemsets que aparecen frecuentemente en las transacciones. Se considera que un itemset es frecuente si su support es superior a un minsup definido.
8. Confidence: Frecuencia en la que dos itemsets X e Y aparecen en una transacción que contiene al itemset X. 
9. Rule: Reglas de la forma X-> Y cuya confidencia es superior a un minconf. Además, XY es un Frequent Item set.
10. Principio de Apiori: Teorema que afirma que, si un itemset es frecuente, todos sus subconjuntos deben de ser frecuentes también. A partir de este principio, si se encuentra un conjunto X que no es frecuente, se puede inferir que todo conjunto que contenga a X tampoco va a ser frecuente. Dicho principio evita computar itemsets que de antemano se sabe no van a ser frecuentes.
11. Principio de monotonicidad: Dada una función f, si X es subconjunto de Y, entonces f(X) <= f(Y). El principio de Apiori se basa en la anti-monotonicidad (lo opuesto a monotonicidad).
12. Máximo frequent Item set: Un itemset frecuente para el cual ninguno de sus itemsets inmediatamente superiores es frecuente.
13. Frequent itemsets cerrados: Se considera un itemset X como cerrado si es frecuente y ninguno de los conjuntos inmediatamente superiores tiene el mismo support count. Sirve para eliminar las reglas redundantes.
14. Big Data: Enfoque para describir enormes cantidades de datos (tanto estructurados, no estructurados o semi-estructurados) que tomaría demasiado tiempo cargarlos a una base de datos relacional. Es decir, es la información que no puede ser procesada con herramientas tradicionales (Por, por ejemplo, requerir una velocidad de respuesta muy rápida). Esto no quiere decir que se deban despreciar las bases de datos convencionales, sino que van de la mano, cada uno cumpliendo funciones diferentes. Existen varios tipos de Big Data, entre los que se encuentra: Web y Social media, Biométricos, Machine to Machine (M2M), Human Generated y Big Transaction Data. Actualmente, la herramienta con liderazgo en popularidad para analizar Big Data es Hadoop, el cual está inspirado en GoogleFile Systems y utiliza el paradigma MapReduce[[2]](#footnote-2).
15. Base de datos Relacional: Base de datos conformada por un conjunto de tablas, que se manipula de acuerdo al modelo de datos relacional (Donde las relaciones son tuplas, lo que se puede ver como si cada tabla fuera una relación).[[3]](#footnote-3)
16. Paradigma Map-Reduce: Framework para trabajar sobre grandes colecciones de datos en varios grupos de computadoras para aprovechar la computación paralela. Está compuesto por dos funciones: Función Map(), la cual mapea los datos de entrada de la función en paralelo y genera una lista de pares (k2,v2). La otra función es la de Reduce(), la cual se llama en paralelo para cada grupo generado en la función Map y retorna un valor final v3.[[4]](#footnote-4)
17. Business Intelligence: Conjunto de metodologías, aplicaciones y tecnologías que permiten reunir, depurar y transformar datos de los sistemas transaccionales e información desestructurada (interna y externa a la compañía) en información estructurada, para su explotación o para su análisis y conversión en conocimiento, dando así soporte a la toma de decisiones sobre el negocio[[5]](#footnote-5).

*Otros desarrollos:*

1. SAP Bussines one: Es un software para gestión de tipo ERP[[6]](#footnote-6) y diseñado para empresas pequeñas-mediana.
2. Hadoop: Herramienta elaborada por el grupo Apache para manipular datos distribuidos logrando un alto paralelismo en el procesamiento. Está compuesto por Hadoop Distributed Files (Los datos son divididos en bloques y distribuidos a través del Cluster), Hadoop Map-Reduce(Aplica el proceso de Map-Reduce a los datos distribuidos) y Hadoop Common(Conjunto de librerías que soportan subproyectos de Hadoop)3.
3. QlikView: Producto para elaborar aplicaciones de analítica interactivas y Dashboards. Permite a los usuarios construir aplicaciones de analítica sin necesidad de tener grandes habilidades de desarrollo, lo que permite actuar fácilmente a los cambiantes requerimientos empresariales[[7]](#footnote-7).

Agregar cosas más específicas sobre Data Mining

1. https://en.wikipedia.org/wiki/Enterprise\_resource\_planning [↑](#footnote-ref-1)
2. https://www.ibm.com/developerworks/ssa/local/im/que-es-big-data/index.html [↑](#footnote-ref-2)
3. https://www.ibm.com/support/knowledgecenter/es/SSEPGG\_8.2.0/com.ibm.db2.udb.doc/admin/c0004099.htm [↑](#footnote-ref-3)
4. https://es.wikipedia.org/wiki/MapReduce [↑](#footnote-ref-4)
5. https://www.sinnexus.com/business\_intelligence/ [↑](#footnote-ref-5)
6. https://en.wikipedia.org/wiki/SAP\_Business\_One [↑](#footnote-ref-6)
7. https://www.qlik.com/es-es/products/qlikview [↑](#footnote-ref-7)