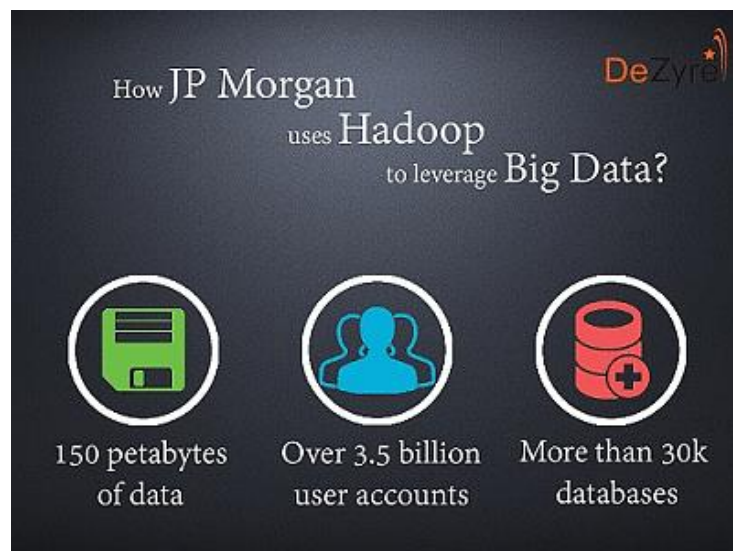


Homework : Big Data Analytics : Hadoop

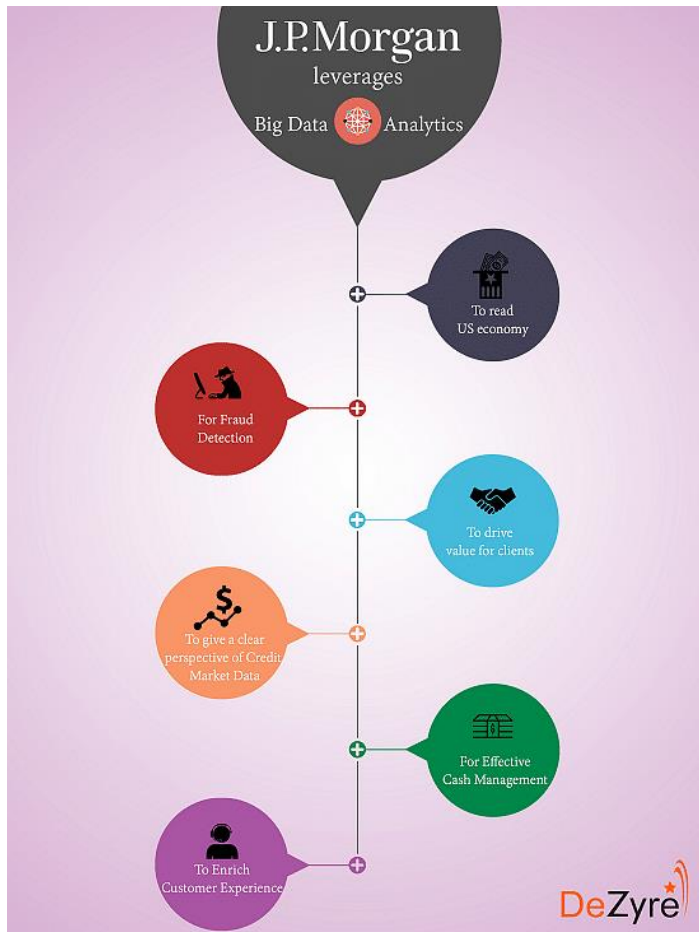
Assignment 1 : นอกจากบริษัท Facebook กับ Yahoo ที่ผมเสนอมาแล้ววันนี้ว่า เป็นบริษัทฯ ที่ใช้ระบบจัดการฐานข้อมูล ของ Apache Hadoop ในการวิเคราะห์ข้อมูลแบบ Un-Structure ดังนั้นจึงอยากทราบว่า มีบริษัทฯ ใดอีกบ้างที่ใช้รูปแบบการประมวลผลเบื้องหลังเป็น Apache Hadoop มาทำ Commercial Distribution ลอง เสนอมาสัก 3 ตัวอย่างว่า บริษัทเหล่านั้นใช้ประโยชน์อะไรจาก Hadoop

Ans : ตัวอย่างของบริษัทที่ใช้ Hadoop เช่น

1. **JP Morgan** หนึ่งในยักษ์ใหญ่สำหรับ US Banking Industry ซึ่งมีปริมาณข้อมูลมากกว่า 150 PB 3.5 Millions User Account และ 30,000 database เลือกใช้ Hadoop สำหรับ Process Customer Transaction data กับ Banking Product นับพันรูปแบบและหลายระบบที่แตกต่างกัน เพื่อให้มีประสิทธิภาพการดำเนินการสูงสุด (Drive Value for Clients & For Effective Cash Management) รวมทั้งนำข้อมูลมาพัฒนาระบบเพื่อเพิ่มประสบการณ์ผู้ใช้ นอกจากนี้ยังใช้มาทำการวิเคราะห์สภาพเศรษฐกิจในอเมริกา และจัดการและ Fraud Detection อีกด้วย

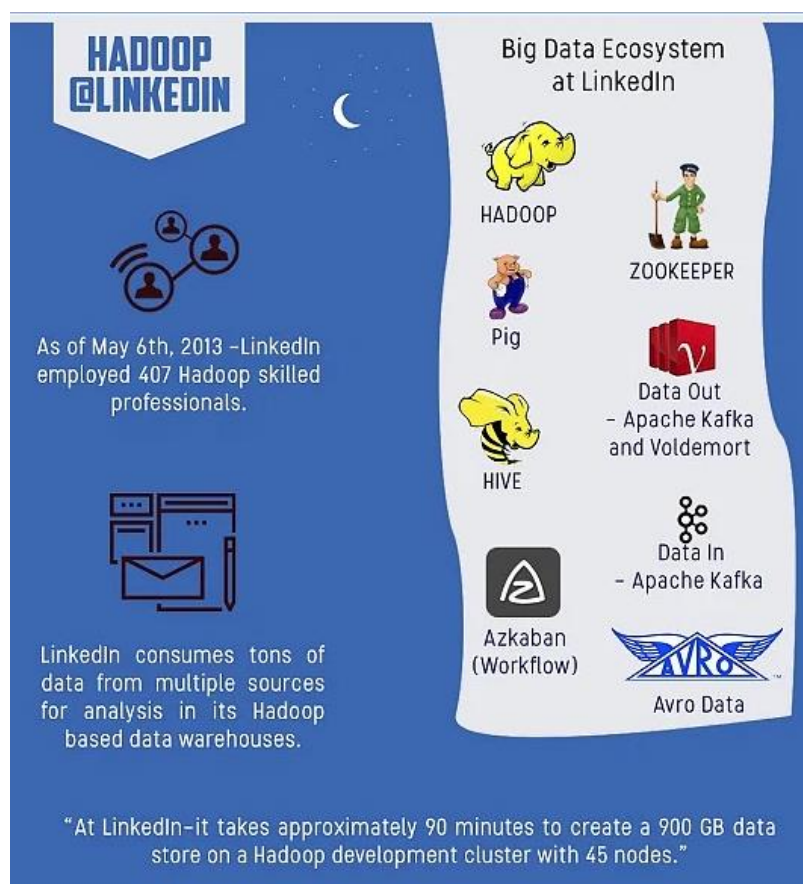
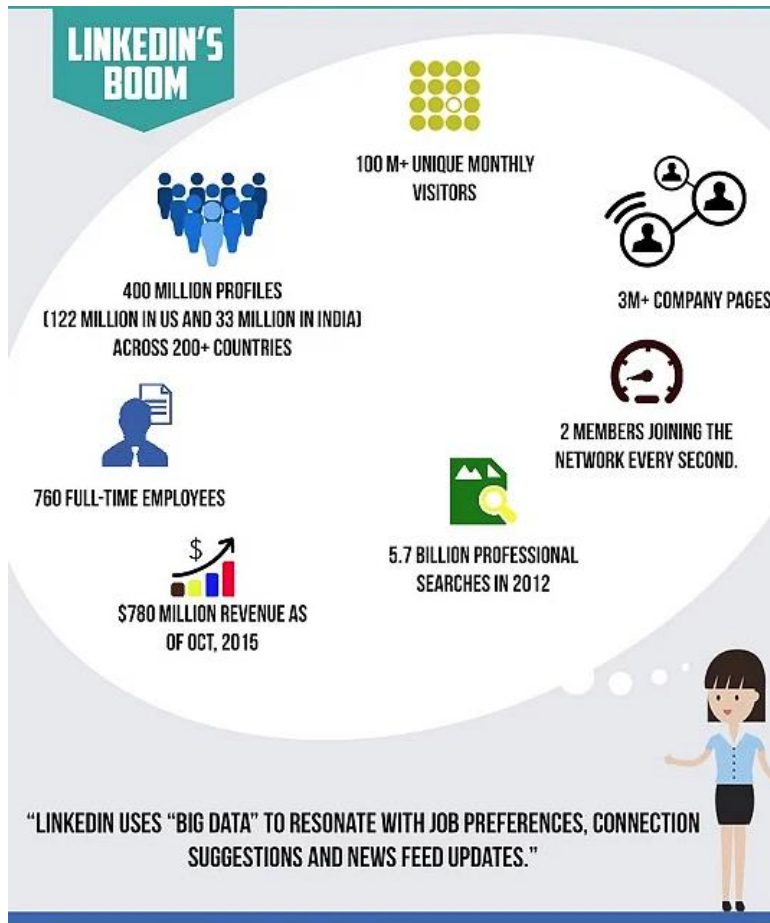


Homework : Big Data Analytics : Hadoop



2. **LinkedIn** ในฐานะ Social Network for Professional ที่ใหญ่ที่สุดในโลก ซึ่งมีข้อมูล User profile มากกว่า 400 ล้าน Profile , มีสมาชิกใหม่ 2 คนทุกวินาที , 5.7 พันล้านครั้งสำหรับการค้นหาพนักงาน (2012) โดย LinkedIn ใช้ Hadoop โดยอาศัยข้อมูลจาก User Profile มากมายมหาศาลนี้มารองรับการทำ Recommendation System มาวิเคราะห์ในเรื่องต่างๆ หลายรูปแบบ ส่งให้แก่ user แต่ละคน เช่น พี่เจอร์ People You May Know , Jobs you may be interested In , Skill Endorsement หรือ News Feed Update

Homework : Big Data Analytics : Hadoop



Homework : Big Data Analytics : Hadoop



3. Ebay หนึ่งในผู้เสนอรายใหญ่ US Ecommerce ใช้ Hadoop สำหรับการทำให้ SEO และการทำ market Research

อ้างอิง :

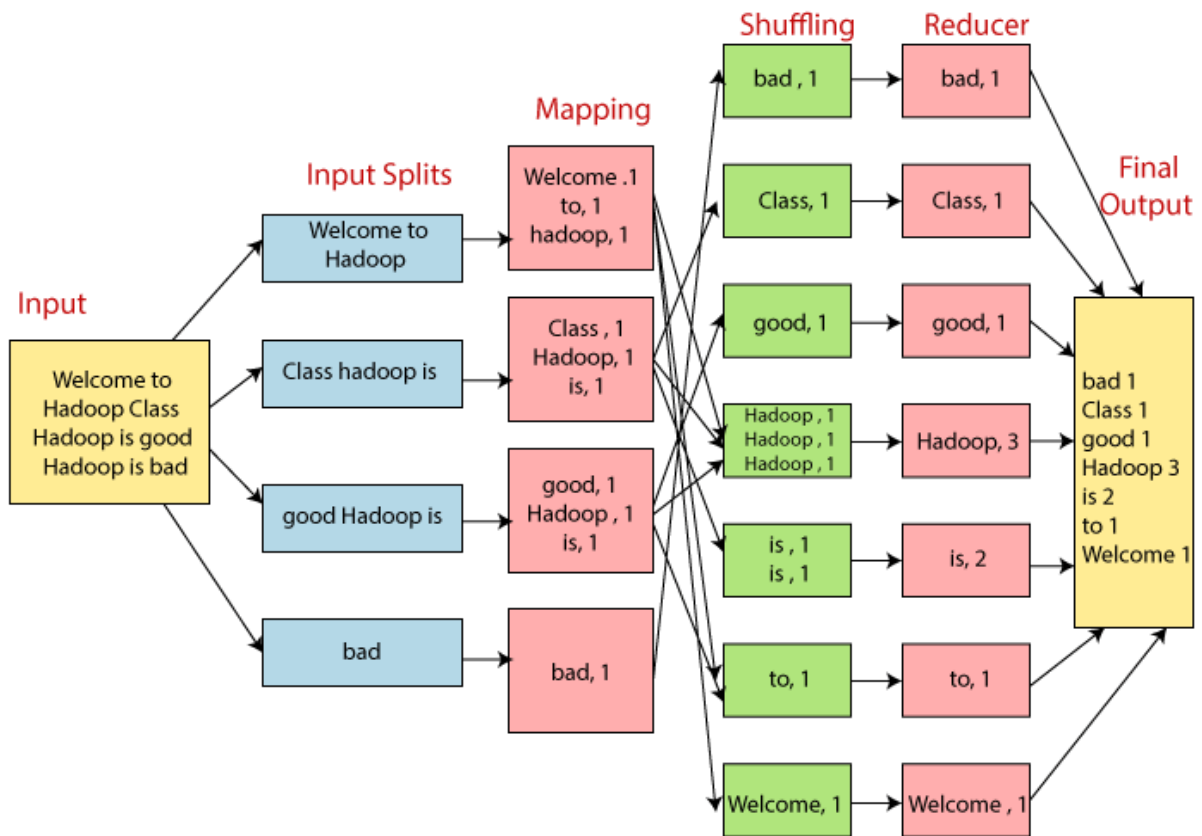
<https://www.projectpro.io/article/top-10-industries-using-big-data-and-121-companies-who-hire-hadoop-developers/69>

<https://www.projectpro.io/article/how-jpmorgan-uses-hadoop-to-leverage-big-data-analytics/142>

<https://www.projectpro.io/article/how-linkedin-uses-hadoop-to-leverage-big-data-analytics/229>

Homework : Big Data Analytics : Hadoop

Assignment 2 : ให้ยกตัวอย่างการประมวลผลข้อมูลที่ใช้หลักการของ MapReduce เพื่อวิเคราะห์ข้อมูล โดยแสดงออกมาในรูปของแผนภาพ ตั้งแต่การรับค่า Input/Splitting/Mapping/Shuffling/Reducing จนได้ผลลัพธ์ในลักษณะ Key-Value



อ้างอิง : <https://www.tutorialandexample.com/mapreduce-tutorial>

Homework : Big Data Analytics : Hadoop

Assignment 3 : ให้ทำการระบุนอย่างละ 3 ข้อว่าเมื่อไหร่ที่ควรนำ Hadoop มาใช้และเมื่อไหร่ที่ไม่ควรนำ Hadoop มาใช้งาน

No.	ควรใช้ Hadoop	ไม่ควรใช้ Hadoop
1	เมื่อต้องทำงานกับข้อมูลขนาดใหญ่ และมีแหล่งที่มาจากหลายแหล่ง (Huge size & Various source)	อยากเอาไปใช้วิเคราะห์ข้อมูล Real Time (เนื่องจาก Hadoop ทำงานในรูปแบบ Batch Processing จึงจะใช้เวลา response time สูง)
2	วางแผนเพื่อ Scale up Database ในอนาคต (เช่น มี Plan ที่จะ Scaleup จาก 10 TB เป็น 1 PB ใน 10 ปีข้างหน้า เป็นต้น)	กะเอามาใช้แทน Infrastructure ดั้งเดิม เช่น ใช้แทน Database ที่มีอยู่แล้ว
3	มีความจำเป็นต้องใช้ Framework ในการจัดการกับข้อมูลกลุ่มเดียวกันอย่างหลากหลาย เพราะ Hadoop มีให้ครบจบในที่เดียว	นำมาใช้กับข้อมูลขนาดเล็ก (เช่น เป็นข้อมูลที่สามารถจัดการด้วย Excel ได้ เช่น Server Daily Log) เพราะอาจมีความยุ่งยาก และใช้เวลานานโดยไม่จำเป็น

อ้างอิง : <https://www.edureka.co/blog/5-reasons-when-to-use-and-not-to-use-hadoop/>