บทที่ 6

ระบบอินพุต/เอาต์พุต และอุปกรณ์จัดเก็บข้อมูล

INPUT/OUTPUT AND STORAGE SYSTEMS

บทนำ

- การจัดเก็บและเรียกใช้ข้อมูลเป็นหนึ่งในงานหลัก ๆ ของระบบคอมพิวเตอร์
- ประสิทธิภาพการทำงานของระบบ I/O จะส่งผลต่อประสิทธิภาพการทำงานโดยรวมของระบบ โดยเฉพาะอย่างยิ่งเมื่อมีหน่วยความจำเสมือน
- หน่วยประมวลผลกลางที่เร็วที่สุดในโลกจะใช้ได้ไม่เต็มศักยภาพถ้าจะต้องเสียเวลารอข้อมูลที่จะประมวลผล
- โปรเซสส่วนใหญ่จะใช้เวลาในการทำงานของซีพียูประมาณ 70% และ เวลาในการรอข้อมูลจากดิสก์ประมาณ 30%
- ต้องพิจารณาว่าการอัพเกรดส่วนประกอบใดจะทำให้ระบบทำงานมีประสิทธิภาพมากที่สุด โดยใช้งบประมาณน้อยที่สุด

บทนำ

- เมื่อเพิ่มความเร็วของคอมพิวเตอร์โดยเพิ่มความเร็วของโปรเซสเซอร์อย่างเดียวโดยไม่สนข้อจำกัดความเร็วในการทำงานกับ I/O จะทำให้ประสิทธิภาพของคอมพิวเตอร์ไม่ดีเท่าที่ควร
- การทำงานของระบบคอมพิวเตอร์จำเป็นต้องมี
  - การรับข้อมูลจากอุปกรณ์รับเข้า (Input Device) เพื่อนำข้อมูลเข้าไปในระบบ
  - การนำข้อมูลออกไปบันทึกบนอุปกรณ์ส่งออก (Output Device)
- I/O : Input/Output จึงเป็นตัวกลางในการติดต่อระหว่างผู้ใช้งานกับคอมพิวเตอร์
- ระบบปฏิบัติการมีหน้าที่จัดการและควบคุมการทำงานของอุปกรณ์รับเข้า/ส่งออกตลอดจนอำนวยความสะดวกแก่ผู้ใช้ในการใช้งานอุปกรณ์ต่างๆ

ประเภทของอุปกรณ์ I/O

- **สามารถแยกได้ 2 ประเภท**
- **1. อุปกรณ์ชนิดข้อมูลเป็นลำดับ (stream)** อุปกรณ์ประเภทนี้ข้อมูลที่ส่งเข้าหรือออกจะเรียงตามลำดับก่อน-หลัง การแบ่งแยกข้อมูลทำได้โดยตรวจสอบลำดับของข้อมูล อุปกรณ์ประเภทนี้สามารถจัดการได้ง่าย เพียงแต่จัดการรับ-ส่งข้อมูลให้ถูกต้องก็เพียงพอแล้ว ตัวอย่างของอุปกรณ์ชนิดนี้ได้แก่
  - คีย์บอร์ด ซึ่งข้อมูลที่ส่งเข้ามาในระบบจะเป็นไปตามลำดับการกดคีย์ คีย์ใดถูกกดก่อนก็จะถูกส่งมาก่อน
  - เครื่องพิมพ์ ข้อมูลที่ถูกส่งออกไปก่อนก็จะถูกพิมพ์ก่อน ข้อมูลที่ถูกส่งไปทีหลังจะถูกพิมพ์ทีหลัง

ประเภทของอุปกรณ์ I/O

- **2. อุปกรณ์ชนิดข้อมูลไม่เป็นลำดับ (non-stream)** อุปกรณ์ประเภทนี้ข้อมูลที่ส่งและรับไม่ขึ้นอยู่กับลำดับการส่ง เราต้องอาศัยข้อมูลเพิ่มเติมเพื่อที่จะแยกแยะข้อมูลแต่ละตัว การจัดการอุปกรณ์ประเภทนี้ระบบปฏิบัติการจะต้องมีวิธีจัดการ โดยเฉพาะขึ้นอยู่กับลักษณะของอุปกรณ์ชนิดนั้น ๆ ตัวอย่างอุปกรณ์ชนิดนี้เช่น
  - จอภาพ ข้อมูลหรือตัวอักษรที่ส่งไปให้จอภาพ แสดงนั้นจะต้องส่งไปในตำแหน่งที่ถูกต้อง ตำแหน่งของตัวอักษรที่แสดงอยู่บนจอภาพแต่ละตัวจะมีแอดเดรสประจำตำแหน่งนั้น ๆ เมื่อเราส่งตัวอักษรไปยังแอดเดรสใด ตัวอักษรก็จะปรากฏอยู่บนจอภาพ ณ ตำแหน่งแอดเดรสนั้น ๆ

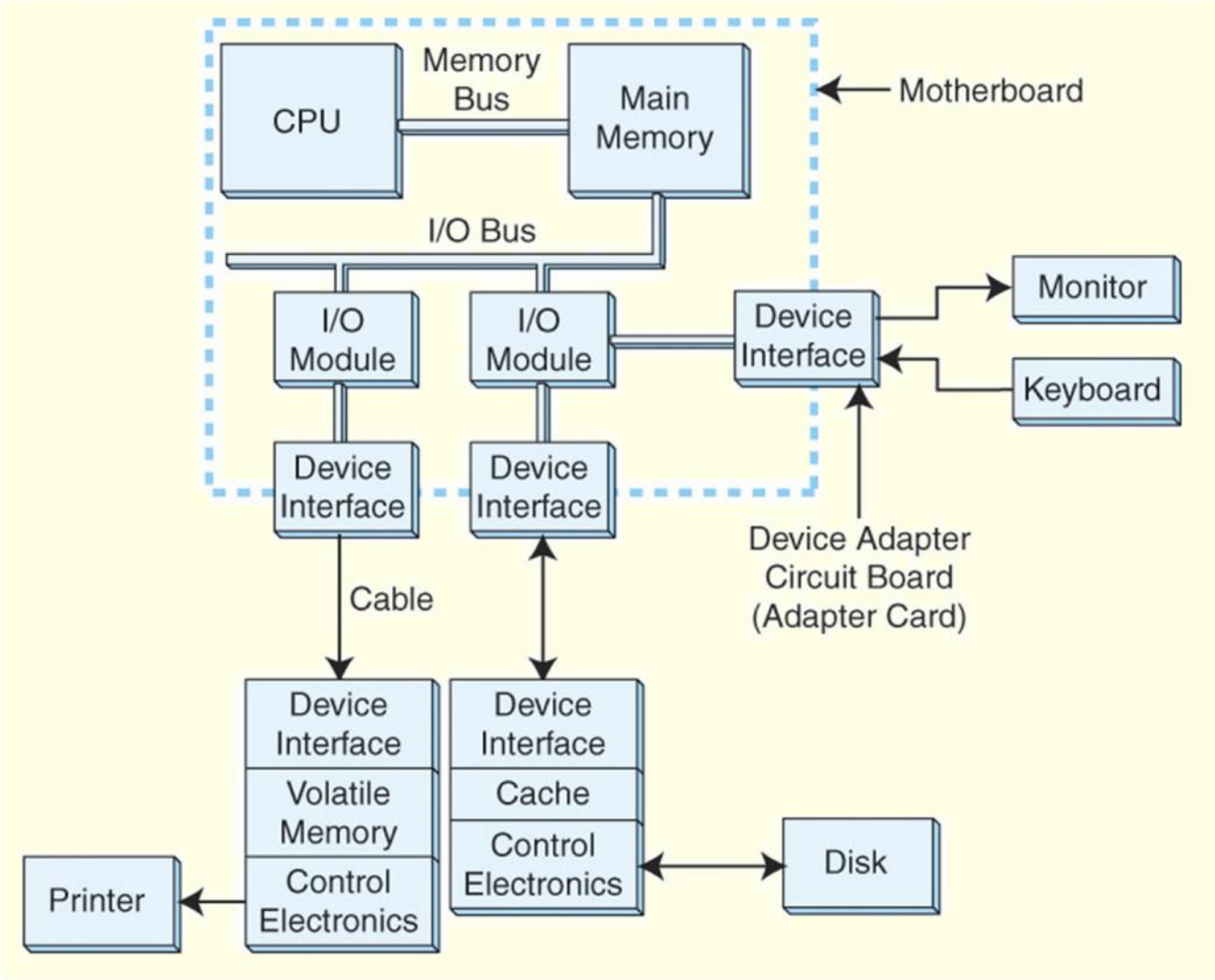
อัตราการถ่ายโอนข้อมูลของอุปกรณ์อินพุต/เอาต์พุตต่าง ๆ

ดีไวซ์	ไอโอ	อัตราการถ่ายโอนข้อมูล (กิโลไบต์ต่อวินาที)
คีย์บอร์ด	อินพุต	0.01
เมาส์	อินพุต	0.02
อินพุตเสียง	อินพุต	0.02
สแกนเนอร์	อินพุต	200
เอาต์พุตเสียง	เอาต์พุต	0.5
เครื่องพิมพ์อิงค์เจ็ต	เอาต์พุต	1.5
เครื่องพิมพ์เลเซอร์	เอาต์พุต	100-1000
การแสดงผล	เอาต์พุต	30000
เครือข่ายเฉพาะบริเวณ	อินพุต หรือเอาต์พุต	200-20000
อ็อพติคอลดิสก์	สื่อจัดเก็บข้อมูล	500-15000
เทปแม่เหล็ก	สื่อจัดเก็บข้อมูล	1000-15000
ดิสก์แม่เหล็ก	สื่อจัดเก็บข้อมูล	2000-60000
แหล่งข้อมูลวิดีโอที่ไม่ได้บีบอัดข้อมูล	อินพุต	1000
แหล่งข้อมูลเสียงที่ไม่ได้บีบอัดข้อมูล	อินพุต	100

สถาปัตยกรรมของ I/O

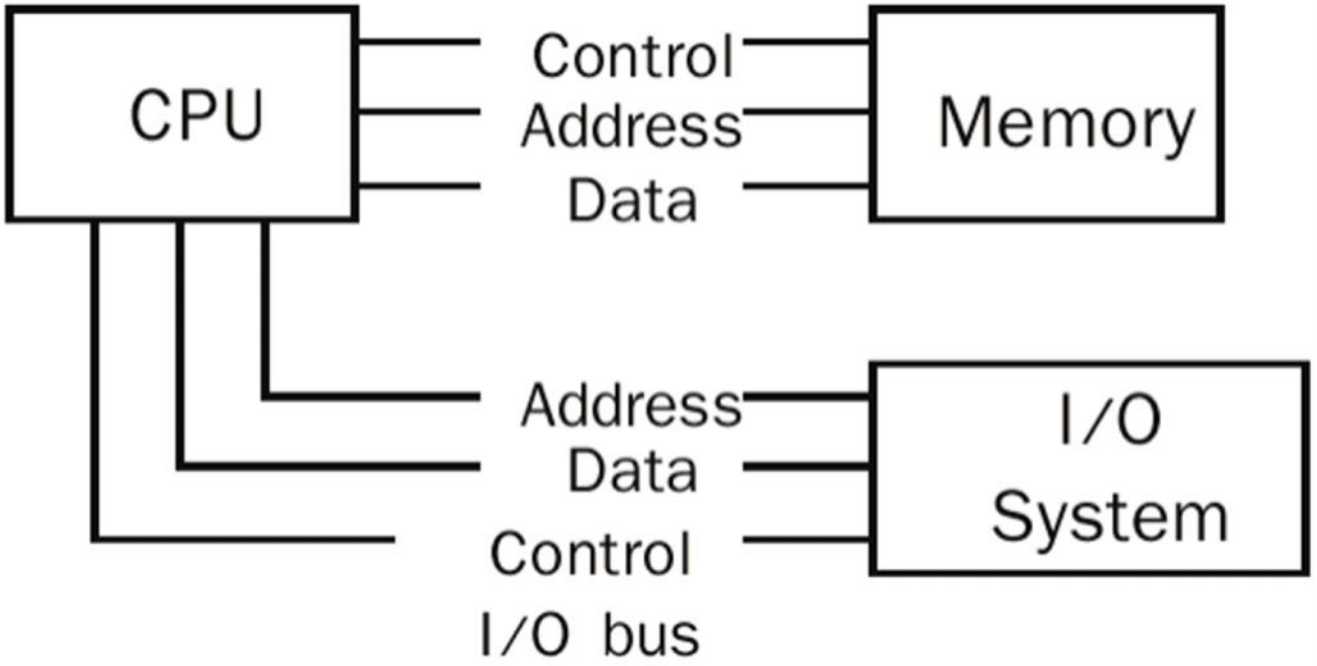
- I/O เป็นระบบที่ทำการโอนย้ายข้อมูลระหว่างอุปกรณ์ภายนอกกับระบบภายใน
- ระบบ I/O ประกอบด้วย
  1. ส่วนของหน่วยความจำหลักที่ใช้ติดต่อกับ I/O
  2. บัสต่าง ๆ ที่ช่วยในการย้ายข้อมูลระหว่างระบบภายในและภายนอก
  3. โมดูลควบคุมภายใน และอุปกรณ์ต่อพ่วง
  4. ส่วนเชื่อมต่อ (Interfaces) กับอุปกรณ์ภายนอก
  5. สายเชื่อมต่อหรือส่วนสื่อสารระหว่างคอมพิวเตอร์กับอุปกรณ์ต่อพ่วง

สถาปัตยกรรมของ I/O



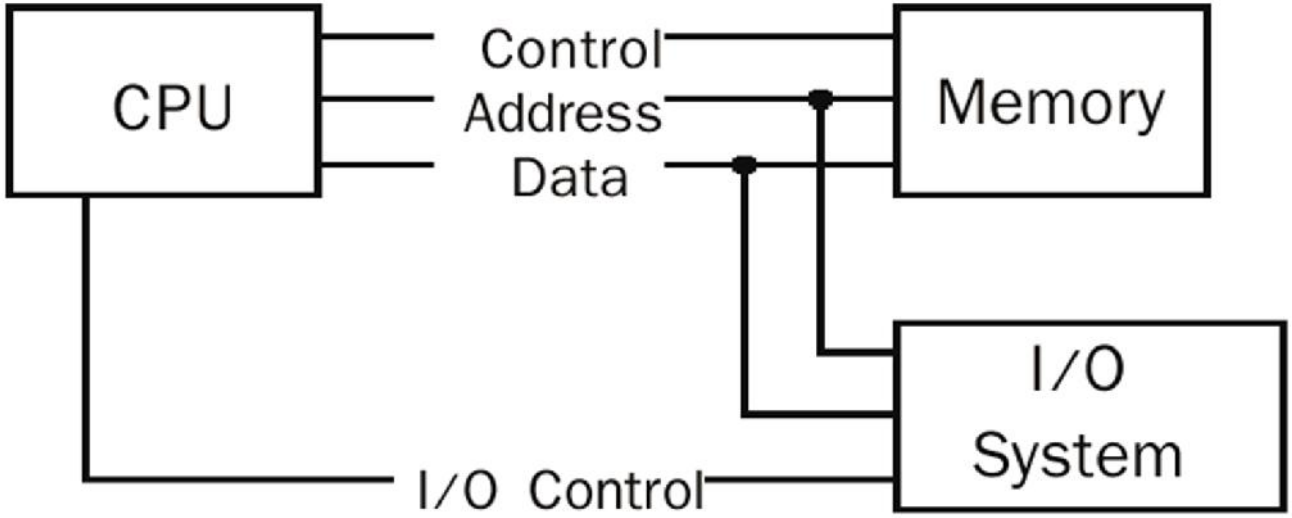
โครงสร้างของ I/O บัส

- **โครงสร้างของไอโอบัสมี 3 ประเภท**
- **Isolated I/O** จะมีชุดของแอดเดรสบัส ดาต้าบัส และคอนโทรลบัส ซึ่งเป็นชุดของระบบไอโอ และจะมีอีกชุดหนึ่งสำหรับหน่วยความจำที่แยกจากกัน



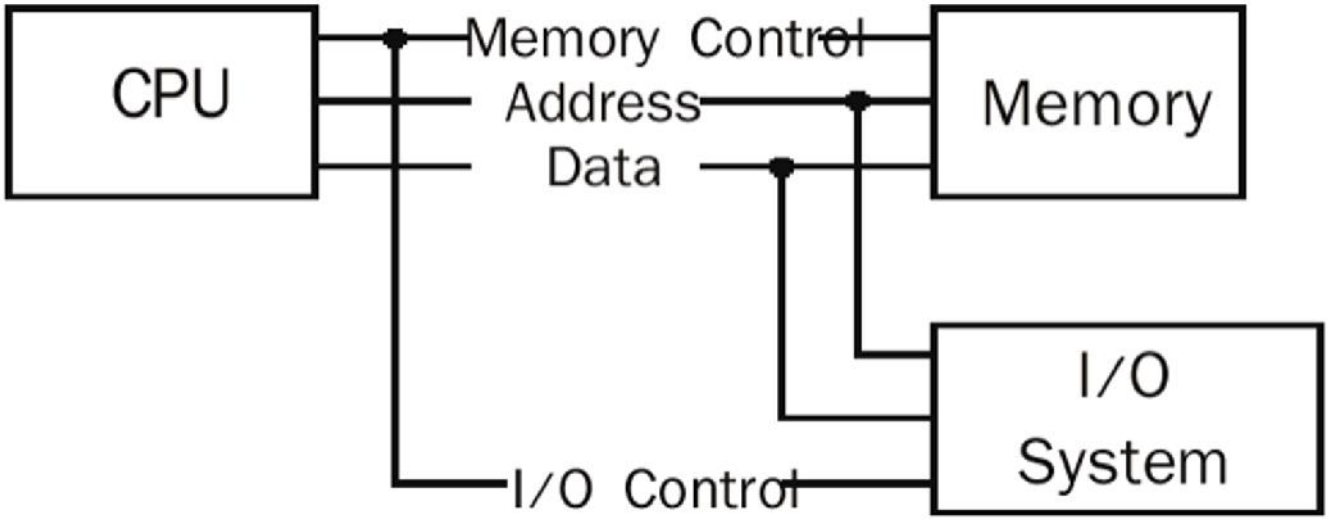
โครงสร้างของ I/O บัส

- **Shared I/O** จะใช้ address bus และ data bus ร่วมกันทั้งระบบไอโอ และหน่วยความจำ แต่มีสัญญาณควบคุมคนละเส้นสำหรับการอ่าน เขียน อินพุต และเอาต์พุต



โครงสร้างของ I/O บัส

- **Memory-map I/O** จะใช้บัสร่วมกันทั้งใน I/O และหน่วยความจำ ช่วงของ memory address จะถูกกันไว้สำหรับ I/O รีจิสเตอร์ด้วย



การควบคุม I/O

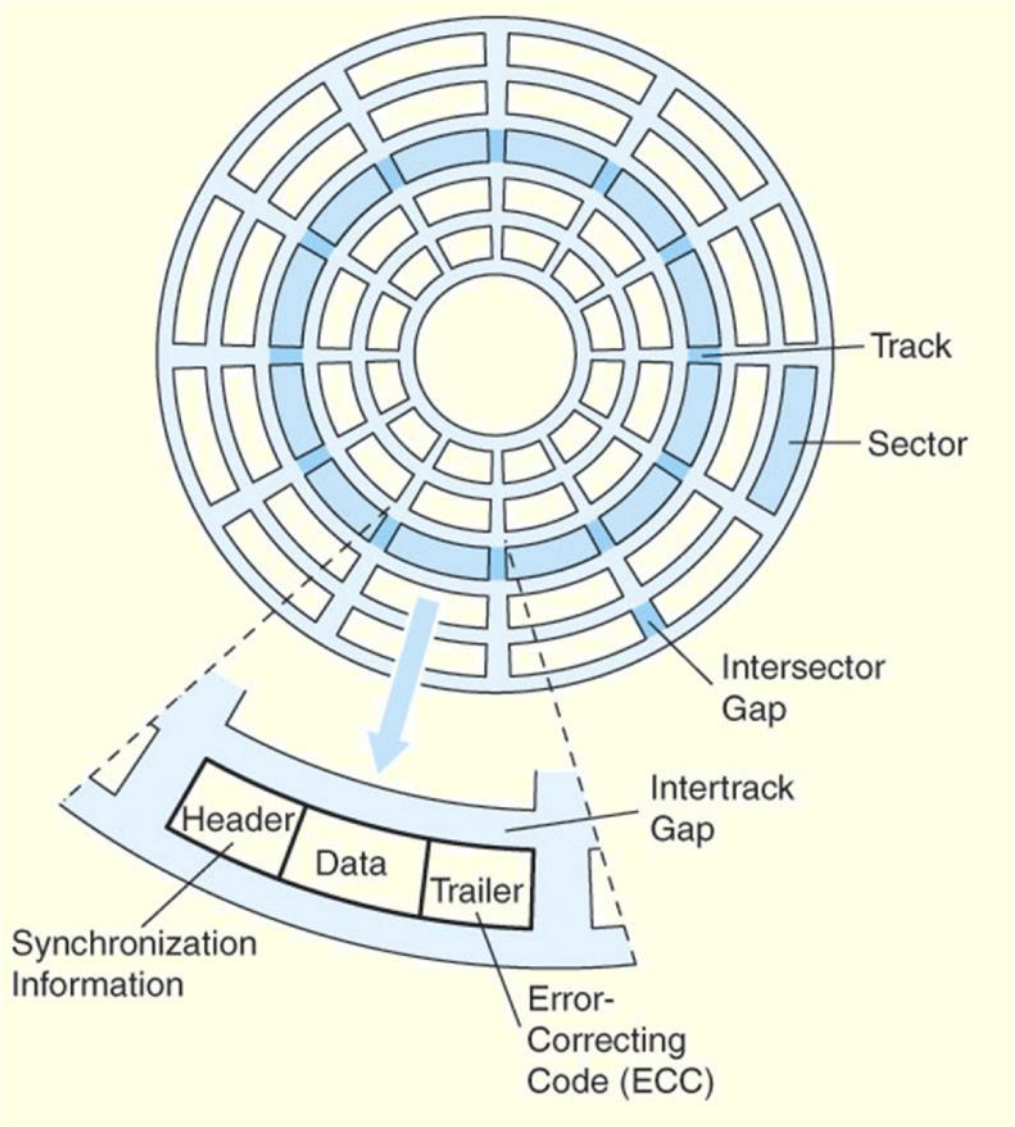
- **การควบคุม I/O มี 4 วิธี คือ**
- **ควบคุมด้วยโปรแกรม (Programmed I/O)** จะกำหนดให้รีจิสเตอร์แต่ละตัวเก็บแอดเดรสของอุปกรณ์ I/O แต่ละอุปกรณ์ และจะคอยตรวจสอบข้อมูลที่ส่งมาจากอุปกรณ์อย่างต่อเนื่อง
- **การอินเทอร์รัพต์ (Interrupt-driven I/O)** โดยอุปกรณ์จะส่งสัญญาณอินเทอร์รัพต์มาที่ CPU เมื่อต้องการส่งข้อมูล
- **การเข้าถึงหน่วยความจำโดยตรง (Direct Memory Access :DMA)** โดยเพิ่มชิปพิเศษในการประมวลผล I/O
- **หน่วยประมวลผลสำหรับ I/O โดยเฉพาะ**

เทคโนโลยีดิสก์แม่เหล็ก (Magnetic disks)

- ดิสก์แม่เหล็กสามารถจัดเก็บข้อมูลได้จำนวนมากและเข้าถึงได้อย่างรวดเร็ว
- ดิสก์ไดร์ฟเป็นอุปกรณ์ที่มีการเข้าถึงข้อมูลแบบสุ่ม (random access storage devices) เนื่องจากทุกบล็อกข้อมูลในดิสก์สามารถเข้าถึงได้ทันที โดยใช้เวลาเท่ากัน
- ปกติเมื่อผู้ใช้ซื้อดิสก์มาใหม่ จะต้องฟอร์แมต (Format) ก่อนที่จะนำไปบรรจุข้อมูล
- การฟอร์แมตดิสก์เป็นการแบ่งดิสก์ออกเป็นส่วน ๆ เพื่อให้คอมพิวเตอร์รู้ว่าตำแหน่งของข้อมูลอยู่ตรงไหนของดิสก์ โดยแบ่งออกเป็นเซ็กเตอร์ (Sector) และแทร็ก (Track)

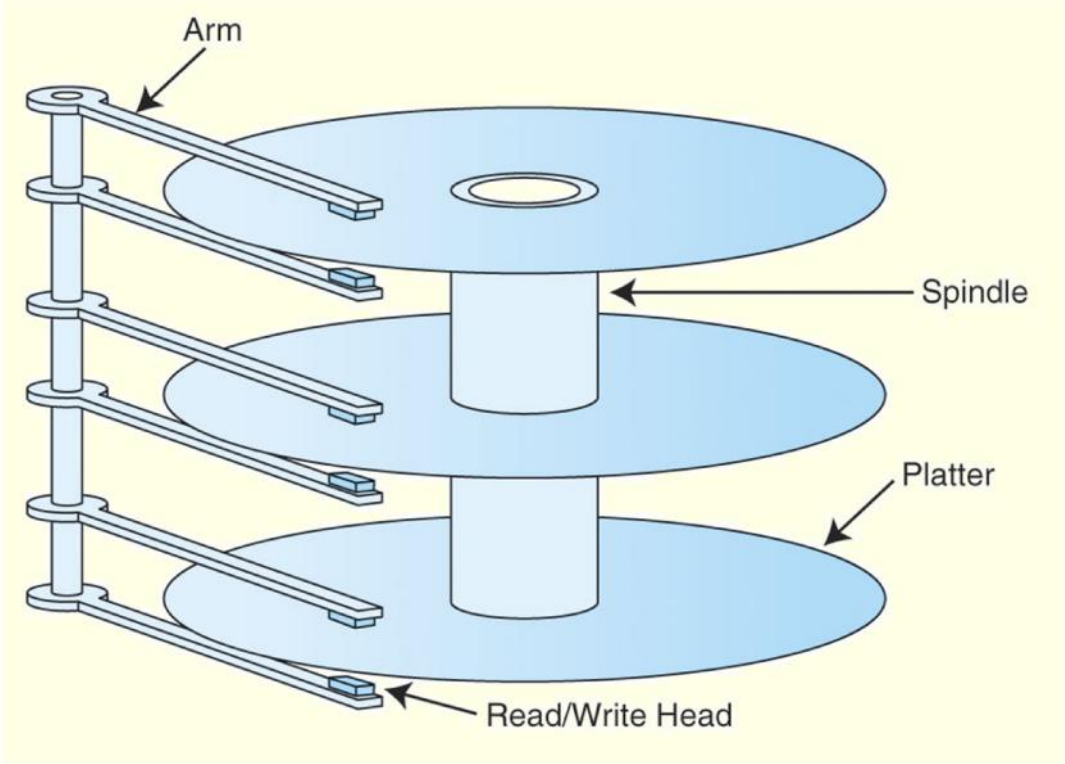
เทคโนโลยีดิสก์แม่เหล็ก (Magnetic disks)

- แทร็กของดิสก์จะเริ่มนับจากวงนอกสุดไปยังด้านใน โดยเริ่มจากแทร็ก 0
- แต่ละแทร็กจะถูกแบ่งออกเป็นเซ็กเตอร์
- การเก็บข้อมูลจะเก็บเป็นบล็อก โดย 1 บล็อกอาจมีมากกว่า 1 เซ็กเตอร์



เทคโนโลยีดิสก์แม่เหล็ก (Magnetic disks)

- แผ่นแม่เหล็กของฮาร์ดดิสก์จะถูกติดตั้งอยู่บนแกนหมุน
- หัวอ่าน/เขียนจะถูกติดตั้งอยู่บนแกนและแกว่งไปมาเพื่ออ่าน/เขียนข้อมูลในแทร็กที่ต้องการ
- แผ่นแม่เหล็กจะหมุนเพื่อให้ เซ็กเตอร์ (sector) ที่ต้องการอยู่ใต้หัวอ่าน
- บล็อกข้อมูลจะถูกกำหนดไว้ที่พื้นผิวตามไซลินเดอร์ (cylinder) และเซ็กเตอร์



เทคโนโลยีดิสก์แม่เหล็ก (Magnetic disks)

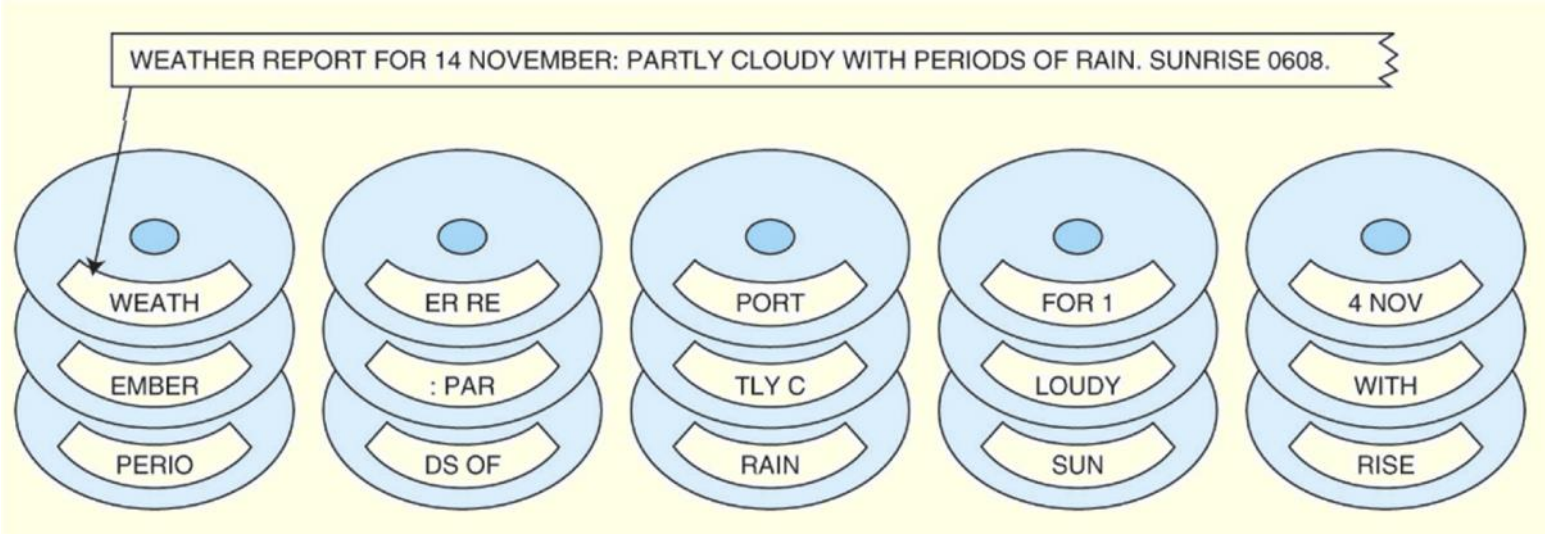
- มีคุณสมบัติทางไฟฟ้าหลายอย่างของดิสก์ที่กำหนดว่าสามารถเข้าถึงข้อมูลได้รวดเร็วเพียงใด
- **Seek time** คือเวลาที่ดิสก์เคลื่อนแกนหัวอ่านให้อยู่เหนือไซลินเดอร์ที่ต้องการ
- **Rotational delay** คือเวลาที่ดิสก์หมุนเพื่อให้เซ็กเตอร์ที่ต้องการอยู่ใต้หัวอ่าน
- เวลาเข้าถึงข้อมูล (access time) = Seek time + rotational delay
- **Transfer rate** คืออัตราการถ่ายโอนข้อมูลจากดิสก์
- **Average latency** คือฟังก์ชันความเร็วในการหมุน

เทคโนโลยีดิสก์แม่เหล็ก (Magnetic disks)

- ลักษณะทางกายภาพ และทางตรรกะ เช่น พื้นผิวแม่เหล็ก และการจัดโครงสร้างในการจัดเก็บข้อมูล เป็นข้อจำกัดด้านความจุในการเก็บข้อมูลของดิสก์
- การจัดโครงสร้างทางตรรกะในการเก็บข้อมูลในดิสก์มีหลายระบบ เช่น FAT, NTFS และ NFS เป็นต้น

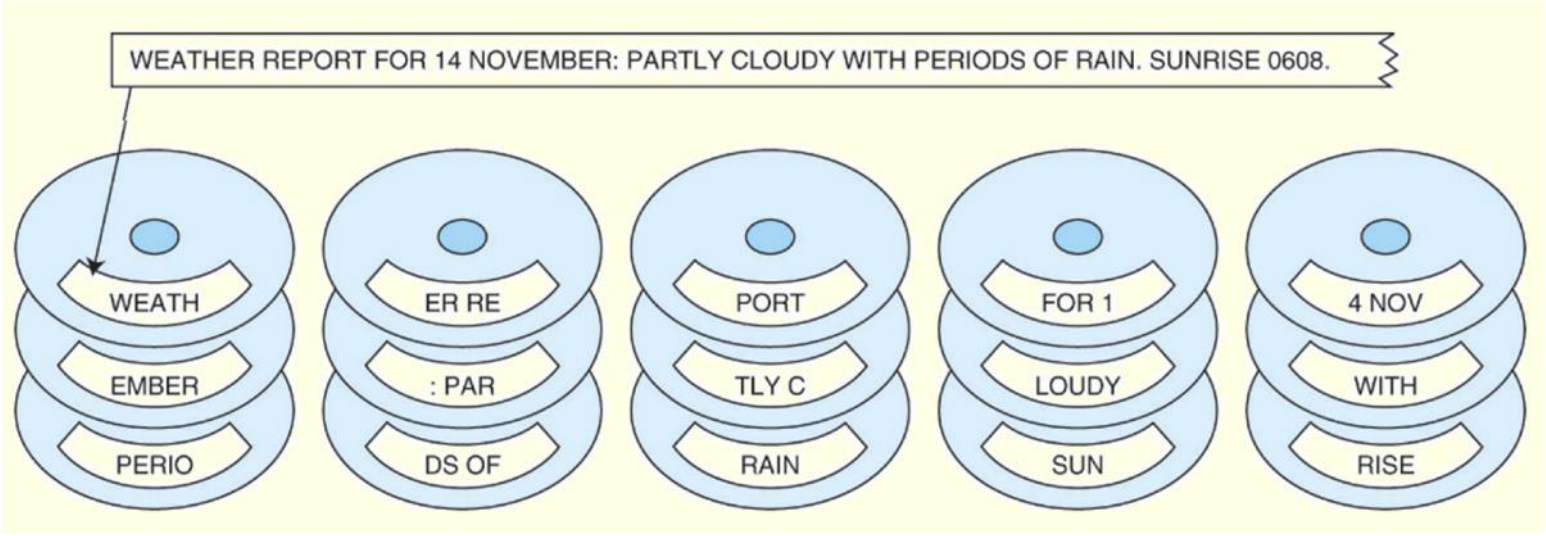
RAID

- RAID (Redundant Array of Inexpensive Disks) คือเทคโนโลยีการนำฮาร์ดดิสก์หลาย ๆ อันมาต่อเข้าด้วยกันเพื่อให้มองเห็นเป็นอันเดียว เพื่อวัตถุประสงค์ต่างๆ เช่น เพิ่มความน่าเชื่อถือของข้อมูล หรือเพิ่มประสิทธิภาพการอ่าน/เขียนข้อมูล



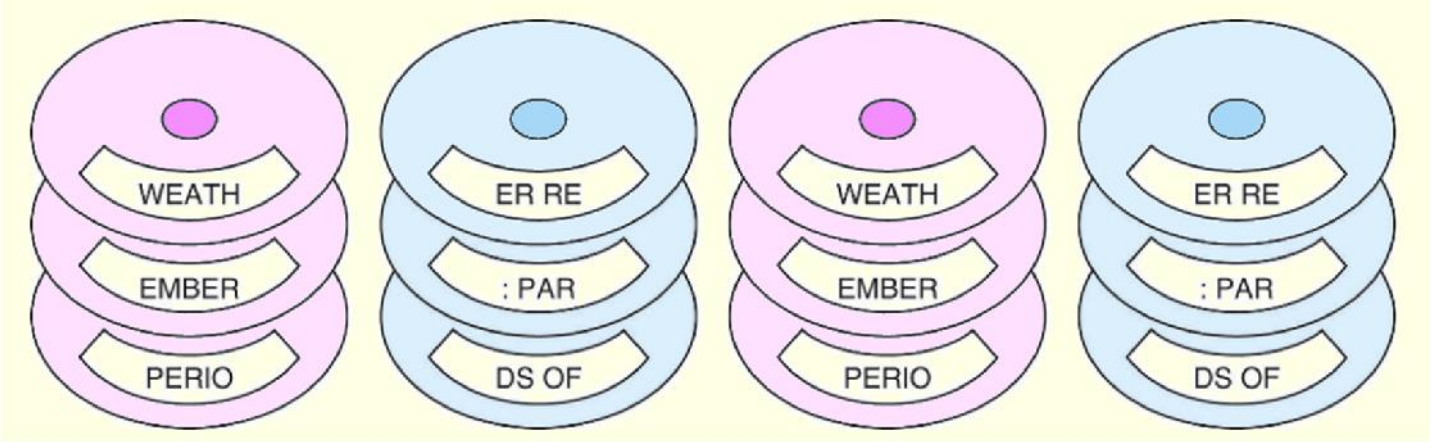
RAID

- RAID จะแบ่งเป็นระดับ (level) โดยที่แต่ละระดับจะมีจุดเด่นที่แตกต่างกัน
- **RAID 0 (striping)** ช่วยให้การบันทึกข้อมูลได้เร็วขึ้น แต่ถ้ามีฮาร์ดดิสก์เครื่องใดเครื่องหนึ่งเสีย จะทำให้ข้อมูลทั้งหมดไม่สามารถใช้งานได้ สมมติมีฮาร์ดดิสก์ 2 เครื่อง เครื่องละ 100 GB จะมีเนื้อที่ในการเก็บข้อมูลทั้งหมด 200 GB



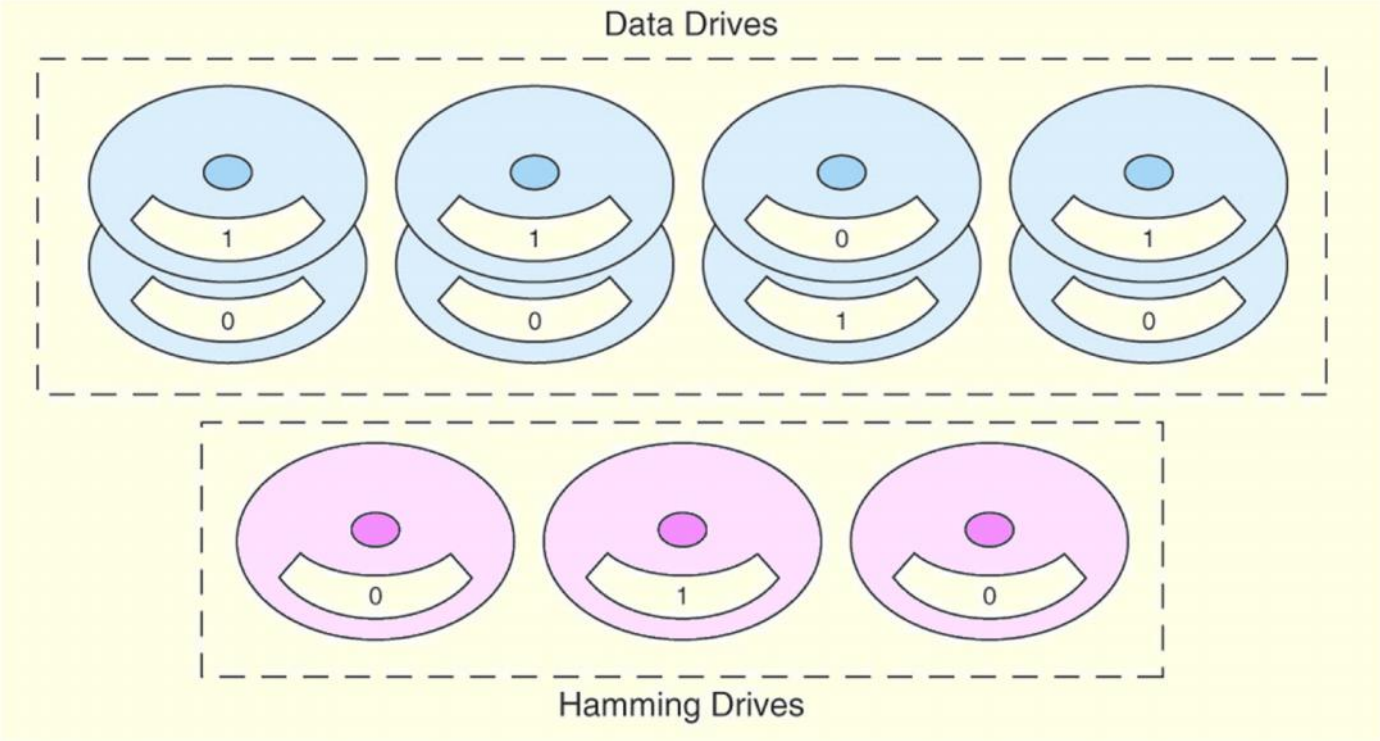
RAID

- **RAID 1 (disk mirroring)** ช่วยให้ข้อมูลมีความปลอดภัย ถ้าฮาร์ดดิสก์เครื่องใดเสีย อีกเครื่องหนึ่งก็จะทำงานแทนได้ สมมติมีฮาร์ดดิสก์ 2 เครื่อง เครื่องละ 100 GB จะมีเนื้อที่ในการเก็บข้อมูลทั้งหมด 100 GB อีก 100 GB จะเป็นดิสก์สำรอง
- จะต้องใช้ฮาร์ดดิสก์เป็นคู่ที่มีความจุเท่ากัน
- ข้อเสียคือเสียค่าใช้จ่ายมากขึ้น



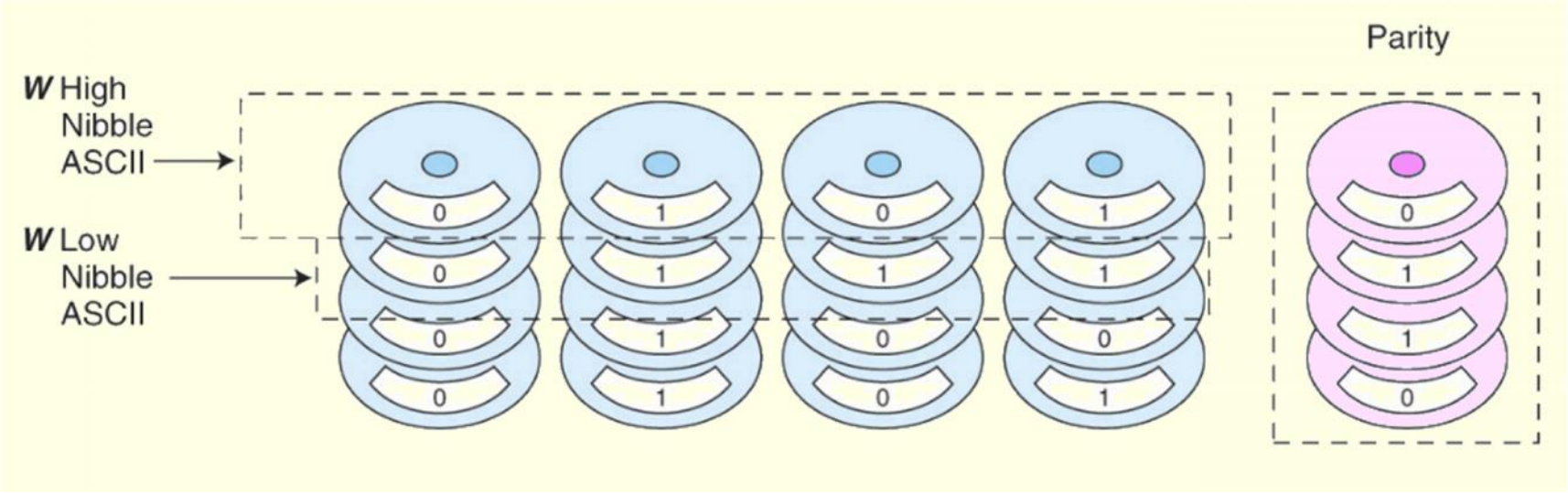
RAID

- **RAID 2** จะใช้ดิสก์สำหรับเก็บข้อมูล (Data drives) 1 ชุด และดิสก์สำหรับตรวจสอบข้อผิดพลาดอีก (Hamming drives) 1 ชุด
- ข้อเสียคือประสิทธิภาพไม่ดี และเสียค่าใช้จ่ายสูง



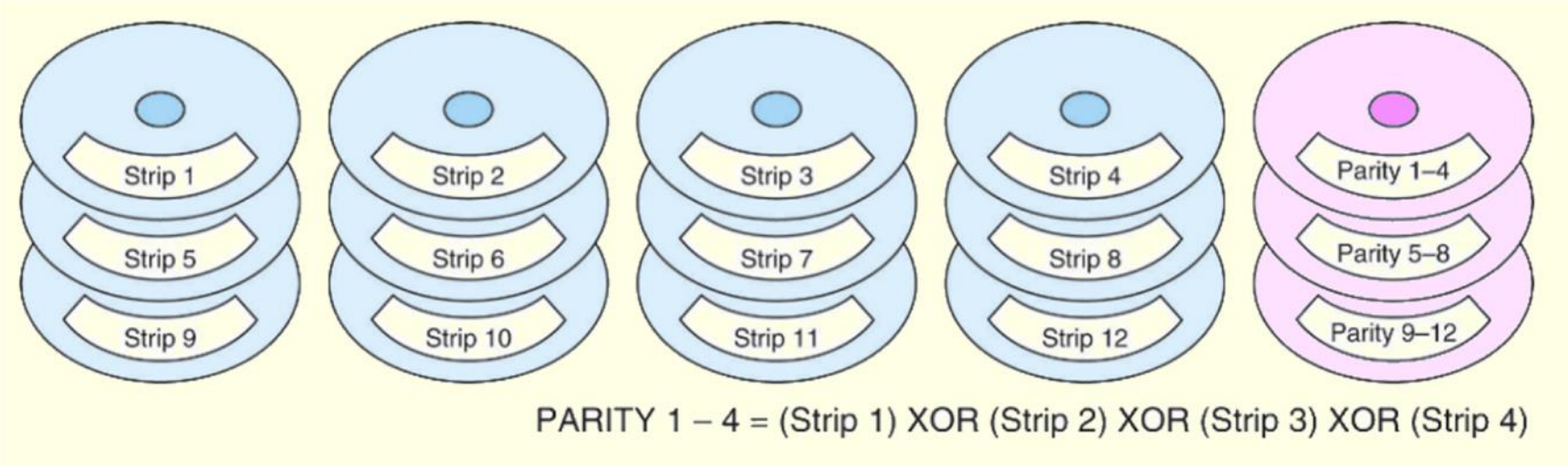
RAID

- **RAID 3 (N +1)** ต้องมีฮาร์ดดิสก์อย่างน้อย 3 ลูก เช่น ฮาร์ดดิสก์ 3 ลูก แต่ละลูกมีความจุ 200 GB จะสามารถจุข้อมูลได้ 400 Gb อีก 200 Gb เก็บไว้สำรองข้อมูลในกรณีที่ลูกแรกหรือ ลูกที่สองเสีย ฮาร์ดดิสก์ลูกที่ 3 จะทำงานให้แทนลูกที่เสียทันที ดังนั้น RAID 3 เหมาะสำหรับใช้ในงานที่มีการส่งข้อมูลจำนวนมาก ๆ เช่นงานตัดต่อ Video เป็นต้น



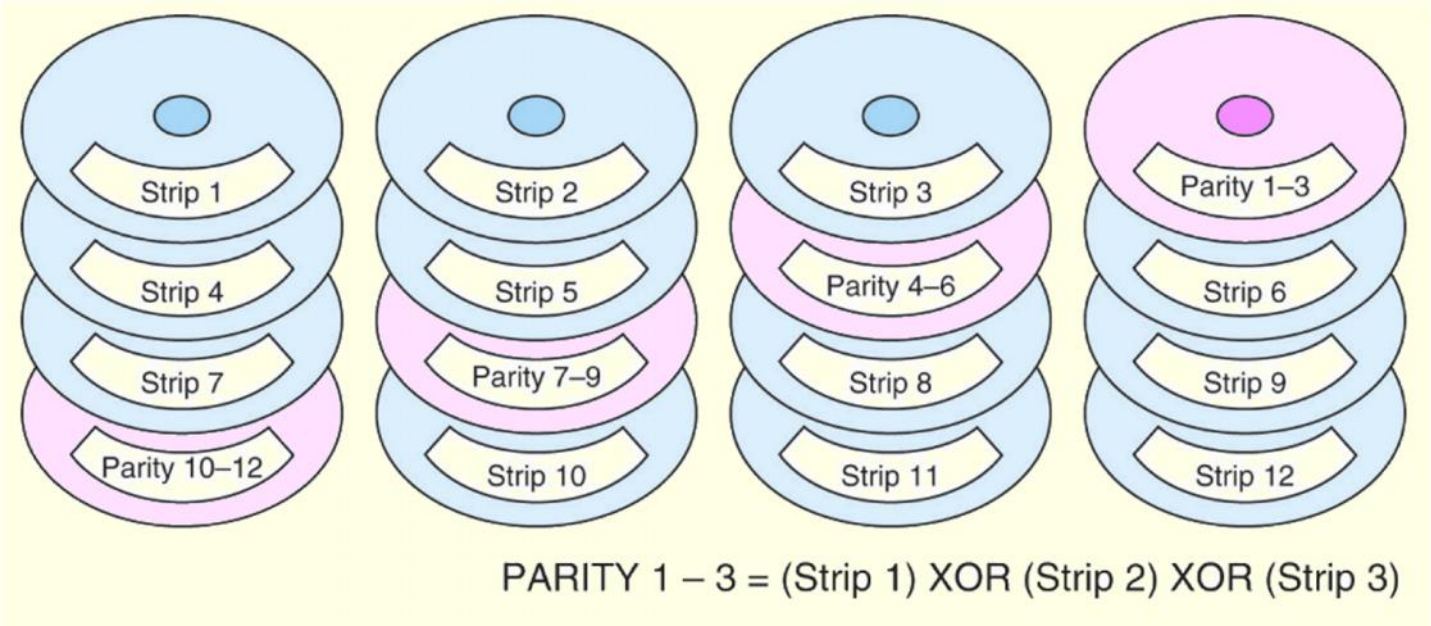
RAID

- **RAID 4** คล้ายกับ RAID ระดับ 3 แต่จะเน้นความสำคัญไปที่ประสิทธิภาพ การทำงานของ application ที่ต่างกันไป เช่น โปรแกรม Database ที่ต้องเกี่ยวข้องกับไฟล์ขนาดใหญ่ที่มีความต่อเนื่องกัน Harddisk แต่ละตัวทำงานได้อย่างอิสระต่อกันมากกว่า RAID 3 ช่วยให้การจัดการเก็บ และการเรียกอ่านข้อมูลที่มีขนาดใหญ่สามารถทำงานได้รวดเร็วยิ่งขึ้น



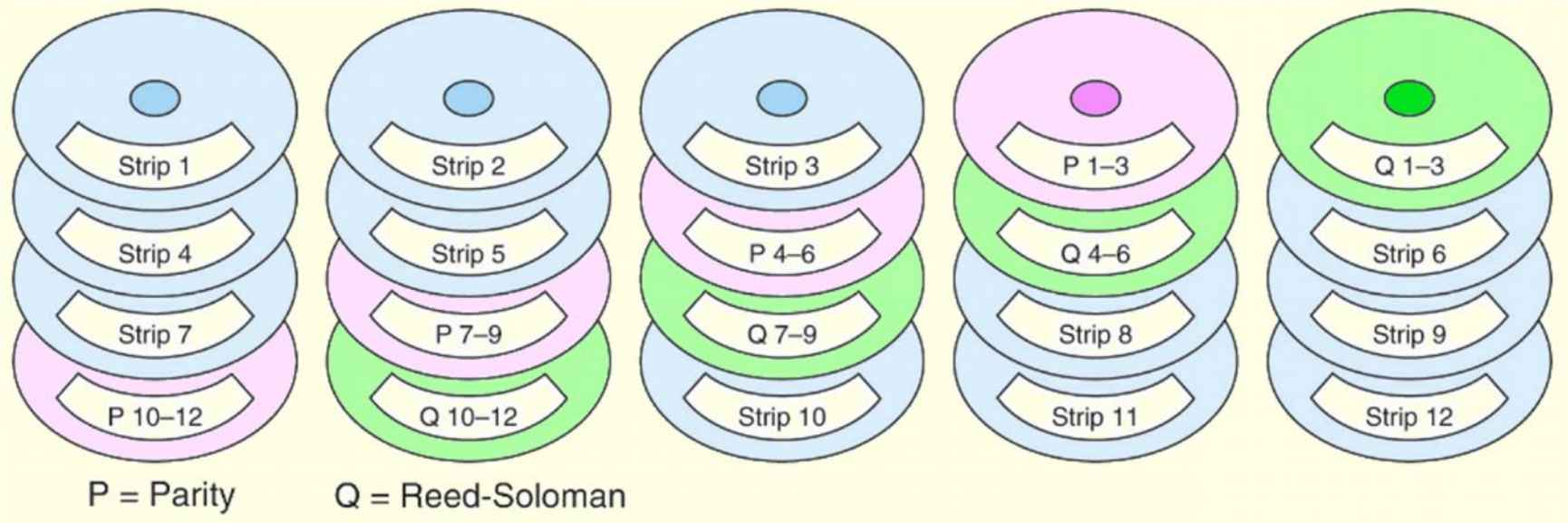
RAID

- **RAID 5 (N +1)** มีความสามารถเช่นเดียวกับ RAID 3 มีจุดเด่นคือ เทคโนโลยี Hot Swap ทำให้สามารถเปลี่ยน harddisk ในกรณีที่เกิดปัญหาได้ในขณะที่ระบบยังทำงานอยู่ เหมาะสำหรับงาน Server ต่าง ๆ ที่ต้องทำงานต่อเนื่อง
- นิยมใช้กันมากในระบบธุรกิจทั่วไป



RAID

- **RAID 6 (N +2)** อาศัยพื้นฐานการทำงานของ RAID 5 แต่จะดีกว่าตรงที่มี backup hard disk สองลูก และยอมให้ทำ Hot Swap ได้พร้อมกัน 2 ตัว ทำให้เหมาะกับงานที่ต้องการความปลอดภัยและเสถียรภาพของข้อมูลที่สูงมาก ๆ

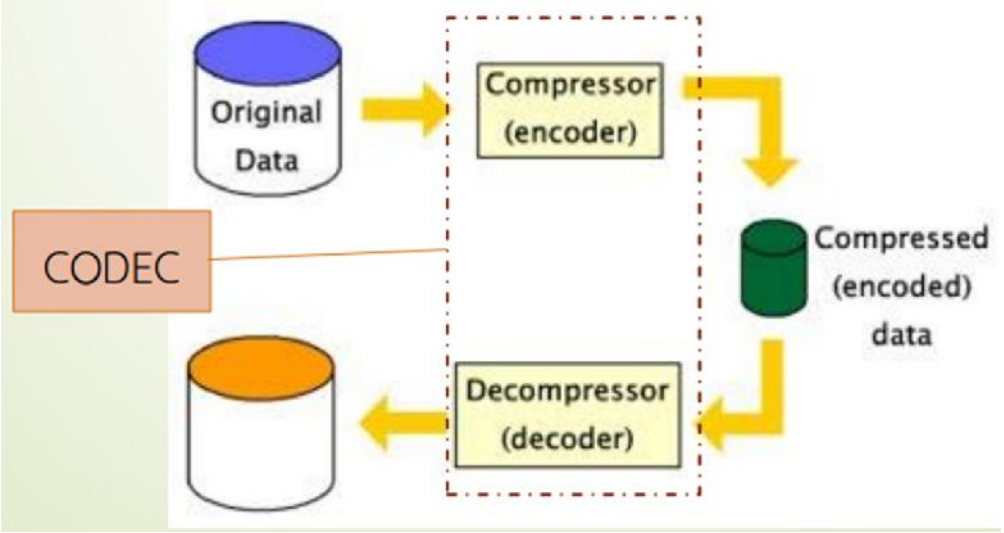


การบีบอัดข้อมูล (Data Compression)

- การบีบอัดข้อมูล คือกระบวนการเข้ารหัสข้อมูลที่เลือกเอาเฉพาะบิตที่จำเป็นเท่านั้น เพื่อประหยัดพื้นที่ในการจัดเก็บ และสามารถรับส่งข้อมูลผ่านเครือข่ายได้รวดเร็วยิ่งขึ้น
- การบีบอัดข้อมูลสำคัญกับระบบจัดเก็บข้อมูลเนื่องจากทำให้สามารถจัดเก็บข้อมูลลงในสื่อเก็บข้อมูลได้มากกว่าข้อมูลที่ไม่ได้บีบอัด
- การบีบอัดข้อมูลช่วยลดเวลาในการโอนย้ายข้อมูลผ่านอินเทอร์เน็ต ซึ่งจะลดขนาดข้อมูลและเวลาในการสื่อสารข้อมูล
- ข้อมูลดิบ (Raw Data) ที่ได้จากเครื่องมือต่าง ๆ นั้นมีขนาดของไฟล์ที่ใหญ่ ซึ่งใช้พื้นที่ในการจัดเก็บมาก จึงจำเป็นต้องมีการบีบอัด เพื่อลดขนาดของข้อมูลลง

การบีบอัดข้อมูล (Data Compression)

- CODEC (Coder-Decoder) เป็นกลไกสำหรับเข้ารหัสและถอดรหัสข้อมูลดิจิตอล เมื่อเข้ารหัสด้วย CODEC ชนิดใด ก็จะต้องถอดรหัสด้วย CODEC ชนิดเดียวกัน
- CODEC จะใช้อัลกอริทึมในการหาข้อมูลที่มีความซ้ำซ้อนกัน (Redundant) และลดความซ้ำซ้อนนั้นลง โดยไม่ส่งผลกระทบต่อคุณภาพของข้อมูล ข้อมูลที่ถูกบีบอัดแล้วก็จะมีนามสกุลของไฟล์แตกต่างกันไปตามวิธีการบีบอัด



ประเภทของการบีบอัดข้อมูล

- Lossless Compression ใช้อัลกอริทึมในการบีบอัดข้อมูลที่ไม่ทำให้ข้อมูลสูญหาย ข้อมูลจึงมีความสมบูรณ์เหมือนต้นฉบับ แต่ก็จะใช้พื้นที่ในการจัดเก็บมาก ตัวอย่างเช่น PNG, TIFF, GIF, AVI
- Lossy Compression จะตัดข้อมูลบางส่วนออกไปเพื่อลดขนาดของไฟล์ โดยข้อมูลที่ซ้ำซ้อนจะถูกตัดทิ้งอย่างถาวร ข้อดีคือขนาดข้อมูลจะลดลงมาก แต่ข้อเสียคือข้อมูลจะไม่ละเอียดเหมือนต้นฉบับ