

บทที่ 1

บทนำ

1.1 ปัญหาและที่มาของการศึกษา

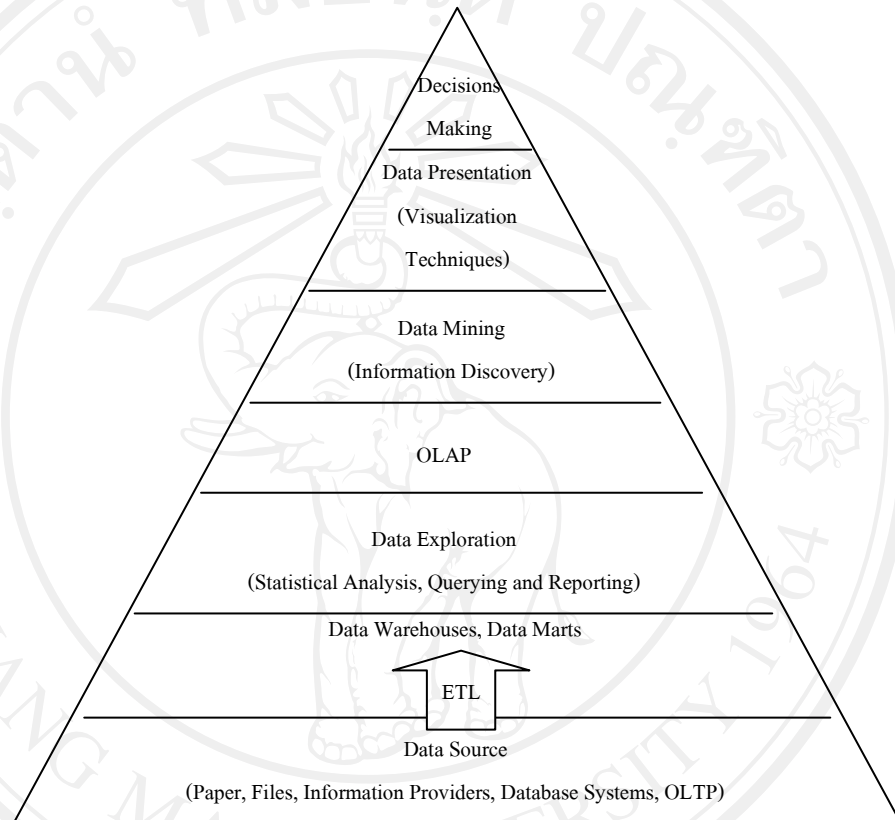
ในสภาวะการแข่งขันปัจจุบันการแข่งขันเพื่อชิงความเป็นหนึ่งมีความรุนแรงมากขึ้นในทุกธุรกิจ และเพื่อให้ตอบรับกับสถานการณ์เหล่านี้ องค์กรจึงต้องนำเทคโนโลยีคอมพิวเตอร์ และระบบจัดการสารสนเทศในลักษณะต่างๆเข้ามาประยุกต์ใช้งาน เพื่อให้ได้มาซึ่งความเป็นหนึ่งขององค์กร

แต่ทว่าการประยุกต์ใช้งานระบบจัดการสารสนเทศส่วนใหญ่ต้องพบกับปัญหามากมาย เช่น การจัดการกับปัญหาระบบฐานข้อมูล (Database System) ที่กระจัดกระจายอยู่ตามแผนกต่างๆ ซึ่งในบางครั้งมีความเข้ากันได้ น้อยมาก นอกจากนี้บางองค์กรยังมีการเจริญเติบโตของข้อมูลมากถึง 20 GB ต่อวัน [5] ซึ่งส่งผลโดยตรงกับความเร็วในการประมวลผลของข้อมูลอย่างชัดเจน นอกจากนี้การจัดสร้างระบบฐานข้อมูลสำหรับการดำเนินการธุรกิจแบบพื้นฐานนั้น อาจยังไม่ได้ตอบคำถามของผู้บริหาร ที่ต้องการนำข้อมูลไปตัดสินใจในเชิงธุรกิจการแข่งขัน การตลาด และการเงิน ซึ่งปัญหานี้สามารถใช้แนวทางของ ปัญญาธุรกิจ (Business Intelligence) เข้ามาบรรเทาหรือแก้ไขได้

ปัญญาธุรกิจ คือ กระบวนการในการใช้ระบบฐานข้อมูลในคลังข้อมูล (Data Warehouse) มาจัดทำรายงานในรูปแบบต่างๆ ที่เหมาะสมกับมุมมองในการวิเคราะห์ และตรงตามความต้องการของผู้ใช้งาน เพื่อใช้ในการตัดสินใจ และวางแผนในทางธุรกิจ โดยปัญญาธุรกิจจะประกอบไปด้วยระบบฐานข้อมูล และโปรแกรมประยุกต์ (Application Program) ด้านการวิเคราะห์ข้อมูลมากมายหลายระบบ เช่น

- คลังข้อมูล (Data Warehouse)
- ดาตามาร์ท (Data Mart)
- การทำเหมืองข้อมูล (Data Mining)
- การวิจัยดำเนินการและวิธีเชิงตัวเลข (Operations Research and Numerical Methods)

- การประมวลผลการวิเคราะห์แบบออนไลน์ (Online Analytical Processing) หรือโอแอลเอพี (OLAP)
- การค้นหา และรายงาน (Search and Report)



รูปที่ 1.1 ภาพความสัมพันธ์ของปัญญาธุรกิจ

จากรูปที่ 1.1 จะเห็นได้ว่าการส่งผ่านข้อมูลต้นทาง (Data Sources) ไปยังคลังข้อมูลนั้นจะผ่านกระบวนการ อีทีแอล (Extract, Transform, and Load หรือ ETL) ซึ่งเป็นกระบวนการที่นำข้อมูลต้นทางหลายๆ แหล่งที่แตกต่างกัน มาจัดรูปแบบให้ถูกต้องสอดคล้องกัน แล้วส่งผ่านไปยังคลังข้อมูล [8] เพื่อทำการประมวลผลที่ชั้นโอแอลเอพี ให้พร้อมใช้งานในชั้นสำรวจข้อมูล (Data Exploration) นำเสนอข้อมูล (Data Presentation) และ การตัดสินใจ (Decision Making) ในขั้นตอนสุดท้าย จากกระบวนการที่ได้กล่าวมานี้ จะเห็นได้ว่าในชั้นคลังข้อมูล และ โอแอลเอพีมีความสำคัญต่อปัญญาธุรกิจเป็นอย่างมาก เพราะว่าถ้ากระบวนการทั้งสองทำงานผิดพลาด หรือเกิดความล่าช้าขึ้น จะส่งผลให้ประสิทธิภาพของปัญญาธุรกิจลดลงได้ การพิจารณาประสิทธิภาพของคลังข้อมูลนั้น จะพิจารณาจากการออกแบบระบบฐานข้อมูล การออกแบบการเชื่อมโยงกันระหว่างตารางใน

ฐานข้อมูล และการใช้งานดัชนี (Index) ต่างๆ นอกจากนี้ยังต้อง พิจารณาถึงความเร็วในการประมวลผลการวิเคราะห์แบบออนไลน์อีกด้วย ซึ่งความเร็วนี้เป็นส่วนสำคัญอย่างมาก คือ ถ้าเวลาที่ใช้ในการประมวลผลในแต่ละครั้งใช้เวลานาน ผลการวิเคราะห์ที่ได้อาจจะมิประโยชน์น้อยมาก เมื่อเทียบกับต้นทุนที่สูญเสียไป ซึ่งเป็นผลเสียอย่างมากในเชิงธุรกิจ [8]

ทางเลือกหนึ่งในการเพิ่มประสิทธิภาพของโอแอลเอพี คือ การเพิ่มความเร็วในการประมวลผล ไม่ว่าจะเป็นการเพิ่มประสิทธิภาพของฮาร์ดแวร์ (Hardware) หรือซอฟต์แวร์ (Software) ซึ่งทั้งสองวิธีนี้ต่างมีข้อดีข้อเสียแตกต่างกันไป เช่น ถ้าเลือกที่จะเพิ่มประสิทธิภาพด้วยฮาร์ดแวร์อาจจะต้องสูญเสียเงินในปริมาณที่มาก และมีอายุการใช้ในเวลาที่จำกัดด้วยสาเหตุด้านการเปลี่ยนแปลงของเทคโนโลยี (Technology) อีกทั้งยังอาจประสบปัญหาเดิมเนื่องจากการเจริญเติบโตของข้อมูลที่รวดเร็วจนเช่นเดิม แต่ถ้าเลือกการเพิ่มประสิทธิภาพด้วยซอฟต์แวร์อาจไม่ต้องเสียเงินมาก แต่จำเป็นต้องดำเนินการด้วยบุคลากรด้านคอมพิวเตอร์ที่มีความรู้ ความเข้าใจในเรื่องการปรับแต่งฐานข้อมูล ทั้งนี้อาจพิจารณาได้ว่าการเพิ่มประสิทธิภาพทางด้านซอฟต์แวร์เป็นแนวทางที่เสียค่าใช้จ่ายต่ำกว่า เมื่อเทียบกับการเพิ่มประสิทธิภาพทางด้านฮาร์ดแวร์ ซึ่งการเพิ่มประสิทธิภาพทางด้านซอฟต์แวร์นั้นกระทำได้โดยการคิดค้น พัฒนา หรือปรับปรุงวิธีการเข้าถึงข้อมูล โดยเลือกใช้วิธีการที่เหมาะสม และได้ประสิทธิภาพสูงที่สุด

วายทรี (Y tree) เป็นดัชนีประเภทหนึ่งที่ถูกออกแบบมาเพื่อแก้ปัญหาการแทรกข้อมูล (Insertion) ที่มีอยู่จำนวนมากเข้าสู่ระบบจัดการฐานข้อมูลโดยใช้เวลาน้อยที่สุด จากการทดลองเมื่อเปรียบเทียบกับดัชนีบีพลัสทรี (B+ tree) แล้ว ดัชนีวายทรีมีสมรรถนะในการแทรกข้อมูลสูงกว่าดัชนีบีพลัสทรีถึง 100 เท่า แต่ยังคงสมรรถนะในการสอบถาม (Query) และการลบ (Deletion) ที่ใกล้เคียงกันกับ ดัชนีบีพลัสทรี [5] ซึ่งสมรรถนะของดัชนีวายทรีนี้ สอดคล้องกับปัญหาของปัญญาธุรกิจในส่วนคลังข้อมูลที่มีมีการเพิ่มข้อมูลลงในระบบฐานข้อมูลจำนวนมาก

โดยทั่วไปปริมาณข้อมูลที่มากมายซึ่งถูกจัดเก็บลงในคลังข้อมูลจะถูกจัดเก็บลงจานบันทึก (Disk) ซึ่งเป็นประเภทจานแม่เหล็ก (Magnetic Disk) อย่างไรก็ตามในช่วงเวลาประมาณ 2-3 ปีที่ผ่านมา แฟลชไดรฟ์ (Flash Drive) ซึ่งเป็นแหล่งเก็บข้อมูลประเภท Electrically-erasable Programmable Read-only Memory (EEPROM) เริ่มได้รับความนิยมในการนำมาเก็บข้อมูลมากขึ้นเรื่อยๆ ด้วยเหตุผลของราคาที่ถูกลง ประสิทธิภาพที่เพิ่มมากขึ้น และความจุที่สูงขึ้นเรื่อยๆ ดังแสดงในรูปที่ 1.2 ซึ่งแสดงให้เห็นว่าราคาของแฟลชไดรฟ์ต่อกิกะไบต์ (Gigabyte) เริ่มลดลงอย่างรวดเร็ว

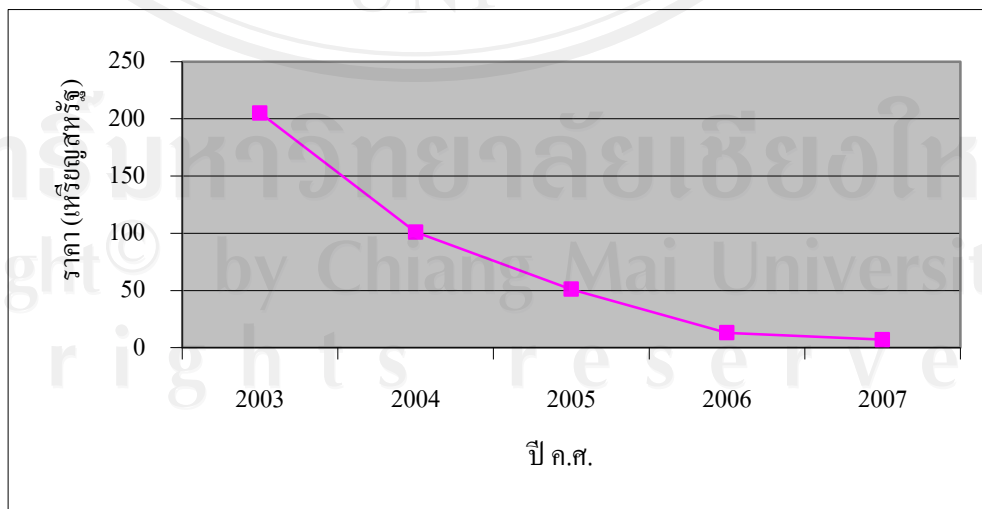
ในปี ค.ศ. 2004-2006 และก็เริ่มลดลงอย่างช้าๆ ในปี ค.ศ. 2006-2007 [4] ซึ่งตรงกันข้ามกับความจุที่เพิ่มขึ้นเรื่อยๆ อย่างรวดเร็ว

แฟลชไดรฟ์ได้ถูกนำไปใช้อย่างกว้างขวาง กับอุปกรณ์หลายๆ ชนิดมาก เช่น Thumb Drives แหล่งเก็บข้อมูลของกล้องถ่ายภาพดิจิทัล หรือฮาร์ดดิสก์ของคอมพิวเตอร์พกพา จากแนวโน้มดังกล่าว อาจกล่าวได้ว่าแฟลชไดรฟ์จะกลายมาเป็นแหล่งเก็บข้อมูลหลักของระบบคอมพิวเตอร์ในอนาคตอีกไม่ช้า [11]

อย่างไรก็ตามลักษณะพิเศษประการหนึ่งที่ทำให้แฟลชไดรฟ์ได้รับความนิยมเพิ่มขึ้น คือ ความเร็วในการเข้าถึงข้อมูลดังแสดงในตารางที่ 1.1 ซึ่งแสดงให้เห็นถึงเวลาในการอ่าน เขียน ลบ ของ Samsung Flash Memory Chip [9] โดยมีอัตราความเร็วการเขียนต่อการอ่านอยู่ที่ 1:2.5 และอัตราความเร็วในการลบต่อการอ่านเป็น 1:18.7

ตารางที่ 1.1 เวลาในการเข้าถึงข้อมูลของแฟลชไดรฟ์ และจานแม่เหล็ก

Media	Access time		
	Read(2KB)	Write(2KB)	Erase(128KB)
Magnetic	12.7 ms	13.7 ms	N/A
Flash Memory	80 μ s	200 μ s	1.5 ms



รูปที่ 1.2 ราคาของ แฟลชไดรฟ์ต่อกิกะไบต์

ข้อแตกต่างประการหนึ่งของ แฟลชไดรฟ์กับจานแม่เหล็ก ก็คือการทำงานแม่เหล็กต้องใช้การหมุนของจานแม่เหล็กเป็นตัวขับเคลื่อนในการอ่าน และเขียนข้อมูล ซึ่งมีผลทำให้การอ่านแบบสุ่ม (Random Read) ของแฟลชไดรฟ์มีความเร็วกว่าถึง 2 เท่า แต่แฟลชไดรฟ์จะมีข้อด้อยเมื่อเปรียบเทียบกับจานแม่เหล็กตรงที่การเขียนแบบสุ่ม (Random write) ซึ่งจะต้องทำการลบข้อมูลในตำแหน่งที่จะเขียนใหม่ก่อนทุกครั้ง จึงทำให้กระบวนการนี้ช้ากว่าจานแม่เหล็ก 2 เท่า [11] เช่นกัน ดังในตารางที่ 1.2

ดังที่ได้กล่าวมาแล้วว่าแฟลชไดรฟ์ แม้จะยังมีสมรรถนะที่ไม่ดีนักในกรณีการเขียนข้อมูล แต่หากพิจารณาจากรูปที่ 1.2 แล้วจะเห็นว่า แนวโน้มของการเปลี่ยนแปลงในด้าน ราคา และความจุของแฟลชไดรฟ์มีแนวโน้มพัฒนาไปในทางที่ดีมากขึ้น ประกอบกับสมรรถนะด้านการเขียนของแฟลชไดรฟ์มีสมรรถนะที่สูงมากเมื่อเทียบกับจานแม่เหล็ก และเป็นไปได้ว่าในที่สุดแฟลชไดรฟ์จะมาแทนจานแม่เหล็กในอีกไม่กี่ปีนี้

ตารางที่ 1.2 การเปรียบเทียบเวลาในการเข้าถึงข้อมูลแบบสุ่ม (จำนวนอินพุตและเอาต์พุตต่อวินาที)

	Samsung 32GB SSD	Intel Extreme 32/64GB SSD	Seagate 7200RPM SATA Magnetic Disk
Random Read	3100	35000	100
Random Write	25	3300	110

1.2 แนวทางการแก้ปัญหา

ในวิทยานิพนธ์นี้ได้เสนอการแนวทางการวิเคราะห์การประมาณค่าสมรรถนะของฟังก์ชันการแทรก และฟังก์ชันการสอบถามของดัชนีวายทรีบนแฟลชไดรฟ์ โดยสมรรถนะที่กล่าวถึงในวิทยานิพนธ์ฉบับนี้หมายถึงความเร็วในการเข้าถึงข้อมูลทั้งการอ่าน และการเขียนข้อมูล โดยจะนำเสนอสมการการคำนวณสมรรถนะทั้งกรณีฟังก์ชันการแทรก และฟังก์ชันการสอบถาม ที่สามารถเปลี่ยนแปลงค่าตัวแปรคุณสมบัติของดัชนีวายทรี และของแฟลชไดรฟ์ได้ นอกจากนี้จะนำเสนอแนวทางการอิมพลีเมนต์ (Implement) ดัชนีวายทรีบนแฟลชไดรฟ์อีกด้วย

ทั้งนี้แฟลชไดรฟ์ที่มีใช้งานอยู่ในปัจจุบันมีอยู่ 2 แบบด้วยกันคือ แฟลชไดรฟ์แบบนอร์ (NOR Flash Drive) และแฟลชไดรฟ์แบบแนนด์ (NAND Flash Drive) ในวิทยานิพนธ์ฉบับนี้เลือกใช้แฟลชไดรฟ์แบบแนนด์เนื่องจากมีคุณสมบัติที่ดีกว่านอร์แฟลชไดรฟ์หลายประการ [7] ตัวอย่างเช่น ความจุสูงกว่า การอ่าน การเขียนข้อมูลที่เร็วกว่า มีอายุการใช้งานที่ยาวกว่า และมีราคาต่ำกว่า เป็นต้น และคุณสมบัติที่แตกต่างอีกประการหนึ่งของแฟลชไดรฟ์แบบแนนด์คือ รองรับการเข้าถึงข้อมูลแบบการเข้าถึงโดยลำดับ (Sequential Access Method) ซึ่งมีผลทำให้การอ่านข้อมูลมีความเร็วสูงกว่าการเขียนข้อมูลหลายเท่าดังแสดงในตารางที่ 1.2

การคำนวณสมรรถนะฟังก์ชันการแทรก และฟังก์ชันการสอบถามของดัชนีวายทรีในวิทยานิพนธ์ฉบับนี้จะใช้แนวทางการคำนวณสมรรถนะของดัชนีแบบ Value-list ตามเอกสาร [5] โดยการคำนวณสมรรถนะแบ่งการคำนวณออกเป็น 2 กรณีตามฟังก์ชัน คือ

1) การคำนวณสมรรถนะของฟังก์ชันการแทรก

การวิเคราะห์สมรรถนะของฟังก์ชันการแทรกเริ่มจาก การกำหนดการประมาณจำนวน โหนดใบของดัชนีวายทรี จากนั้นให้พิจารณาเวลาของจำนวนโหนดที่จะต้องถูกอ่าน และเขียนเมื่อมีการแทรกเอ็นทรีดัชนี (Index Entry) เข้าสู่ดัชนีวายทรี และรวมไปถึง การพิจารณาความลึกของดัชนีวายทรี ทั้งนี้จะนำค่าของคุณสมบัติเฉพาะของแฟลชไดรฟ์มาร่วมพิจารณาด้วย โดยคุณสมบัติเหล่านั้นได้แก่ เวลาการอ่าน เวลาการเขียน เวลาการลบข้อมูล และเวลาการเข้าถึงข้อมูล

2) การคำนวณสมรรถนะของฟังก์ชันการสอบถาม

การวิเคราะห์สมรรถนะของฟังก์ชันการสอบถามมีความคล้ายกับฟังก์ชันการแทรก เพียงแต่ฟังก์ชันการสอบถามมีองค์ประกอบของการอ่านเท่านั้น เริ่มต้นด้วยการ ประเมินเวลาของการสอบถามแบบช่วง (Range Query) ที่ต้องการคำตอบเป็นชุดข้อมูล (String) ซึ่งชุดข้อมูลนี้เกิดจากการทราเวิร์ด (Traverse) แบบปรกติ และเข้าถึงโหนดต่อไปด้วยพอยเตอร์ (Pointer) จนกว่าจะถึงจุดสิ้นสุดของการสอบถามแบบช่วง และจะแบบเป็นซิงค์เกอร์พาท (Single Path) ส่วนคุณสมบัติเฉพาะของแฟลชไดรฟ์ที่เลือกนำมาใช้ คือ เวลาการอ่าน และเวลาการเข้าถึงข้อมูลเท่านั้น

ในขั้นตอนสุดท้ายของการวิเคราะห์สมรรถนะของดัชนีวายทรีบนแฟลชไดรฟ์นั้นจะได้แนวทางของสมการ ที่มีความสัมพันธ์กับองค์ประกอบต่างๆ ตามแนวทางการวิเคราะห์ที่กล่าวมาแล้วข้างต้นดังสมการที่ (1.1)

$$y = f(\text{ytree}) * f(\text{flash}) \quad (1.1)$$

เมื่อ

y คือ สมรรถนะของดัชนีวายทรีบนแฟลชไดรฟ์
 f(ytree) คือ ฟังก์ชันการเข้าถึงข้อมูลของดัชนีวายทรี ซึ่งอาจมีความแตกต่างกันระหว่างฟังก์ชันการแทรก และฟังก์ชันการสอบถาม
 f(flash) คือ ฟังก์ชันคุณสมบัติเฉพาะของแฟลชไดรฟ์ โดยคุณสมบัติที่เลือกใช้อาจแตกต่างกันระหว่างฟังก์ชันการแทรก และฟังก์ชันการสอบถาม

1.3 สรุปสาระสำคัญของเอกสารที่เกี่ยวข้อง

R. Elmasri และคณะ [1] ได้นำเสนอความรู้เกี่ยวกับความรู้พื้นฐาน การออกแบบ และการใช้งานระบบจัดการฐานข้อมูล

เศรษฐพงศ์ [2] ได้นำเสนอความรู้เกี่ยวกับ ความหมาย ประเภท ข้อดี และข้อเสียของคลังข้อมูล

L. Adam [3] ได้นำเสนอข้อมูลประเภทของแฟลชไดรฟ์ คุณสมบัติต่างๆ และสถิติราคาของแฟลชไดรฟ์ในอดีต

J. Christopher และคณะ [4] ได้นำเสนอดัชนีวายทรีซึ่งเป็นดัชนีโครงสร้างต้นไม้ซึ่งมุ่งเน้นการปรับปรุงสมรรถนะ สำหรับการแทรกข้อมูล ให้มีความรวดเร็วเพิ่มมากขึ้น และนำเสนอแนวทางการวิเคราะห์สมรรถนะดัชนีวายทรีของฟังก์ชันการแทรก และฟังก์ชันการสอบถาม เพื่อปรับปรุงสมรรถนะของดัชนีวายทรีให้มีความเหมาะสมกับขนาดของข้อมูล

M-Systems Flash Disk Pioneers, Ltd [5] ได้นำเสนอข้อมูลการเปรียบเทียบคุณสมบัติต่างๆ ของแฟลชไดรฟ์แบบนอร์ และแฟลชไดรฟ์แบบแนนด์

Mfatix Company Limited [6] ได้ให้ความหมาย คำจำกัดความ ลักษณะ และรูปแบบ ของ ปัญญาธุรกิจ

T. Sai และคณะ [7] ได้นำเสนอดัชนีโครงสร้างต้นไม้แบบใหม่ เพื่อแก้ไขปัญหาการ ปรับปรุงล่าช้าของดัชนีบีพลัสทรี เมื่อใช้งานบนแฟลชไดรฟ์ และนำเสนอสมการ การวัดสมรรถนะ ของการปรับปรุงข้อมูล

K. Sebastian [8] ได้นำเสนอดัชนีโครงสร้างอาร์ทรี ซึ่งเป็นดัชนีที่ใช้แก้ปัญหาของข้อมูล หลายมิติ โดยจะพิจารณาวัตถุเป็นรูปสี่เหลี่ยมผืนผ้า นอกจากนั้นยังกล่าวถึงคุณสมบัติของดัชนี อาร์ทรีอีกด้วย

L. Yinan และคณะ [9] ได้นำเสนอดัชนีโครงสร้างต้นไม้แบบเอฟดีทรี (FD-tree) ซึ่งเป็น ดัชนีต้นไม้แบบใหม่ที่ใช้งานบนแฟลชไดรฟ์ เพื่อแก้ปัญหาคาร์ดิแฟลชไดรฟ์เขียนข้อมูลล่าช้า

1.4 วัตถุประสงค์ของการศึกษา

1.4.1 เพื่อวิเคราะห์สมรรถนะของดัชนีวายทรี เมื่อใช้งานบนแฟลชไดรฟ์

1.4.2 เพื่อหาพารามิเตอร์ที่เหมาะสมที่สุดของดัชนีวายทรี เมื่อใช้งานบนแฟลชไดรฟ์

1.5 ประโยชน์ที่ได้รับจากการศึกษาเชิงทฤษฎี และ/หรือ เชิงประยุกต์

เนื่องจากองค์กร หรือผู้ที่เกี่ยวข้องกับการประมวลผลข้อมูลในปริมาณที่มากบนแฟลช ไดรฟ์ และต้องการผลลัพธ์ที่รวดเร็ว เพื่อให้ทันต่อการนำไปใช้งานสามารถนำผลลัพธ์ และวิธีการ รวมไปถึงการปรับปรุงสมรรถนะที่ได้นำเสนอ ไปประยุกต์ใช้งานเพื่อให้ได้การประมวลผลที่ รวดเร็ว และมีสมรรถนะสูงสุด

1.6 ขอบเขตการทำวิจัย

1.6.1 การประเมินสมรรถนะ หมายถึงความเร็วในการเข้าถึงข้อมูลด้วยหน่วยวัดเป็นวินาที

1.6.2 นำเสนอแนวทางการพัฒนาดัชนีวายทรี

1.7 วิธีการทำวิจัย

1.7.1 ศึกษาทฤษฎีเกี่ยวกับแฟลชไดรฟ์

1.7.2 ศึกษาทฤษฎีเกี่ยวกับดัชนีวายทรี

1.7.3 นำเสนอแนวทางการพัฒนาดัชนีวิทยุที่ตามคำแนะนำในเอกสารของดัชนีวิทยุที่

1.7.4 จำนวนสมรรถนะของฟังก์ชันการแทรก และฟังก์ชันการสอบถาม

1.7.5 สรุปผลการวิจัยการใช้ดัชนีวิทยุบนแพลตฟอร์ม พร้อมเสนอแนะการคำนวณ

พารามิเตอร์ ที่ให้สมรรถนะการใช้งานสูงขึ้น

1.7.6 จัดทำ และเสนอรายงานวิทยานิพนธ์

1.8 เครื่องมือในการพัฒนา

ในวิทยานิพนธ์นี้การพัฒนาขั้นตอนวิธีทั้งหมดจะใช้เครื่องมือในการพัฒนาดังนี้

1.8.1 ฮาร์ดแวร์

- Lenovo ThinkPad X200si (Core 2 Duo SU7300 1.3 GHz)
- DDR3 4 Gigabytes Main memory
- Samsung MMDOE56G5MXP-0VB00 2.5" 256GB SATA Solid State Drive

1.8.2 ซอฟต์แวร์

- Ubuntu 9.10
- JDK 1.6
- NetBean 6.8