

記錄編號	14575
狀態	G0401346184
助教查核	
索書號	
學校名稱	輔仁大學
系所名稱	資訊管理學系
舊系所名稱	
學號	401346184
研究生(中)	林子揚
研究生(英)	
論文名稱(中)	基於Hadoop Hive之Big Data效率分析研究
論文名稱(英)	Efficiency Analysis of Big Data Based on Hadoop Hive
其他題名	
指導教授(中)	蔡明志
指導教授(英)	
校內全文開放日期	
校外全文開放日期	
全文不開放理由	
電子全文送交國圖	
國圖全文開放日期	
檔案說明	
電子全文	
學位類別	碩士
畢業學年度	103
出版年	103
語文別	中文
關鍵字(中)	雲端運算、海量資料、Hadoop、Hive
關鍵字(英)	Cloud computing、Big Data、Hadoop、Hive
摘要(中)	<p>資訊科技快速發展，大量且多元的資料無時無刻在產生，但以目前企業IT技術，面對TB等級的資料量，根本難以負荷，無法符合即時、準確的需求。因此本論文將實際架設一套Hadoop分散式運算平台，處理結構化的海量資料集，解決傳統資料庫處理海量資料時，速度及效率上所面臨的瓶頸，並歸納出在固定的資料量及限定的時間內，找出最符合經濟效益（節點個數、處理器核心數、記憶體量）的節點成本。本研究前測階段會使用5台一般電腦實體機器搭配小量資料集進行測試，瞭解不同資料量的效率；而實驗階段為了更精準的控制資源使用量，因此我們會使用5台伺服器等級的電腦並搭配虛擬化技術來調配資源，並使用共有四千多萬筆資料的海量資料集進行實驗，並會針對節點個數、資料量、處理器核心數、記憶體使用量進行研究，使本研究可以歸納出特定程度的海量資料分析效率，並期許對於後續研究及企業有實質的貢獻。</p>
摘要(英)	<p>The rapid development of information technology (IT), a large and diverse data generating all the time, but the current IT in enterprise, it's difficult to process large volumes of data in TB level. Therefore, this paper will practically set up a Hadoop distributed computing platform for processing structured data, and solving the efficiency bottleneck of traditional datasets w</p>

	<p>hen dealing with large volumes of data. In conclusion, we will conclude that the most cost-effective interval (number of nodes, processor cores, memory usage) within a fixed amount of data and the limited time. In this paper, pretest phase we use 5 PC (personal computer) and a small set of datasets, to understand the different amount of data efficiency. Implementation phase we use five server-class computer and a dataset that more than 40 million records, and focused on number of nodes, quantity of data, number of CPU cores, memory usage, so this study can be summarized the large volumes of data analysis in specific level of efficiency, Expectations have substantial contribution for the future research and corporate.</p>
論文目次	
參考文獻	
論文頁數	
附註	
全文點閱次數	
資料建置時間	
轉檔日期	
全文檔存取記錄	
異動記錄	