

專題演講

演講者：柯英毅研究員、許智程研究員(群暉科技-軟體開發部)

演講題目：High Scale Storage (高拓展性儲存技術)

演講摘要：

Topic 1	Distributed filesystem	<p>隨著全球資料量的大幅成長，現存的單機儲存技術也面臨越來越嚴苛的挑戰：在 PB 等級的資料量下，單機儲存與資料保護的成本皆會大幅度的提升。如何使用儲存伺服器進行低成本的水平擴展，來增進效能、儲存容量與資料防護性，已是當今巨量資料潮流下所必需面對的課題。</p> <p>本議程我們將從開放原始碼的分散式檔案系統---Gluster Filesystem 做為起點，探討分散式檔案系統的儲存架構、容錯能力及與現有應用程式相容等議題；並同時分享此架構在商業產品中的實際應用和效能優化。</p>
Topic 2	Deduplication	<p>隨著資料量的爆炸性成長，如何有效率地使用儲存空間更顯得重要。然而，這龐大資料內容中真的是完全不重複的嗎？以資料備份應用為例，同一備份任務每次備份的資料內容中經常有很大的重複性。</p> <p>而作為儲存伺服器，不同使用者也有可能各自儲存了相似內容的檔案。如果可以有效率地找出有重複的資料區段，便能大幅減少儲存空間的使用進而降低儲存成本。</p> <p>資料重複刪除技術(Data De-duplication)便是為此而生，其將含括：如何判定資料內容是否重複、如何從茫茫資料海中找到相同的資料內容、如何儲存與存取重複/不重複的資料區塊。</p>

講者簡歷：

柯英毅 研究員：

國立台灣大學 電機工程學系研究所(計算機科學組)

許智程 研究員：

國立台灣大學 資訊工程博士