

ฉุริเคช สุกสิ : การเพิ่มผลผลิตภพของการทดสอบซอฟต์แวร์แบบกระจายโดยใช้จุด
ตรวจสอบ (A PRODUCTIVITY IMPROVEMENT OF DISTRIBUTED SOFTWARE
TESTING USING CHECKPOINTS) อาจารย์ที่ปรึกษา : ผู้ช่วยศาสตราจารย์ ดร.ชาญวิทย์
แก้วกลี, 116 หน้า.

จากความก้าวหน้าของเทคโนโลยีการเก็บข้อมูลและอัตราการสร้างข้อมูลที่เกิดขึ้นได้อย่าง
รวดเร็วในปัจจุบันทำให้เกิดยุคของข้อมูลขนาดใหญ่ แต่เครื่องมือที่ออกแบบมาก่อนหน้านั้นพบว่า
ไม่เหมาะสำหรับการประมวลผลข้อมูลขนาดใหญ่ที่เก็บอยู่ในปัจจุบัน ดังนั้นเพื่อที่จะประมวลผล
ข้อมูลเหล่านั้นจึงมีการพัฒนาเครื่องมือที่ใช้ประมวลผลงานในลักษณะงานที่ต้องการประมวลผล
ข้อมูลที่สามารถขยายได้ เครื่องมือหนึ่งที่เป็นที่นิยมในการใช้งานคือ Apache Spark ซึ่งเน้นการ
ประมวลผลบนหน่วยความจำ ทำให้ทำงานได้อย่างรวดเร็ว แต่อย่างไรก็ตามพบว่ายังไม่มีการ
การควบคุมคุณภาพซอฟต์แวร์ที่สามารถทดสอบกับกรณีทดสอบบนข้อมูลจริงที่มีขนาดใหญ่ได้
นักพัฒนาจึงเลือกที่จะใช้ข้อมูลตัวแทนชุดข้อมูลมาเพื่อทดสอบ ทำให้นักพัฒนาบางรายเลือกที่จะ
ไม่ทดสอบซอฟต์แวร์เลยเพื่อลดความยุ่งยากและต้องใช้เวลาทดสอบนาน หรือกระทั่งไม่สามารถ
ทดสอบได้บนเครื่องคอมพิวเตอร์เครื่องเดียว

งานวิจัยในเอกสารนี้จึงได้นำเสนอ Distributed Test Checkpointing (DTC) สำหรับกรอบ
งาน Apache Spark ซึ่งจะช่วยให้สามารถทดสอบซอฟต์แวร์เป็นกรณีทดสอบแบบเชิงหน่วย บน
ข้อมูลขนาดใหญ่ที่ยังคงความถูกต้อง และใช้เวลาทดสอบที่ผู้ใช้สามารถยอมรับได้ จากการทดลอง
พบว่ากรอบงาน DTC จะใช้เวลาทำงานใกล้เคียงหรือสูงกว่าเมื่อเทียบกับกลไกสร้างจุดตรวจสอบ
ดั้งเดิมของ Spark ในกรณีทดสอบแรก และในกรณีทดสอบถัดมากลับสามารถลดระยะเวลาได้อย่าง
มาก กรณีการตั้งค่ากรอบงาน DTC ที่ดีที่สุดพบว่าสามารถลดเวลาการประมวลผลได้ถึงร้อยละ 450
ถึงร้อยละ 500 และยังสามารถลดการใช้พื้นที่เก็บจุดตรวจสอบได้มากกว่า 19.7 เท่า

สาขาวิชา วิศวกรรมคอมพิวเตอร์
ปีการศึกษา 2560

ลายมือชื่อนักศึกษา 
ลายมือชื่ออาจารย์ที่ปรึกษา 

BHURIDECH SUDSEE : A PRODUCTIVITY IMPROVEMENT OF DIS-
TRIBUTED SOFTWARE TESTING USING CHECKPOINTS. THESIS
ADVISOR : ASST. PROF. CHANWIT KAEWKASI, Ph.D., 116 PP.

DISTRIBUTED CHECKPOINTING/APACHE SPARK/BIG DATA TESTING/
SOFTWARE TESTING;

The advancement of storage technologies and the fast-growing number of generated data have made the world moved into the Big Data era. In this past, we had many data mining tools but they are inadequate to process Data-Intensive Scalable Computing workloads. The Apache Spark framework is a popular tool designed for Big Data processing. It leverages in-memory processing techniques that make Spark up to 100 times faster than Hadoop. Testing this kind of Big Data program is time consuming. Unfortunately, developers lack a proper testing framework, which cloud help assure quality of their data-intensive processing programs, while saving development time.

We propose Distributed Test Checkpointing (DTC) for Apache Spark. DTC applies unit testing to the Big Data software development life cycle and reduce time spent for each testing loop with checkpoint. From the experimental results, we found that in the subsequence rounds of unit testing, DTC dramatically speed the testing time up to 450 – 500% . In case of storage, DTC can cut unnecessary data off and make the storage 19.7 times saver than the original checkpoint of Spark.

School of Computer Engineering

Academic Year 2017

Student's Signature

Advisor's Signature

