



Dropbox

จัดทำโดย

นาย พันวัชร เฉลิมวงศ์ 5709611957

รายงานนี้เป็นส่วนหนึ่งของรายวิชา

Operating System 2 (CS447)

มหาวิทยาลัยธรรมศาสตร์

ภาคเรียนที่ 1 ปีการศึกษา 2560

สารบัญ

หัวข้อ	หน้า
การให้บริการของ Dropbox	1
ปัญหาด้าน Scalability ของ Dropbox	1
การแก้ไขปัญหาด้าน Scalability ของ Dropbox	2
ผลลัพธ์ที่ได้เมื่อ Dropbox ใช้วิธีการดังกล่าว	4
สิ่งที่สัมพันธ์กับเนื้อหาที่ได้ศึกษามา	5
บรรณานุกรม	7

1. การให้บริการของ Dropbox

หลายปีก่อนคนส่วนใหญ่เรียก Dropbox ว่า “Magic Pocket”(กระเป๋าที่เก็บรวบรวมทุกอย่างที่เรามีและสามารถพกไปทุกที่ได้) เพราะว่า Dropbox ถูกออกแบบมาให้สามารถเก็บไฟล์ทั้งหมดของคุณไว้ในที่เดียว และ Dropbox ได้พัฒนาขึ้นจากจุดเริ่มต้นง่ายๆจนกลายเป็นหนึ่งในแพลตฟอร์มการทำงานร่วมกันที่มีประสิทธิภาพและแพร่หลายมากที่สุดในโลก ซึ่งจากที่กล่าวมานั้นส่งผลให้ Dropbox มีข้อมูลที่ต้องจัดเก็บมากขึ้นเรื่อยๆ ส่งผลให้เกิดปัญหาต่างๆตามมา

2. ปัญหาด้าน Scalability ของ Dropbox

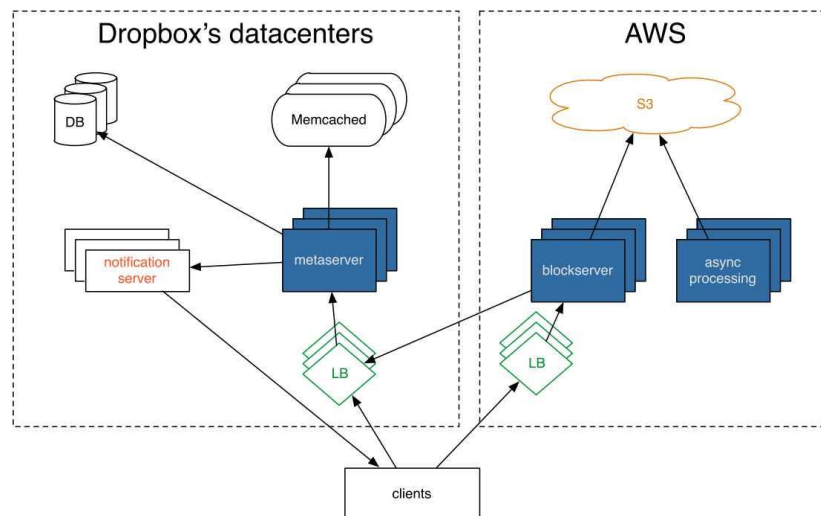
นับตั้งแต่ปี 2008 Dropbox ก็เติบโตขึ้นอย่างมหาศาล ในปี 2011 พวกเขาต้องบันทึกไฟล์ 1 ล้านไฟล์ในทุก 15 นาที และในปี 2012 พวกเขามีข้อมูลของผู้ใช้อยู่ที่ประมาณ 40 petabytes จนกระทั่งปี 2016 มีผู้สมัครใช้งานมากกว่า 500 ล้านราย และมีข้อมูลของผู้ใช้มากถึง 500 petabytes (เลข 5 และตามด้วย 0 ถึง 17 ตัว) ซึ่งนั่นมากกว่าตัวอักษรของหนังสือทั้งหมดในหอสมุดแห่งชาติถึง 14,000 เท่า จะเห็นได้ว่าในปี 2016 พวกเขาเติบโตขึ้นมากกว่าปี 2012 ถึง 12 เท่า

ด้วยความที่ Dropbox เติบโตขึ้นอย่างรวดเร็วในระยะเวลาอันสั้นทั้งในด้านขนาดของข้อมูลและในด้านจำนวนผู้ใช้งาน ทำให้พวกเขาประสบปัญหาในการขยายพื้นที่จัดเก็บข้อมูลให้สามารถรองรับอัตราการเพิ่มขึ้นอย่างรวดเร็วของข้อมูล ซึ่งเมื่อมีข้อมูลจำนวนมากๆทำให้มีอีกปัญหาที่ตามมาคือความเร็วในการค้นหาข้อมูลของผู้ใช้ลดลงและส่งผลให้ Availability ลดลง และจากการที่มีผู้ใช้จำนวนมากขึ้นเรื่อยๆส่งผลให้มีการไหลของข้อมูลมากขึ้นอย่างมากส่งผลให้ดาต้าเซ็นเตอร์ที่มีอยู่ไม่สามารถปรับให้รองรับการไหลที่เพิ่มขึ้นได้ทันทำให้เกิดเหตุการณ์ที่ผู้ใช้ไม่สามารถใช้งานระบบได้

3. การแก้ไขปัญหาด้าน Scalability ของ Dropbox

ในปี 2012 Dropbox ได้ทำ hybrid infrastructure เชื่อมต่อ ตัวดาต้าเซ็นเตอร์ของพวกเขาเข้ากับ AWS

2012



สืบทอดจาก (<https://www.infoq.com/presentations/dropbox-infrastructure>)

- ในส่วนของ AWS นั้นจะใช้ในส่วนของ Amazon S3 และมีการทำ blockserver เพื่อกรองให้การไหลของข้อมูลไปยัง Amazon S3 นั้นไม่มากเกินไป โดย blockserver จะทำการตรวจสอบว่าการเรียกดูนั้นเป็นรูปแบบไหน
- ในส่วนของตัวดาต้าเซ็นเตอร์ของ Dropbox มีการทำ Metaserver เป็นตัวจัดการเรื่องต่างๆ
- ในตัว Metaserver มีการนำ Metadata มาใช้เพื่อให้การค้นหาเร็วขึ้น

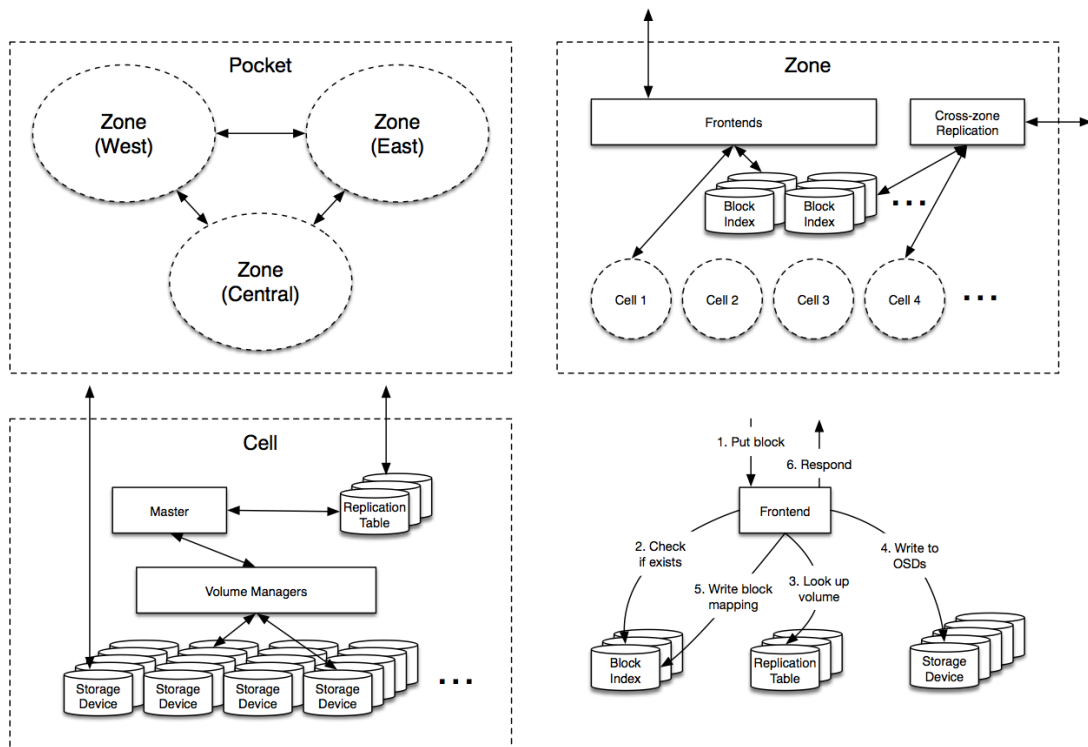
และเนื่องจากการที่ AWS เป็น service ที่มี Availability Reliability และ Scalability สูงพอที่จะมั่นใจได้ว่าสามารถรองรับการขยายของ Dropbox ในอนาคตได้โดยไม่เกิดปัญหาใดๆ

การ scaling ในส่วน database

- ใช้การขยายในรูปแบบ Horizontal scaling
- มีการทำ Sharding
- มีการใช้ Memcached

- มีการใช้ Replication

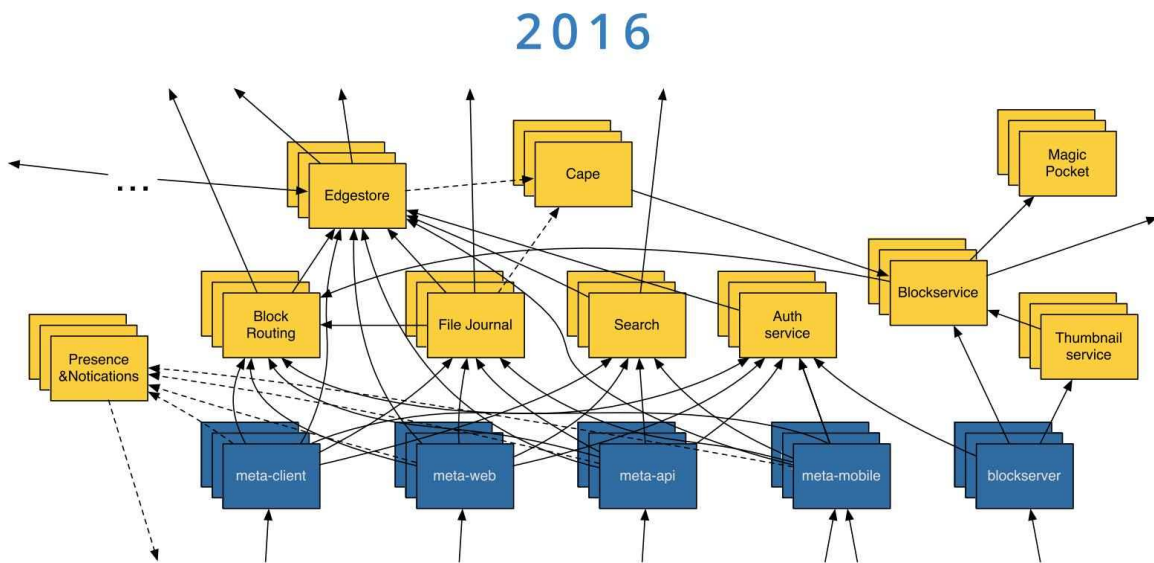
ในปี 2013 Dropbox ได้มีการริเริ่มโปรเจกขนาดใหญ่ที่ชื่อว่า Magic Pocket เพื่อจะมาทำหน้าที่แทนในส่วนของ Amazon S3



สืบทอดจาก (<https://www.infoq.com/presentations/dropbox-infrastructure>)

- ในส่วนของดาต้าเซ็นเตอร์นั้นมีการแบ่งออกเป็น 3 พื้นที่คือ center west east โดยแต่ละดาต้าเซ็นเตอร์ก็จะส่งผ่านข้อมูลหากัน

ในปี 2016 ได้มีการแยกส่วนการทำงานของ meta server ออกเป็นส่วนๆและได้นำ magic pocket มาใช้แทนใน ส่วนของ Amazon S3 แต่บางส่วนก็ยังส่งไปยัง Amazon S3 อยู่



สืบทัดจาก (<https://www.infoq.com/presentations/dropbox-infrastructure>)

4. ผลลัพธ์ที่ได้เมื่อ Dropbox ใช้วิธีดังกล่าว

1. แม้ว่าจะมีผู้ใช้งานในระบบเยอะ Dropbox ก็ยังสามารถทำงานได้รวดเร็ว
2. กรณีมีเซิร์ฟเวอร์ใดเซิร์ฟเวอร์หนึ่งดับไประบบก็จะยังสามารถทำงานได้เหมือนเดิม
3. สามารถรองรับอัตราการขยายตัวของข้อมูลและจำนวนผู้ใช้ได้
4. ผู้ใช้ไม่ต้องกังวลว่าข้อมูลจะหายไปทั้งหมด
5. การไหลของข้อมูลไปยังเซิร์ฟเวอร์ของ Dropbox ลดลงอย่างมาก

5. สิ่งที่สัมพันธ์กับเนื้อหาที่ได้ศึกษามา

- Availability

Dropbox นั้นมีความต้องการที่จะแก้ปัญหาต่างๆที่เกิดขึ้นเพื่อไม่ให้เกิดเหตุการณ์ที่ระบบนั้นไม่สามารถใช้งานได้ มีความต้องการต้องการให้ระบบใช้งานได้มากกว่า 99.99% ของเวลาทั้งหมด นอกจากนี้ Dropbox ยังคำนึงถึงความเร็วในการทำงานต่างๆในระบบเพื่อให้สามารถทำงานตามความต้องการของผู้ใช้ได้ดีที่สุด

- Scalability

Dropbox ได้มีการปรับขีดความสามารถให้สามารถรองรับจำนวนผู้ใช้งานและขนาดของข้อมูลเพื่อไม่ให้เกิดปัญหาเกี่ยวกับในเรื่องของ Availability โดยเมื่อมีความต้องการเพิ่มมากขึ้น Dropbox ก็จะขยายความสามารถในการรองรับให้สูงขึ้น

- Horizontal Scalability

เวลาที่ Dropbox ขยายความสามารถในการรองรับ Dropbox จะใช้วิธีการเพิ่มจำนวนเซิร์ฟเวอร์ขึ้น

- Content Delivery Network

ไม่ว่าจะเป็นตอนที่ Dropbox ใช้ Amazon S3 หรือตอนที่ใช้โปรเจกต์ Magic pocket ของตนเองนั้น การเรียกข้อมูลของผู้ใช้จะตรงไปยังเซิร์ฟเวอร์ที่อยู่ใกล้ผู้ใช้นั้นๆมากที่สุด

- Caching

Dropbox ได้มีการใช้ Memcached ในส่วนของ database โดย Memcached ก็จะทำให้การเก็บค่าคีย์ต่างๆไว้เพื่อประหยัดเวลาและทรัพยากรของระบบในการค้นหาข้อมูลต่างๆ

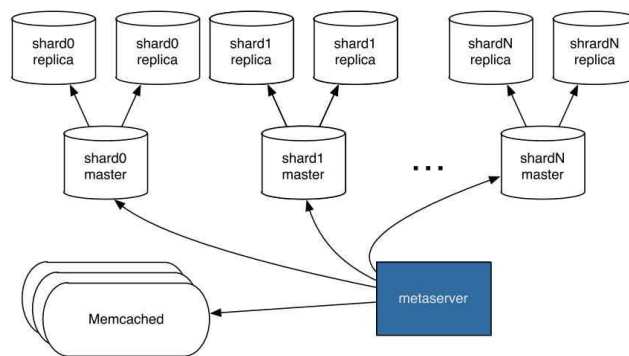
- Replication

ก็จะเห็นได้ตั้งแต่ในส่วนโครงสร้างของการ Scaling Database ของ Dropbox ว่ามีการนำ Replication มาใช้ โดยหน้าที่ของ Replication ก็คือการคัดลอกข้อมูลจาก Database หนึ่ง ไปยัง Database อื่นๆ(จาก Master ไปยัง บรรดา Slave) ซึ่งในที่นี่เพื่อเป็นการป้องกันการสูญหายของข้อมูลของผู้ใช้

- Sharding

Dropbox มีการแบ่งข้อมูลออกเป็นส่วนย่อยๆ จำนวนนับไม่ถ้วนโดยแต่ละส่วนไม่ได้ขึ้นต่อกัน เพื่อที่เวลา Database ตัวใดผิดพลาดก็จะแทบไม่ส่งผลกระทบต่อการใช้งานเลย

SCALING DATABASES



สืบทอดจาก (<https://www.infoq.com/presentations/dropbox-infrastructure>)

บรรณานุกรม

- <http://highscalability.com/blog/2011/3/14/6-lessons-from-dropbox-one-million-files-saved-every-15-minu.html>
- <https://blogs.dropbox.com/tech/2016/03/magic-pocket-infrastructure/>
- <https://www.infoq.com/presentations/dropbox-infrastructure>
- <https://blogs.dropbox.com/tech/2016/05/inside-the-magic-pocket/>