



รายงาน

ESPN Scalability

จัดทำโดย

นายศุภชัย ตริประทีป 5809610669

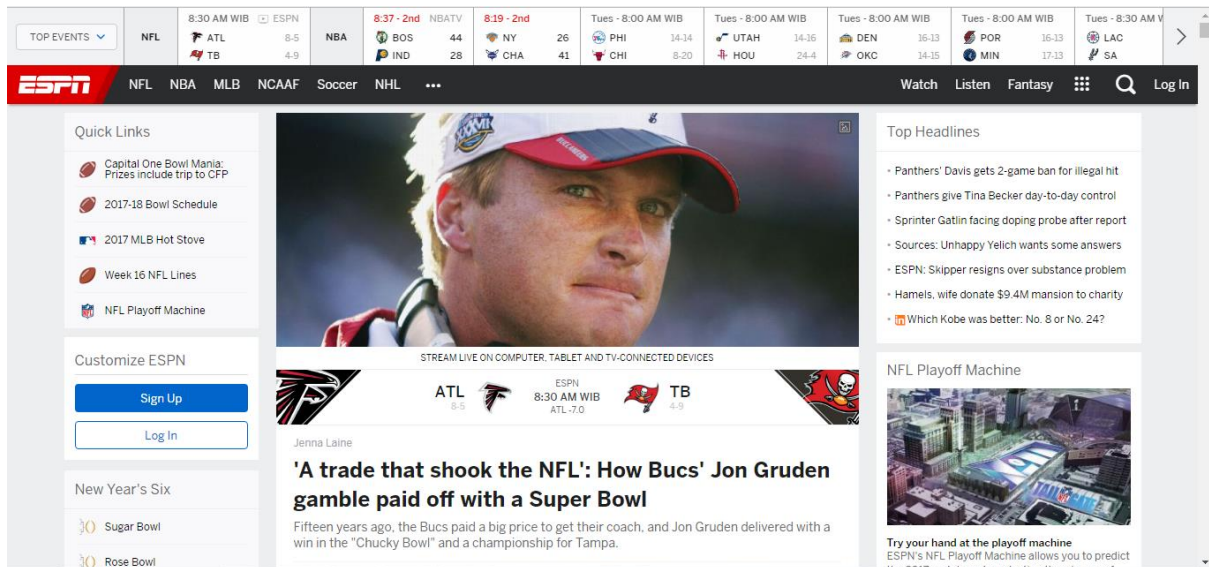
เสนอ

ดร. ประภาพร รัตนธารง

รายงานนี้เป็นส่วนหนึ่งของวิชา คพ. 447 ระบบปฏิบัติการ 2
ภาคเรียนที่ 1 ปีการศึกษา 2560

บริษัทดังกล่าวให้บริการอะไร

ESPN คือ สถานีเคเบิลทีวีแห่งหนึ่งของอเมริกา ให้บริการในการผลิตและแพร่ภาพรายการเกี่ยวกับกีฬาทุกชนิดตลอด 24 ชั่วโมง นอกจากนี้ ESPN ยังมีเว็บไซต์และแอปพลิเคชันที่ให้บริการเกี่ยวกับกีฬาอีกหลากหลายรูปแบบ เช่น ถ่ายทอดสด ตารางกิจกรรมของกีฬาประเภทต่างๆ วิเคราะห์ข้อมูลข่าวสารเกี่ยวกับกีฬา เป็นต้น



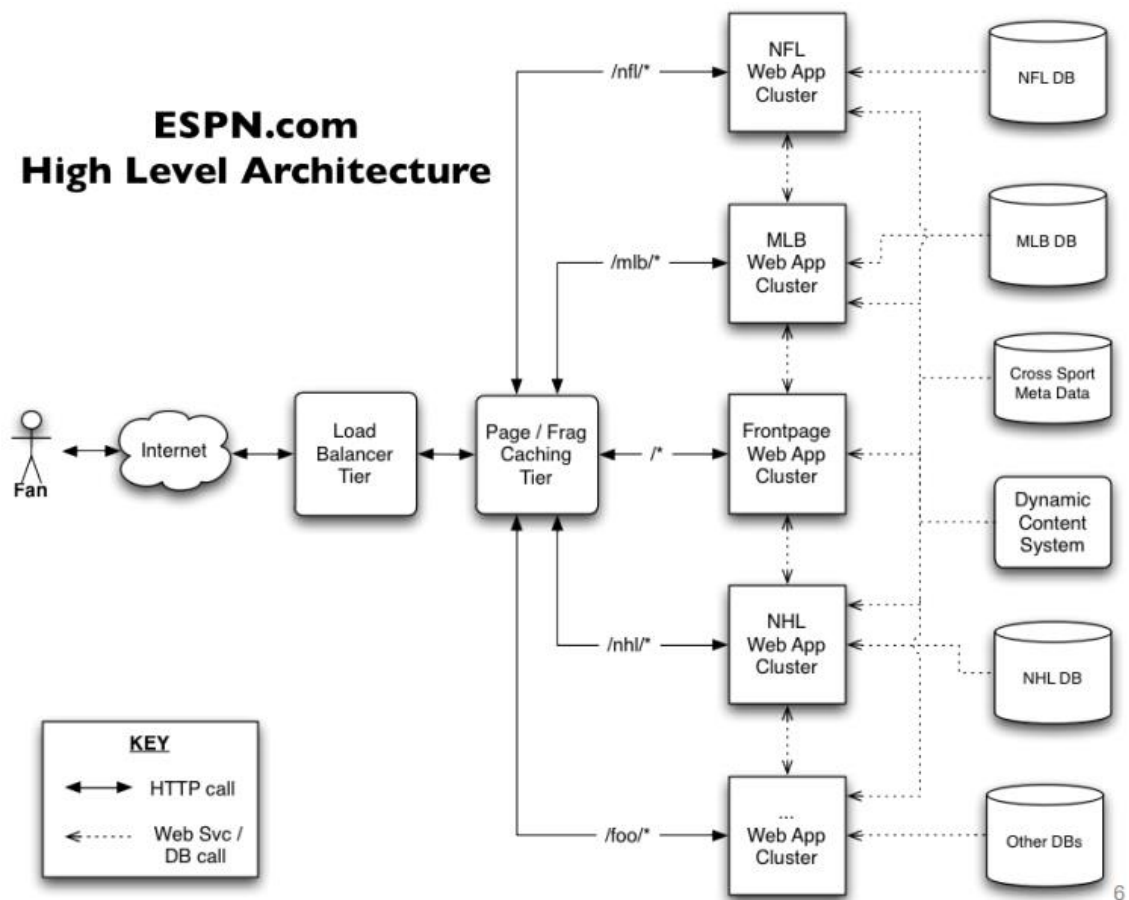
1.ตัวอย่างเว็บไซต์ ESPN.com

ประเด็นปัญหาเกี่ยวกับ Scalability ที่ประสบในการให้บริการ

ESPN ได้มีความคิดที่จะพัฒนาเว็บไซต์โดยมีเป้าหมายที่จะให้บริการผู้ให้บริการทุกที่ทุกเวลาบนอุปกรณ์รูปแบบใดก็ได้ พร้อมข้อมูลที่ถูกต้องและทันเวลาโดยสามารถเข้าถึงเนื้อหาที่ลึกซึ้งได้ คะแนนสถิติ และเนื้อหาที่ลึกจะต้องถูกต้องและรวดเร็ว ไม่มีการหยุดทำงาน

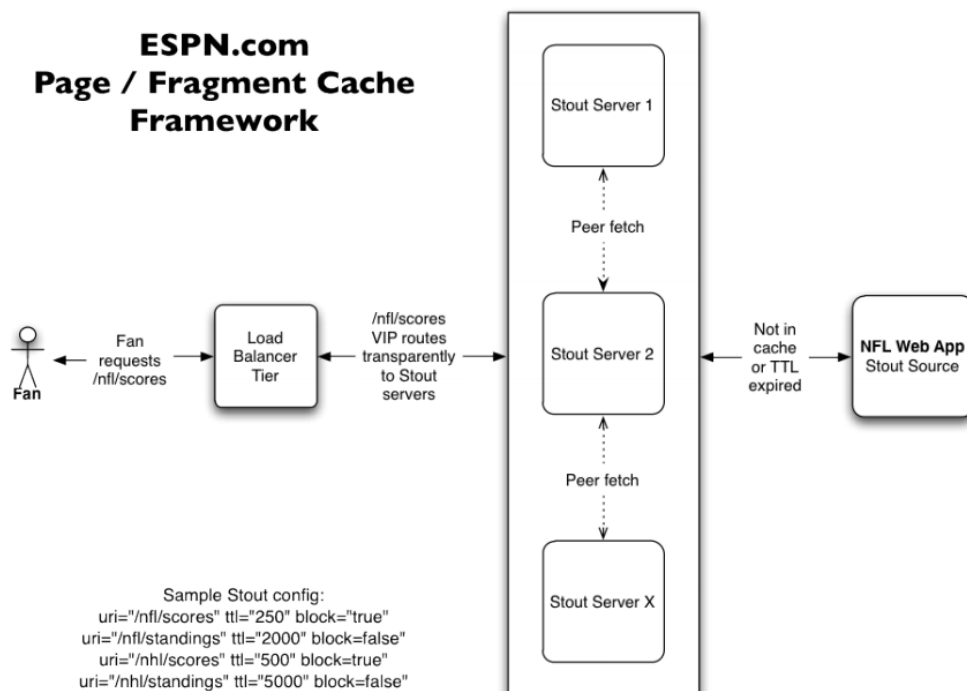
เทคนิควิธี หรือสถาปัตยกรรมที่ใช้ในการแก้ปัญหา

1. High Level Architecture คือรูปแบบสถาปัตยกรรมที่มีตัวจัดการกลาง ทำหน้าที่ในการรับข้อมูลจากแอปพลิเคชันผู้ใช้เข้ามา แล้วทำการส่งข้อมูลที่ได้อไปยังแอปพลิเคชันจำลองเพื่อประมวลผล โดยแต่ละแอปพลิเคชัน จะมีdatabaseแยกเป็นของตัวเอง



2. High Level Architecture

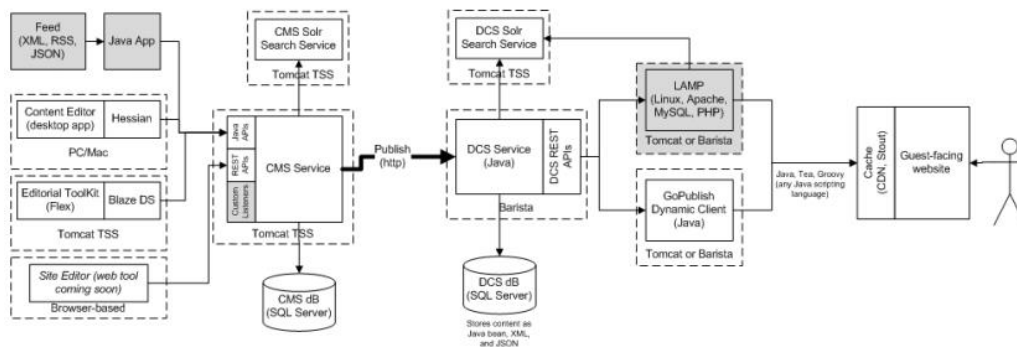
2. Page Caching Framework คือการใช้ Framework ในการเก็บข้อมูลเว็บไซต์ที่ได้มีการใช้ข้อมูลไปได้ไม่นาน โดยจะมี TTL เพื่อตรวจสอบว่าเก็บข้อมูลไว้นานแค่ไหนแล้ว เมื่อ TTL ถึงจุดที่กำหนดก็จะทำการไปดึงข้อมูลจากเซิร์ฟเวอร์



3. Page Caching Framework

3. GoPublish system เป็นระบบที่มีไว้ใช้จัดการกับเนื้อหาบนหน้าเว็บ โดยมีส่วนประกอบสำคัญคือ Content Management Service (CMS) ที่ใช้ในการจัดการเนื้อหาของหน้าเว็บ และ Dynamic Content Service (DCS) ที่จะส่งข้อมูลหน้าเว็บไปได้ตรงกับประเภทของผู้ใช้ โดยแบ่งจากชนิดของภาษาโปรแกรมของเครื่องผู้ใช้

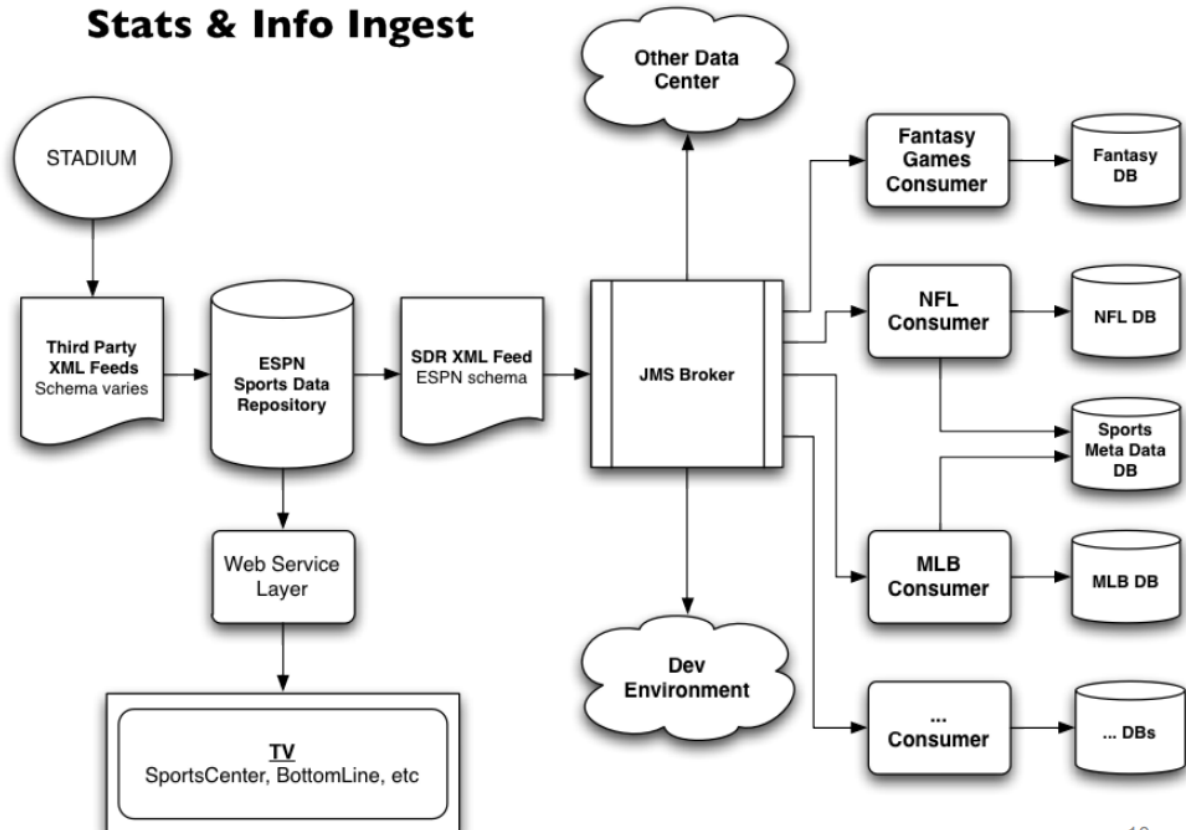
ESPN.com / Disney Dynamic Content System



4. GoPublish system

4. Data Ingest คือ การนำข้อมูลเข้ามาจากแหล่งอื่น ESPN จะมีการรับข้อมูลจากแหล่งข้อมูลที่เราเชื่อถือได้โดยที่สามารถนำข้อมูลมาแสดงผลแบบ real time ได้ หรือ สามารถส่งผ่านข้อมูลแบบเป็นชุดได้

ESPN.com Stats & Info Ingest



5. Data Ingest

ผลที่ได้เมื่อใช้เทคนิควิธีดังกล่าว

1. High Level Architecture เป็นการแบ่งให้เซิร์ฟเวอร์ทำงานในserviceที่แตกต่างกัน ทำให้ระบบทำงานได้เร็วขึ้น และเมื่อเกิดข้อผิดพลาดก็จะไม่ไปกระทบต่อส่วนอื่นๆ

2. Page Caching Framework ทำให้หน้าเว็บสามารถแสดงผลได้เร็วยิ่งขึ้น และช่วยลดภาระของเซิร์ฟเวอร์หลักลงได้เป็นอย่างมาก

3. GoPublish ทำให้หน้าเว็บมีการอัปเดตอยู่ตลอดเวลา เพราะมีการเปลี่ยนหน้าเว็บให้เป็นแบบ Dynamic

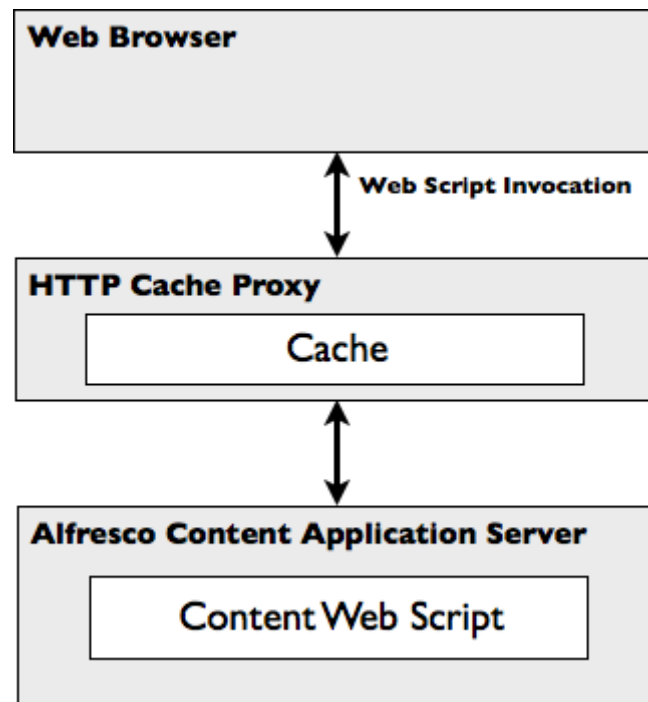
4. Data Ingest ช่วยให้ข้อมูลของ ESPN มีความน่าเชื่อถือมากยิ่งขึ้น เพราะมี3rd party ช่วยตรวจสอบให้

อภิปรายความสัมพันธ์กับเนื้อหาที่เรียนในชม.บรรยาย

1. มีการใช้ horizontal scaling ใน High Level Architecture โดยการเพิ่มเซิร์ฟเวอร์ขึ้นไปตามจำนวนประเภทของกีฬา เมื่อมีผู้เข้ามา load balancer ก็จะทำการส่งข้อมูลไปให้แก่เซิร์ฟเวอร์ของกีฬานั้นๆ ที่ผู้ใช้ต้องการ



2. Caching ใน Page Caching Framework และใน Cache จะมี TTL ในการบอกว่าเมื่อไรที่ Cache จะต้องไปขอข้อมูลจากเซิร์ฟเวอร์ใหม่อีกครั้ง



3. มีการทำหน้าเว็บแบบ Dynamic ใน GoPublish ส่วน DCS ที่จะส่งหน้าเว็บให้ผู้ใช้แต่ละคนไม่เหมือนกัน โดยจะเปลี่ยนไปตามชนิดของภาษาโปรแกรมของเครื่องผู้ใช้

4. ข้อมูลทั้งหมดมี Reliability สูง เพราะมี 3rd party ตรวจสอบให้



แหล่งอ้างอิงที่น่าเชื่อถือ

- <http://highscalability.com/blog/2013/11/4/espns-architecture-at-scale-operating-at-100000-duh-nuh-nuhs.html>
- <https://www.infoq.com/presentations/Architecture-Scale-ESPN>
- https://qconsf.com/sf2011/dl/qcon-sanfran2011/slides/SeanComerford_MaxProtectScalabilityAndCachingAtESPNCom.pdf
- <https://stackoverflow.com/questions/3887460/high-level-architecture-hla-versus-distributed-interactive-simulation-dis>
- <http://www.mktbrief.com/2015/02/personalization.html>
- <http://whatis.techtarget.com/definition/data-ingestion>