

ผู้ให้บริการ Cloud ที่มี SLA ความพร้อมใช้งาน 99.999% (Five Nines)

SLA 99.999% คืออะไร?

SLA ระดับ **99.999%** หรือที่เรียกกันว่า “**Five Nines**” หมายถึงระบบหรือบริการที่มีเวลาพร้อมใช้งานเกือบสมบูรณ์แบบ โดย **downtime** (ช่วงเวลาที่ระบบหยุดให้บริการ) จะไม่เกินประมาณ **5.26 นาทีต่อปี** (หรือราว 0.44 นาทีต่อเดือน) ¹ ซึ่งถือว่าเป็นความพร้อมใช้งานในระดับสูงมาก มักต้องการสำหรับระบบที่มีความสำคัญวิกฤต (mission-critical) เช่น ระบบการเงิน ระบบฉุกเฉิน หรือ แพลตฟอร์มที่ธุรกิจไม่สามารถหยุดชะงักได้เลย

Availability	Downtime / Year	Downtime / Month	Downtime / Week	Downtime / Day
99.999%	5.256 Minutes	0.438 Minutes	0.101 Minutes	0.014 Minutes
99.995%	26.28 Minutes	2.19 Minutes	0.505 Minutes	0.072 Minutes
99.990%	52.56 Minutes	4.38 Minutes	1.011 Minutes	0.144 Minutes
99.950%	4.38 Hours	21.9 Minutes	5.054 Minutes	0.72 Minutes
99.900%	8.76 Hours	43.8 Minutes	10.108 Minutes	1.44 Minutes
99.500%	43.8 Hours	3.65 Hours	50.538 Minutes	7.2 Minutes
99.250%	65.7 Hours	5.475 Hours	75.808 Minutes	10.8 Minutes
99.000%	87.6 Hours	7.3 Hours	101.077 Minutes	14.4 Minutes

ตารางเปรียบเทียบเปอร์เซ็นต์ความพร้อมใช้งาน (*Availability*) กับเวลาที่ระบบหยุดทำงานสูงสุดที่ยอมรับได้ต่อปี/เดือน/สัปดาห์/วัน โดย SLA 99.999% หมายถึงระบบหยุดทำงานได้เพียงประมาณ 5 นาทีต่อปี (ซึ่งคิดเป็นประมาณ 0.438 นาทีต่อเดือน) ¹

ด้านล่างนี้คือรายชื่อผู้ให้บริการ Cloud ชั้นนำระดับโลกที่ประกาศหรือรับประกัน SLA ความพร้อมใช้งาน **99.999%** (หรือใกล้เคียง) สำหรับบริการบางประเภทของคุณ พร้อมคำอธิบายว่าแต่ละรายใช้สถาปัตยกรรมหรือเทคนิคใดเพื่อให้บรรลุ SLA ระดับนี้ รวมถึงการระบุประเภทบริการ (SaaS, PaaS, IaaS) และข้อมูลเกี่ยวกับต้นทุน/การลงทุนที่เกี่ยวข้องในการสร้างระบบโครงสร้างพื้นฐานให้มีความน่าเชื่อถือสูงระดับ Five Nines

Amazon Web Services (AWS)

AWS มีบริการบางส่วนที่ให้ **SLA 99.999%** หรือมากกว่า ซึ่งส่วนใหญ่เป็นบริการโครงสร้างพื้นฐาน (IaaS) และแพลตฟอร์ม (PaaS) ที่ได้รับการออกแบบให้มีความทนทานสูงมาก:

- **Amazon DynamoDB Global Tables** – บริการฐานข้อมูล NoSQL แบบกระจายหลายภูมิภาค (multi-Region) ของ AWS ที่ให้ SLA **99.999% uptime** ² เมื่อใช้ฟีเจอร์ Global Tables (ตารางแบบหลายภูมิภาค) ซึ่งถือเป็นบริการแบบ PaaS (Database as a Service). สาเหตุที่ DynamoDB สามารถให้ความพร้อมใช้งานระดับนี้ได้เพราะ **สถาปัตยกรรม multi-active, multi-Region replication** ที่มีการทำสำเนาข้อมูลข้าม Region โดยอัตโนมัติ เมื่อใดที่ Region หนึ่งมีปัญหา ระบบจะยังคงให้บริการจาก DynamoDB table ที่อยู่ใน Region อื่นๆ โดยไม่ต้องรอการ failover ³ ⁴ ทำให้แอปพลิเคชันสามารถอ่าน/เขียนข้อมูลที่สำรองไว้ใน Region อื่นได้ทันที ส่งผลให้บริการไม่หยุดชะงักแม้เกิดเหตุ Region ล่ม

- **AWS Key Management Service (KMS)** – บริการจัดการกุญแจเข้ารหัสของ AWS (ประเภท PaaS ด้านความปลอดภัย) ที่มี **SLA 99.999%** สำหรับทุกคำขอเรียกใช้ API ของ KMS ⁵ . เพื่อให้บรรลุ SLA ระดับนี้ AWS KMS ถูกออกแบบให้ **ไม่มี single point of failure** ภายในแต่ละ Region: โครงสร้างพื้นฐานของ KMS ในแต่ละ Region มีการทำซ้ำ (replicate) อย่างน้อย **3 Availability Zones (AZ)** และสามารถให้บริการคำขอจาก AZ ใดก็ได้ภายใน Region นั้นๆ ⁶ ซึ่งหมายความว่าหากโหนดหรือดาต้าเซ็นเตอร์ใดล้มเหลว ระบบจะสลับไปใช้อีก AZ แทนโดยไม่กระทบต่อการให้บริการ นอกจากนี้ AWS KMS ยังอัปเดตซอฟต์แวร์แบบ **zero downtime** (ไม่มีช่วงหยุดบริการเพื่อบำรุงรักษา) ผ่านกลไก rolling update อัตโนมัติ ทำให้ผู้ใช้ไม่ประสบปัญหาระบบหยุดเลย ⁷ ⁵
- **Amazon Route 53 (DNS)** – บริการระบบชื่อโดเมน (DNS) แบบกระจายของ AWS ที่ให้ **SLA 100% uptime** สำหรับการตอบสนองคำขอ DNS ⁸ (ซึ่งถือว่าสูงยิ่งกว่า 99.999%). Route 53 เป็นบริการแบบ IaaS (Infrastructure) ด้านเครือข่ายที่บรรลุ SLA สูงสุดได้ด้วยการใช้ **เครือข่าย Anycast DNS servers กระจายทั่วโลก** ที่มีการทำงานสำรองซ้อนทับกันอย่างสมบูรณ์ เมื่อมี DNS server ใดล้มหรือเครือข่ายในบางพื้นที่ขัดข้อง การออกแบบแบบ Anycast จะทำให้คำขอ DNS ของผู้ใช้ถูกส่งไปยังเซิร์ฟเวอร์ DNS ตัวอื่นที่ยังทำงานอยู่โดยอัตโนมัติ ซึ่ง AWS กล้าการันตีถึงขั้น **ไม่มีช่วง Downtime เลย** (หากมีแม้แต่นาทีเดียวก็ถือว่าผิด SLA และจะชดเชยเครดิตตามสัดส่วน) ⁸ . การบรรลุ 100% SLA นี้ต้องอาศัยโครงสร้างพื้นฐาน DNS ที่มีความซับซ้อนสูงมาก กระจายอยู่หลายภูมิภาคทั่วโลก และใช้ **โพรโตคอล BGP Anycast** เพื่อให้ DNS Query ของผู้ใช้ถูกตอบกลับจากเซิร์ฟเวอร์ที่ใกล้ที่สุดที่ยังว่างอยู่เสมอ

หมายเหตุ: บริการหลักอื่นๆ ของ AWS โดยทั่วไปมี SLA ต่ำกว่า 99.999% (เช่น EC2, EBS มักอยู่ที่ ~99.99% หรือน้อยกว่า) แต่ AWS เปิดโอกาสให้ลูกค้าสามารถสร้างสถาปัตยกรรมที่มีความทนทานสูงขึ้นเองได้ เช่น การดีพลอย VM ข้ามหลาย AZ หรือหลาย Region ซึ่งหากออกแบบดีอาจบรรลุความพร้อมใช้งานที่เข้าใกล้ 5 นัยน์ได้ แม้ AWS ไม่ได้ให้ SLA ระดับนั้นโดยตรงสำหรับบริการเดี่ยว

Microsoft Azure

Microsoft Azure มีบริการหลายอย่าง que เมื่อกำหนดค่าในรูปแบบ **High Availability (HA)** สูงสุดแล้วจะได้รับ SLA ในระดับ **99.999%** เช่นกัน ส่วนมากเป็นบริการแบบแพลตฟอร์ม (PaaS) โดยเฉพาะบริการด้านข้อมูลและเครือข่าย:

- **Azure Cosmos DB (NoSQL Database)** – บริการฐานข้อมูล NoSQL แบบกระจายของ Azure ที่ขึ้นชื่อเรื่องความทนทานและ latency ต่ำทั่วโลก. Azure รับประกัน SLA **99.999%** สำหรับ **บัญชี Cosmos DB ที่ตั้งค่าเป็นหลายภูมิภาคพร้อมเขียนหลายแห่ง (multi-region writes)** ⁹ ¹⁰ ซึ่งเป็น configuration ระดับสูงสุดของ Cosmos DB. วิธีที่ Azure บรรลุ SLA นี้คือการใช้ **การทำซ้ำข้อมูลข้ามภูมิภาค (global replication)** พร้อม **multi-master writes (active-active)** – ทุก Region จะมีสำเนาข้อมูลและรองรับการเขียนของตัวเอง และข้อมูลจะถูก replicate ไปยังทุก Region อื่นแบบเรียลไทม์ ¹⁰ ¹¹ . ดังนั้นถ้าเกิด Region ใดล้ม ผู้ใช้ยังสามารถอ่าน/เขียนที่ Region อื่นได้ทันที (ระบบจะจัดการ failover การเขียนให้อัตโนมัติ) ทำให้ฐานข้อมูลยัง “always online” อยู่เสมอ นอกจากนี้ Cosmos DB ยังออกแบบให้ **ไม่มีการหยุดบริการระหว่าง maintenance** (patching, upgrade) โดยระบบจะจัดการเบื้องหลังแบบโปร่งใสต่อผู้ใช้. Cosmos DB ถือเป็นบริการประเภท PaaS (Database-as-a-SaaS) ที่มีความพร้อมใช้งานสูงมากระดับองค์กร
- **Azure Managed Redis (Enterprise)** – บริการแคชข้อมูลในหน่วยความจำ (In-memory cache) ที่ Azure ร่วมพัฒนากับ Redis Inc. โดยเปิดตัว GA ในปี 2025. ในโหมด **Active-Active multi-region** Azure Managed Redis ให้ SLA ความพร้อมใช้งานสูงสุด **99.999%** ¹² ซึ่งเหนือกว่าแผนบริการแคชเดิมของ Azure เอง. การบรรลุ Five Nines นี้ทำได้ผ่านการ **ทำคลัสเตอร์ Redis กระจายข้ามหลาย Region แบบ Active-Active** โดยข้อมูลแคชจะทำสำเนาและ sync กันระหว่าง Region ตลอดเวลา ดังนั้นหาก Region ใดล้ม ระบบแคชยังให้บริการจากอีก Region ได้โดยไม่สูญเสียข้อมูลแคช (ด้วยการออกแบบ conflict-free replication ของ Redis). Azure Managed Redis เป็นบริการแบบ PaaS (Data caching service). หมายเหตุ: Azure ยังมีบริการ **Azure Cache for Redis** รุ่นมาตรฐานซึ่งใช้หลักการ replica

ภายใน Region (master/slave) ให้ SLA 99.9%–99.95% เท่านั้น ¹³ แต่ Azure Managed Redis (ซึ่งเป็นบริการรุ่นใหม่) ได้ยกระดับด้วยสถาปัตยกรรม Active-Active ข้าม Region จนได้ SLA 99.999% ดังกล่าว

- **Azure Managed Instance for Apache Cassandra** – บริการฐานข้อมูล NoSQL (Cassandra) แบบ Managed ที่ Azure ให้บริการ (PaaS) สำหรับรันคลัสเตอร์ Cassandra แบบยืดหยุ่น. สำหรับระบบที่ต้องการ HA สูง Azure ระบุว่าสามารถตั้งค่า Managed Cassandra ให้วางโหนดข้ามหลาย Availability Zone หรือหลาย Region และรองรับ SLA สูงสุดประมาณ **99.995%–99.999%** (โดย 99.999% จะได้เมื่อใช้งานหลาย Region อย่างเต็มรูปแบบ) ¹⁴ ¹⁵. โครงสร้างของบริการนี้คล้ายกับ Cosmos DB คืออาศัยการทำ **replication ข้ามศูนย์ข้อมูลหลายแห่ง** และมีการจัดการ failover อัตโนมัติ. เช่น การใช้งาน **Cassandra MI แบบ multi-region cluster** ที่ข้อมูลแต่ละ keyspace ทำซ้ำในหลายภูมิภาคพร้อมกัน จะช่วยลดโอกาส downtime ให้เหลือน้อยมาก (ระดับห้าหลัก) เนื่องจากไม่มีจุดล้มเหลวจุดเดียว ทั้งนี้ SLA 99.999% ของบริการนี้ถือเป็นการรับประกันในเชิงการออกแบบ (design target) โดย Azure (ข้อมูลบางส่วนจาก Azure SLA Board) ¹⁶
- **บริการ DNS และ Traffic Manager ของ Azure** – เช่นเดียวกับ AWS, Azure มีบริการเครือข่ายที่ให้ SLA สูงสุดถึง **100%** ได้แก่ **Azure DNS** (บริการจัดการโดเมน) ที่รับประกันการตอบ DNS จาก name server อย่างน้อยหนึ่งตัว **100% ของเวลา** หากผู้ใช้กำหนด DNS ให้ใช้งาน name servers ทั้ง 4 ชุดที่ Azure กำหนดไว้ ¹⁷. และ **Azure Traffic Manager** (บริการ DNS-based load balancing ข้าม region) ที่มี SLA 100% เช่นกัน ¹⁸. บริการเหล่านี้เป็นประเภท IaaS (Infrastructure Networking) ที่บรรจุ SLA สูงมากด้วยการใช้ **โครงสร้างพื้นฐาน DNS Anycast และ Load Balancer แบบกระจายทั่วโลก** เช่น Azure DNS มี name server หลักหลายตัวกระจายตามภูมิภาค (Azure ระบุว่าหากกำหนดให้ลูกค้า query DNS ผ่านทั้ง 4 nameserver จะรับประกัน 100% uptime) ¹⁷. ส่วน Traffic Manager เองก็ทำงานบนเครือข่าย DNS globally distributed เช่นกัน ดังนั้นจึงไม่มีช่วง downtime ที่บริการจะไม่ตอบสนอง (หากมี node ล่มก็มี node อื่นทดแทนได้ตลอดเวลา)

หมายเหตุ: โดยทั่วไป Azure จะให้ SLA มาตรฐาน 99.9%–99.99% สำหรับบริการ PaaS ส่วนใหญ่เมื่อใช้งานใน Region เดียว แต่ Azure เปิดให้ลูกค้า **เพิ่มความทนทานด้วยตนเอง** เช่น เปิด **Zone Redundancy** (กระจายทรัพยากรในหลาย Availability Zone ภายใน Region) จะเพิ่ม SLA ของบางบริการเป็น 99.95% หรือ 99.99% ⁹; หรือการ **เพิ่ม Region สำรอง** (Active-Standby) ซึ่งมักไม่ได้รับ SLA รวมโดยตรง แต่การออกแบบ multi-region ที่ดีสามารถลด downtime ได้มาก หากลูกค้าจัดการ failover เองอย่างถูกต้อง อย่างไรก็ตาม มีบริการไม่กี่อย่างที่ Azure รับประกัน SLA ห้าหลักโดยตรงตามที่กล่าวมาข้างต้น

Google Cloud Platform (GCP)

Google Cloud (GCP) เป็นผู้ให้บริการที่เน้นจุดเด่นด้านความน่าเชื่อถือของโครงสร้างพื้นฐานแบบ global distributed จึงมีหลายบริการระดับแพลตฟอร์มที่ให้ SLA สูงถึง **99.999%** เมื่อใช้ในรูปแบบที่รองรับ:

- **Google Cloud Spanner** – บริการฐานข้อมูลเชิงสัมพันธ์แบบกระจาย (NewSQL Database) ของ Google ที่ขึ้นชื่อเรื่อง **consistency และ availability สูง**. สำหรับ **Cloud Spanner แบบ Multi-Regional** (ฐานข้อมูลที่กระจายข้อมูลในหลายภูมิภาค) GCP ให้ SLA $\geq 99.999\%$ ¹⁹ และแม้แต่ **Spanner แบบ Dual-Regional** (2 Region) ก็รับประกัน $\geq 99.999\%$ เช่นกัน ¹⁹. Spanner เป็นบริการประเภท PaaS (Database) ที่สามารถบรรจุ SLA สูงขนาดนี้ด้วยสถาปัตยกรรม **Distributed SQL** ของ Google: ข้อมูลถูก replicate แบบ **synchronous** ไปยังหลาย datacenter ในหลาย Region พร้อมกัน ผ่านโปรโตคอล **Paxos/TrueTime** ของ Spanner ที่ทำให้ทุกการเขียนยืนยันเมื่อข้อมูลถูกเขียนใน replica โหนดส่วนใหญ่ (quorum) ในหลายที่แล้ว. นอกจากนี้ Spanner **ไม่มี downtime สำหรับ maintenance** เพราะ Google ทำ rolling upgrade และใช้เทคนิค **Transparent failover** ระหว่างโหนด. GCP ออกแบบ Spanner ให้ **ไร้ช่วงหยุดวางแผน (no planned downtime)** อย่างแท้จริงสำหรับลูกค้า ²⁰. ดังนั้นแม้มีการอัปเดตระบบหรือปัญหาใน region ใด Spanner multi-region ยังให้บริการต่อจาก region อื่นได้ (RPO=0, RTO ใกล้ 0)

- **Google Cloud Bigtable** – บริการฐานข้อมูล NoSQL แบบคอลัมน์กว้างขนาดใหญ่ของ GCP (PaaS). เดิม Bigtable ให้ SLA ต่ำกว่าบริการอื่น แต่ปัจจุบัน (หลังปี 2021) GCP ได้ปรับเพิ่ม SLA สำหรับ Bigtable สูงสุดถึง **99.999%** เช่นเดียวกับ Firestore และ Spanner ²¹ โดยเงื่อนไขคือ Bigtable instance นั้นต้องถูกตั้งค่า **แบบหลายคลัสเตอร์ (Multi-Cluster Routing)** ที่กระจายใน 3 Region ขึ้นไป ²² . การตั้งค่า **multi-cluster** จะทำให้ Bigtable มีการ **replicate** ข้อมูลระหว่างคลัสเตอร์ในแต่ละ Region และใช้การ **routing** อัตโนมัติ: แอปพลิเคชันจะเชื่อมต่อผ่าน **endpoint** ที่ระบบจะส่งคำขอไปยังคลัสเตอร์ที่พร้อมทำงาน (ถ้าคลัสเตอร์ใดใน Region หนึ่งล่มหรือช้า ระบบจะ **route** ไปอีกคลัสเตอร์ใน Region อื่น). แนวทางนี้ลดโอกาส downtime เหลือน้อยมาก. Bigtable จึงถือเป็น PaaS ที่ลูกค้าสามารถเลือกปรับแต่ง HA เอง – **Single cluster** ใน 1 Zone จะได้ SLA ~99.9%, แต่ **multi-cluster** ข้าม 3 Region ให้ SLA 99.999% ²² . ทั้งนี้ GCP ได้ระบุว่าการออกแบบ Bigtable ภายในมีการใช้งาน **infrastructure** เดียวกันที่ Google ใช้ในบริการสำคัญ (เช่น Search, YouTube) ซึ่งมีการทำ Geo-replication และ automation** เพื่อรองรับระดับ availability สูงสุด

- **Google Cloud Firestore (Datastore)** – บริการฐานข้อมูล NoSQL แบบ document ของ Google (ให้บริการผ่านทั้ง GCP และ Firebase) ซึ่งมีโหมดการใช้งาน **multi-region instance**. GCP ให้ SLA **99.999%** สำหรับ **Cloud Firestore** แบบ **Multi-Region** และ **99.99%** สำหรับแบบ **Regional** ²³ . Firestore multi-region บรรลุ five nines ด้วยการ **เก็บข้อมูลแบบหลายสำเนาข้ามภูมิภาค** (โดยแต่ละ multi-region location ของ Firestore เช่น “nam5” หรือ “eur3” จะประกอบด้วย Region หลัก 2 แห่งและ Region สำรอง 1 แห่งที่อยู่คนละภูมิภาคกัน) ข้อมูลทุกชิ้นจะถูกบันทึกลงในอย่างน้อย 3 data center ที่กระจายภูมิศาสตร์ ทำให้ทนทานต่อการสูญเสียทั้ง data center หรือแม้แต่ทั้ง Region. หาก Region หลักล่ม ระบบ Firestore จะสลับการให้บริการไปยัง Region สำรองอัตโนมัติ ผู้ใช้แทบไม่รู้สึกลง downtime. Firestore ยังออกแบบเป็นบริการ **serverless** ที่ Google จัดการ scaling และ failover ให้ทั้งหมด จึงไม่มี maintenance downtime จากฝั่งผู้ใช้. ด้วยสถาปัตยกรรมนี้ Google จึงมั่นใจการันตี SLA ห้าเก้าสำหรับ Firestore multi-region ²³

- **Google Cloud DNS** – บริการ DNS แบบ Hosted ของ GCP (ประเภท IaaS ด้านเครือข่าย) ซึ่งเช่นเดียวกับคู่แข่ง Google ได้โฆษณา **SLA 100%** สำหรับ Cloud DNS โดยระบุว่าจะมีการตอบสนองคำถาม DNS จาก authoritative name server อย่างน้อยหนึ่งตัว **ตลอดเวลา 100%** ²⁴ . Cloud DNS บรรลุสิ่งนี้ด้วยโครงสร้างพื้นฐาน DNS ของ Google ที่กระจายอยู่ทั่วโลก และใช้ **Anycast** ในการประกาศ IP ของ name server ร่วมกันหลายแห่ง. เมื่อผู้ใช้ส่ง DNS query มา ระบบ Anycast จะนำทางคำขอนั้นไปยังศูนย์ข้อมูลใกล้ที่สุดซึ่งยังให้บริการ DNS ได้. Google ระบุว่าบริการ Cloud DNS เป็นบริการหนึ่งในไม่กี่ตัวที่ “ให้ 100% availability SLA” ใน Google Cloud ²⁴ . (แน่นอน ในทางปฏิบัติ 100% หมายถึงหากมี downtime ใดๆ ก็จะได้รับเครดิตชดเชยตาม SLA แต่เป้าหมายคือไม่มี downtime เลย)

หมายเหตุ: GCP ยังมีบริการอื่นที่มี HA สูง เช่น **Google Cloud Load Balancing** ที่เป็น global load balancer อาจมี SLA ~99.99%, หรือ **Google Kubernetes Engine (GKE)** ที่เมื่อเปิดใช้งาน regional clusters (หลาย zone) จะได้ SLA สูงกว่า single zone เป็นต้น แต่บริการที่มีการันตีระดับ 99.999% โดยตรงจะจำกัดอยู่ในกลุ่ม database และ infrastructure core ที่กล่าวมา. Google มีแนวคิดคล้าย AWS/Azure คือให้ลูกค้าออกแบบระบบหลาย region เองเพื่อความทนทาน แต่สำหรับบางบริการที่ Google สร้างมาให้ global ตั้งแต่แรก (Spanner, Firestore, Bigtable) จึงให้ SLA สูงกว่าปกติ

IBM Cloud

IBM Cloud ในฐานะผู้ให้บริการคลาวด์สำหรับองค์กร มักเน้นเรื่องความน่าเชื่อถือเช่นกัน แต่ **โดยทั่วไป IBM Cloud จะประกาศ SLA อยู่ที่ประมาณ 99.9%–99.99%** สำหรับบริการส่วนใหญ่. อย่างไรก็ตาม IBM ได้วางสถาปัตยกรรมบริการบางอย่างให้รองรับ

availability สูงถึงระดับ **99.999%** แม้จะไม่ค่อยได้โฆษณาเป็น SLA ตรงๆ (บางกรณีระบุเป็น SLO ซึ่งไม่ผูกพันการชดเชย). ตัวอย่างบริการของ IBM Cloud ที่มีการออกแบบเพื่อให้ได้ห้าเก้า ได้แก่:

- **IBM Cloud DNS Services** – บริการจัดการ DNS บน IBM Cloud (เทียบได้กับ Route 53/Azure DNS) ที่ IBM ออกแบบให้ทำงานแบบ **global anycast** เช่นกัน. IBM ระบุไว้ในเอกสารว่าบริการ DNS นี้ถูกออกแบบมาเพื่อให้มี **Availability Target (SLO) 99.999%** ²⁵ ²⁶ โดยโครงสร้างประกอบด้วย DNS Server หลายตัวกระจายอยู่ใน **หลาย Multi-Zone Region (MZR)** ทั่วโลก (เช่น ดัลลัส, วอชิงตัน ดี.ซี., ลอนดอน, แฟรงก์เฟิร์ต, โตเกียว ฯลฯ) และใช้ Global Load Balancer ช่วยกระจายกราฟฟิการจัดการ DNS config ด้วย ²⁷ ²⁸ . หาก Region ใดเกิดปัญหา ระบบจะเปลี่ยนเส้นทางคำขอไปยัง Region อื่นโดยอัตโนมัติ ทำให้ลูกค้าใช้งาน API และบริการ DNS ได้ต่อเนื่อง แม้ IBM จะไม่ได้การันตี SLA 100% แบบเป็นทางการ (เพราะ 99.999% ถูกระบุเป็นเป้าหมายการออกแบบ SLO และ “not a warranty” ในเอกสาร ²⁶) แต่ประวัติการให้บริการ DNS ของ IBM Cloud ก็พยายามรักษาระดับใกล้เคียงห้าเก้าเสมอ
- **IBM Cloud Direct Link** – บริการเชื่อมต่อเครือข่ายส่วนตัวเข้าสู่ IBM Cloud (คล้าย AWS Direct Connect, Azure ExpressRoute) ซึ่ง IBM ระบุ SLO การออกแบบไว้ที่ **99.999%** เช่นกันในกรณีที่ลูกค้าเชื่อมต่อแบบ Dual-hub (มีสำรองสองจุดขึ้นไป) ²⁹ . โดย IBM จะมีคู่ link ที่อยู่คนละศูนย์ข้อมูล หากเส้นหนึ่งล้มอีกเส้นจะยังทำงาน ทำให้การเชื่อมต่อไม่หลุด (คล้ายกับ telco ที่มี SLA เครือข่ายใกล้เคียงห้าเก้าเมื่อเชื่อมต่อวงจรสำรอง). อย่างไรก็ตาม เช่นเดียวกับ DNS, ค่านี้มักเป็น SLO ภายใน. **IBM Cloud Internet Services (CIS)** ที่เป็นบริการ CDN/DNS/ความปลอดภัย (ร่วมกับ Cloudflare) ก็ได้รับการออกแบบให้ highly available สูงมากเช่นกัน ซึ่งคาดการณ์ว่ามีเป้าหมายใกล้เคียง 99.99%–99.999% (CIS ใช้โครงข่ายของ Cloudflare ที่มี PoP ทั่วโลกกว่า 200 แห่ง)
- **บริการโครงสร้างพื้นฐาน IBM Cloud อื่นๆ** – IBM Cloud มีแนวคิด **Multi-Zone Region (MZR)** คือแต่ละ Region ของ IBM จะมีอย่างน้อย 3 Availability Zone ที่แยกกัน (เช่น Region โตเกียวมี 3 DC แยกกัน). บริการอย่าง **IBM Cloud Databases (เช่น IBM Cloudant, IBM Db2, etc.)** หากลูกค้าตั้งค่าให้ **กระจายสำเนาข้อมูลลงทุก AZ (Zone redundancy)** ก็จะได้รับ SLA สูงสุดที่ IBM ให้ไว้ (~99.99%). และหากต้องการสูงกว่านั้น IBM มักแนะนำให้ทำ Active-Active ข้าม Region ด้วยตนเอง (ซึ่ง IBM อาจไม่การันตี SLA ระดับห้าเก้าแต่ลูกค้าสามารถออกแบบเพื่อหวั่ง downtime ต่ำมากได้). พอสรุปได้ว่า IBM Cloud **ยังไม่มีบริการที่ประกาศ SLA 99.999% แบบมีการชดเชยโดยตรง**, แต่มีการ **ออกแบบระบบ (architecture)** ให้รองรับถึงระดับนั้นในบางบริการ. ลูกค้ากลุ่มองค์กรของ IBM มักจะประเมินร่วมกับ IBM เป็นกรณีๆ ไปถึงความเสี่ยงและ SLO ที่ต้องการ แล้วออกแบบระบบ (เช่น ระบบเมนเฟรม Z Cloud หรือ Power Systems cloud) ให้มี HA ตามเป้า ซึ่งบางเคสอาจอ้างว่าได้ 99.999% แต่ไม่ได้อยู่ใน SLA สาธารณะที่ประกาศทั่วไป

ต้นทุนและการลงทุนในการบรรลุ SLA 99.999%

การจะสร้างระบบหรือบริการให้มีความพร้อมใช้งานระดับ **ห้าเก้า** ต้องอาศัยทั้ง **ทรัพยากรโครงสร้างพื้นฐานที่ซับซ้อนอย่างมาก** และ **กระบวนการดำเนินงานที่ยืดเยื้อและอัตโนมัติ** ซึ่งทั้งหมดนี้มี **ต้นทุนสูงมาก**. อุตสาหกรรมมักกล่าวกันว่า “ทุกๆ 9 ที่เพิ่มขึ้น จะเพิ่มต้นทุนขึ้นประมาณ 10 เท่า” ³⁰ — กล่าวคือ การจะขยับจาก 99.9% (สามเก้า) ไปสู่ 99.99% หรือ 99.999% ต้องลงทุนเพิ่มทั้งฮาร์ดแวร์, ดาต้าเซ็นเตอร์, ทีมวิศวกร, และระบบการจัดการอัตโนมัติอีกหลายเท่าตัว.

เหตุผลที่ต้นทุนสูงเป็นเพราะต้องมี **ความซ้ำซ้อน (redundancy)** ในทุกองค์ประกอบ: ผู้ให้บริการ Cloud ต้องมี **ดาต้าเซ็นเตอร์หลายแห่งในหลายภูมิภาค** ที่ข้อมูลและบริการเดียวกันทำงานซ้ำกันไปหมด ทุกจุดต้องสามารถรับภาระแทนกันได้ทันทีเมื่อส่วนใดส่วนหนึ่งล้มเหลว. ยกตัวอย่างเช่น AWS KMS ใช้การทำงานพร้อมกันใน **3 AZ** ภายในหนึ่ง Region ⁶ ดังนั้นหากต้องการแก้ปัญหาในระดับทั้ง Region ล่ม ก็ต้องมีสำรองใน Region อื่นเพิ่มด้วย. ในกรณี DynamoDB Global Tables หรือ Cosmos DB จะต้อง **เก็บข้อมูลหลายชุดในหลาย Region** เสมอ ซึ่งหมายถึงค่าใช้จ่ายด้านที่เก็บข้อมูลและการส่งข้อมูลข้ามประเทศที่สูงขึ้นมาก. จากตาราง SLA Cosmos DB ด้านล่าง จะเห็นว่าการเปิด **Availability Zone** ต้องเพิ่มทรัพยากร ~25% (1.25x) และการเพิ่ม Region อีกเท่าตัวจะคูณค่าใช้จ่ายตามจำนวน Region (เช่น 2 Region ~2x) ³¹ และถ้าเปิด multi-region writes (เพื่อ 5 นัยน์) ก็จะยิ่ง

เพิ่ม overhead เข้าไปอีก. กล่าวคือการได้ SLA 99.999% ใน Cosmos DB ต้องจ่ายค่า RU/s และ storage อย่างน้อย ~2.5 เท่าของการใช้ single region แบบไม่มี AZ

นอกจากฮาร์ดแวร์และที่ตั้งแล้ว ผู้ให้บริการ Cloud ยังต้องลงทุนใน **ระบบซอฟต์แวร์สำหรับตรวจจับและแก้ไขปัญหาด้านอัตโนมัติแบบเรียลไทม์** เพราะเมื่อเป้าหมาย downtime คือน้อยกว่า ~5 นาทีต่อปี “เวลา” คือสิ่งสำคัญมาก – แทนที่ไม่มีเวลาสำหรับการแก้ไขปัญหาแบบ manual อีกต่อไป ³² . ตัวอย่างเช่น AWS ต้องพัฒนา **ระบบตรวจจับความผิดปกติและ rollback การปรับปรุงที่อัตโนมัติ** โดยทันทีที่ระบบเริ่มเกิด error จากการ deploy code ใหม่ ระบบ CD ของ AWS จะ rollback โค้ดนั้นเองภายในไม่กี่วินาที ก่อนที่วิศวกร on-call จะกันรู้ตัวด้วยซ้ำ ³³ . สิ่งเหล่านี้ต้องใช้การลงทุนสร้าง **pipeline และ orchestration** ขึ้นสูง เพื่อให้ทุกขั้นตอน (deployment, failover, recovery) เกิดขึ้นโดยอัตโนมัติและรวดเร็ว. ผู้ให้บริการยังต้องมี **ทีม SRE/วิศวกร** คอยเฝ้าระวังตลอด 24/7 (หมุนเวียนกันตลอดทั้งปีไม่มีหยุด) เพื่อตรวจสอบเหตุผิดปกติที่ automation จัดการไม่ได้. บริษัทขนาดใหญ่บางแห่งชี้ว่าเพื่อขยับจากระบบสามเท่าไปห้าเท่า อาจต้อง **เพิ่มจำนวนพนักงานปฏิบัติการขึ้นหลายเท่า** เพื่อครอบคลุมงานดูแลระบบระดับนี้ตลอดเวลา ³⁴ ³⁵

โดยสรุป ผู้ให้บริการ Cloud ชื่อนำสามารถให้ SLA 99.999% ได้เฉพาะบริการที่พวกเขาได้ลงทุนสร้างสถาปัตยกรรมแบบ *Fault-Tolerant* สูงสุด เช่น ระบบฐานข้อมูลหรือระบบ DNS กระจายข้ามโลก ซึ่งต้องมี **การลงทุนระดับโครงสร้างพื้นฐานมหาศาล** (ถ้าตีเป็นเตอร์หลายแห่ง, ลิขสิทธิ์หรือค่าความเร็วสูงซ้ำซ้อน, อุปกรณ์ฮาร์ดแวร์สำรองทุกจุด) และ **การพัฒนาซอฟต์แวร์ระบบขั้นสูง** (algorithm จัดการ replicated state, consensus protocol, anycast routing, etc.) รวมถึง **การทดสอบ DR (Disaster Recovery) เป็นประจำ** เพื่อมั่นใจว่าระบบสำรองจะทำงานได้จริงเมื่อเกิดเหตุฉุกเฉิน. ทั้งหมดนี้เป็นต้นทุนที่ทำให้บริการ five nines มักมีราคาแพง และโดยมากผู้ให้บริการจะจำกัดวงไว้เฉพาะบริการที่ลูกค้ายอมจ่ายแพงเพื่อ HA สูงสุด หรือเป็นส่วน critical ที่ขาดไม่ได้ในโครงสร้าง Cloud (เช่น DNS, Key Management). สำหรับลูกค้าทั่วไป การเลือก SLA 99.99% (four nines) มักจะเพียงพอและคุ้มค่ากว่ามากในแง่ต้นทุนเมื่อเทียบกับห้าเท่าที่ค่าใช้จ่ายเพิ่มแบบทวีคูณ ³⁰

ด้านล่างเป็นตารางสรุปเปรียบเทียบผู้ให้บริการแต่ละราย, บริการที่มี SLA ~99.999%, เปอร์เซนต์ SLA และประเภทของบริการนั้น ๆ เพื่อเป็นข้อมูลประกอบการพิจารณา:

ผู้ให้บริการ Cloud	บริการ (ผลิตภัณฑ์)	SLA Uptime	ประเภทบริการ
AWS (Amazon Web Services)	Amazon DynamoDB (Global Tables)	99.999% ²	PaaS – NoSQL Database
	AWS Key Management Service (KMS)	99.999% ⁵	PaaS – Security / Key Management
	Amazon Route 53 (DNS)	100% ⁸ (≈ 99.999%)	IaaS – DNS (Networking)
Microsoft Azure	Azure Cosmos DB (Multi-region writes)	99.999% ⁹ ¹⁰	PaaS – NoSQL Database
	Azure Managed Redis (Active-Active)	99.999% ¹²	PaaS – In-memory Cache
	Azure Managed Cassandra MI	99.999% (สูงสุด) ³⁶	PaaS – NoSQL Database
	Azure DNS	100% ¹⁷ (≈ 99.999%)	IaaS – DNS Service

ผู้ให้บริการ Cloud	บริการ (ผลิตภัณฑ์)	SLA Uptime	ประเภทบริการ
Google Cloud (GCP)	Azure Traffic Manager	100% ¹⁸ (≈ 99.999%)	IaaS – Global Load Balancing (DNS)
	Cloud Spanner (Multi/Dual-region)	99.999% ¹⁹	PaaS – Distributed SQL DB
	Cloud Bigtable (Multi-cluster, 3+ regions)	99.999% ²²	PaaS – NoSQL Database
	Cloud Firestore (Multi-region)	99.999% ²³	PaaS – NoSQL Document DB
	Cloud DNS	100% ²⁴ (≈ 99.999%)	IaaS – DNS (Networking)
IBM Cloud	IBM Cloud DNS Services	99.999% (เป้าหมาย/ SLO) ²⁵ ²⁶	IaaS – DNS (Networking)
	IBM Cloud Direct Link (Dual)	99.99%–99.999% (SLO) ²⁹	IaaS – Network Connectivity
	อื่นๆ (ออกแบบ High Availability)	สูงสุด ~99.99%	PaaS – Database, etc.
Oracle Cloud (OCI) (ข้อมูลเสริม)	Autonomous Database (with Data Guard)	99.995% ³⁷ ³⁸	PaaS – Database
	Oracle NoSQL Database Cloud	99.995% ³⁸	PaaS – NoSQL Database
UpCloud (ผู้ให้บริการรายอื่น)	UpCloud Cloud Servers (ตั้งแต่ พ.ศ. 2025)	99.999% ³⁹	IaaS – VM/Storage

หมายเหตุ: ผู้ให้บริการรายอื่นๆ เช่น *Alibaba Cloud* ก่อนหน้านี้ประกาศ SLA สูงสุด ~99.995% สำหรับ instances แบบ multi-zone ⁴⁰ (ยังไม่ถึง 99.999%), *Oracle Cloud* มี SLA สูงสุด 99.995% ในบริการฐานข้อมูลและ NoSQL บางตัว ³⁷, ส่วน *UpCloud* ผู้ให้บริการ Cloud จากยุโรปรายงานว่าให้ SLA 99.999% แก่ VM ทุกเครื่องตั้งแต่ปี 2025 ³⁹. อย่างไรก็ตาม “ห้าเก้า” ยังคงเป็นระดับ SLA ที่ท้าทายและมีน้อยบริการมากที่จะรับประกัน จึงมักถูกใช้เป็นจุดขายเฉพาะบริการที่ออกแบบมาพิเศษ และลูกค้าที่ต้องการระดับนี้ควรตรวจสอบเงื่อนไข SLA อย่างละเอียด (เช่น ต้องใช้หลาย Region/Zone พร้อมกัน, ต้องปฏิบัติตามแนวทางที่ผู้ให้บริการกำหนด เช่น config DNS ครอบคลุม nameserver เป็นต้น) เพื่อให้การรับประกัน SLA มีผลจริง

¹ Achieving “five nines” in the cloud for justice and public safety | AWS Public Sector Blog

<https://aws.amazon.com/blogs/publicsector/achieving-five-nines-cloud-justice-public-safety/>

² Amazon DynamoDB Service Level Agreement

<https://aws.amazon.com/dynamodb/sla/>

³ ⁴ global-tables

<https://aws.amazon.com/dynamodb/global-tables/>

5 6 7 Resilience in AWS Key Management Service - AWS Key Management Service

<https://docs.aws.amazon.com/kms/latest/developerguide/disaster-recovery-resiliency.html>

8 Running recovery-oriented applications with Amazon Route 53 Application Recovery Controller, AWS CI/CD tools, and Terraform | Networking & Content Delivery

<https://aws.amazon.com/blogs/networking-and-content-delivery/running-recovery-oriented-applications-with-amazon-route-53-application-recovery-controller-aws-ci-cd-tools-and-terraform/>

9 31 High availability (Reliability) in Azure Cosmos DB for NoSQL | Microsoft Learn

<https://learn.microsoft.com/en-us/azure/reliability/reliability-cosmos-db-nosql>

10 11 Distribute data globally with Azure Cosmos DB | Microsoft Learn

<https://learn.microsoft.com/en-us/azure/cosmos-db/distribute-data-globally>

12 Azure Managed Redis Launches as Enterprise-Ready, Fully Managed Cloud Data Platform | Windows Forum

<https://windowsforum.com/threads/azure-managed-redis-launches-as-enterprise-ready-fully-managed-cloud-data-platform.366807/>

13 Azure Cache for Redis

<https://azure.microsoft.com/en-us/products/cache>

14 16 18 36 Azure SLA Board

<https://azurecharts.com/sla>

15 Azure Managed Instance for Apache Cassandra vs. FlashGrid

<https://slashdot.org/software/comparison/Azure-Managed-Instance-for-Apache-Cassandra-vs-FlashGrid/>

17 Azure DNS SLA: Updated to 100% - Microsoft Azure

<https://azure.microsoft.com/en-us/updates/azuredns100sla/>

19 Cloud Spanner Service Level Agreement (SLA) | Google Cloud

<https://cloud.google.com/spanner/sla>

20 Google Cloud Spanner update includes SLA that promises less than ...

<https://techcrunch.com/2017/11/14/google-cloud-spanner-update-include-sla-that-promises-less-than-five-minutes-of-downtime-per-year/>

21 22 Google Cloud Improves SLA for Bigtable and Adds New Security Features - InfoQ

<https://www.infoq.com/news/2021/06/bigtable-sla-security/>

23 NoSQL for the serverless age: Announcing Cloud Firestore general availability and updates - Google Developers Blog

<https://developers.googleblog.com/nosql-for-the-serverless-age-announcing-cloud-firestore-general-availability-and-updates/>

24 How to leverage global address resolution using Cloud DNS in ...

<https://strategicfocus.com/2021/05/27/how-to-leverage-global-address-resolution-using-cloud-dns-in-google-cloud-vmware-engine/>

25 26 27 28 IBM Cloud Docs

<https://cloud.ibm.com/docs/dns-svcs?topic=dns-svcs-ha>

29 Understanding high availability for Direct Link - IBM Cloud Docs

<https://cloud.ibm.com/docs/dl?topic=dl-ha>

30 What is Five 9s Availability? Do you really need 99.999% Server Uptime?

<https://www.nobl9.com/resources/do-you-really-need-five-nines>

32 33 Down for less than four minutes a month: how AWS deploys code

<https://graphite.dev/blog/how-amazon-deploys-code>

34 35 domain name system - Achieving Five Nines - Server Fault

<https://serverfault.com/questions/466280/achieving-five-nines>

37 Autonomous Database now provides a 99.995% availability SLA ...

<https://blogs.oracle.com/datawarehousing/post/autonomous-database-updated-sla>

38 Oracle Cloud Infrastructure Service Level Agreement (SLA)

<https://www.oracle.com/cloud/sla/>

39 Introducing The Five 9's - New And Improved Service Level - UpCloud

<https://upcloud.com/blog/introducing-the-five-9s-new-and-improved-service-level>

40 Alibaba Cloud upgrades SLA for multi-zone instances - DCD

<https://www.datacenterdynamics.com/en/news/alibaba-cloud-upgrades-sla-multi-zone-instances/>