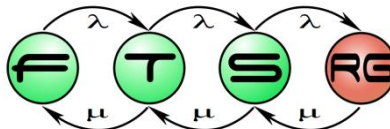


Felhő számítástechnika

A kiberfizikai rendszerek (VIMIMA02) tárgyban

Kocsis Imre, ikocsis@mit.bme.hu

2017. ősz



Felhő számítástechnika

Mi van ma a „felhőben”?

Virtuális gép
(Amazon EC2)

Adatbázis
(Amazon RDS)

...

Alkalmazás
(Lotus)

Alkalmazáserver
(Google App Engine)

Cloud
Computing

Trend: funkciók/képességek (internet-elérésű)
szolgáltatásként (is) hozzáférhetőek legyenek

Definíció...?

A „számítási felhők” egy modell, amely lehetővé teszi a hálózaton keresztül való, kényelmes és széles körű hozzáférést konfigurálható számítási erőforrások egy megosztott halmazához.

- NIST 800-145 alapján
- Tulajdonságok, szolgáltatási és telepítési modellek

Eredetiben

*Cloud computing is a **model** for enabling ubiquitous, convenient, **on-demand network access** to a **shared pool** of configurable computing resources (e.g., networks, servers, storage, applications, and services) that can be **rapidly provisioned and released** with minimal management effort or service provider interaction.*

- NIST 800-145

Alapvető tulajdonságok

- Széles körű hálózati hozzáférés
 - Nem csak az Internet
- Igény szerinti önkiszolgálás
- „Resource pooling”
 - „Multi-tenant model”: több bérlő egyszerre
 - Dinamikus ügyfelekhez rendelés
 - Bérlői kontroll: legfeljebb magasabb absztrakciós szinten

Alapvető tulajdonságok

- Rugalmas fel- és leskálázás
 - Látszólag végtelen,
 - akármikor előfizethető erőforrások
- Mért szolgáltatások
 - Szolgáltatás/erőforrás „használata”
 - Sokszor: használat alapú számlázás

Szolgáltatás-terminológia

"Algorithm as a Service"

Alkalmazás
SaaS

Szoftver-környezet / -platform
PaaS

Számítási erőforrás
IaaS

Adattárolás
DaaS

Kommunikáció
CaaS

Szoftver kernel

Lehet együtt is IaaS (pl.
NIST terminológia)

Firmware / Hardware
HaaS

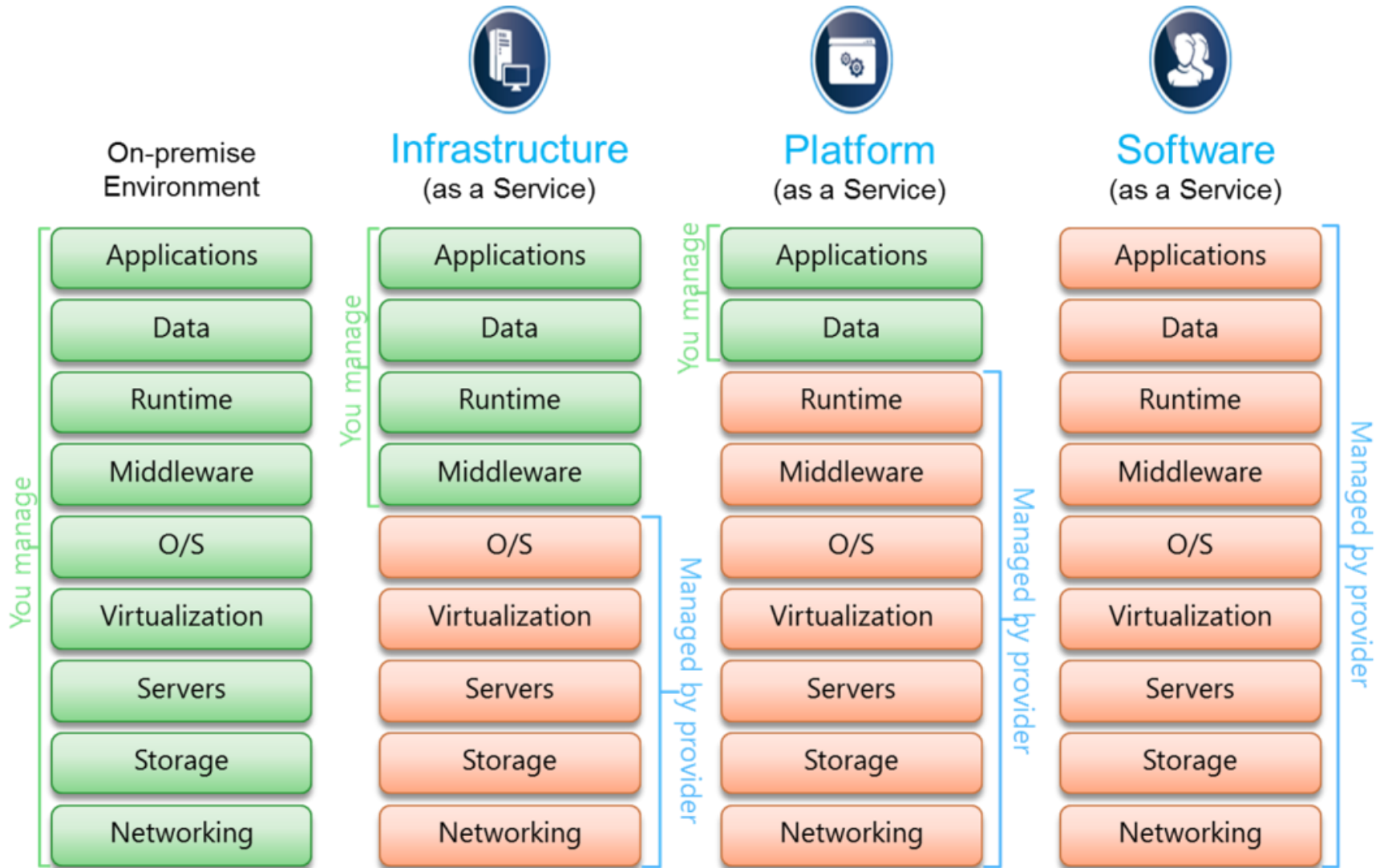
SaaS

- Képesség: szolgáltató ***alkalmazásainak*** használata
 - Hozzáférés: jellemzően vékony kliens
 - Nem új koncepció
- Példák
 - Google Apps
 - Salesforce CRM
 - LotusLive
 - Microsoft Business Productivity Online Suite (BPOS)
- Néhány igen sikeres terület: kollaboráció, könyvelés, CRM, ERP, HRM, CM, PM, ...

- Képesség: saját/beszerzett alkalmazás telepítése bérelt **futtatókörnyezetbe**
 - Adott környezeti szolgáltatások
 - Adott használható API-k, nyelvek
 - Konfigurálható környezet
 - Korlátozhatja az alkalmazás-modellt
- Google AppEngine
- Microsoft Windows Azure Platform
- Amazon Beanstalk

- **Képesség: alapvető számítási erőforrások foglalása**
 - A felhasználó „tetszőleges” szoftvert futtat
 - Jellemzően logikai/virtuális erőforrások
 - Kontroll: OS, tárolás, alkalmazások, hálózati aspektusok *egy része*
- **Amazon Elastic Compute Cloud (EC2)**
 - Xen alapú virtualizáció
 - Egyre teljesebb ökoszisztéma
 - Az alapszolgáltatás: „tömegtermék”
 - Érdekesség: gépidőre licitálás („bidding”)

Szolgáltatásmodellek összehasonlítása



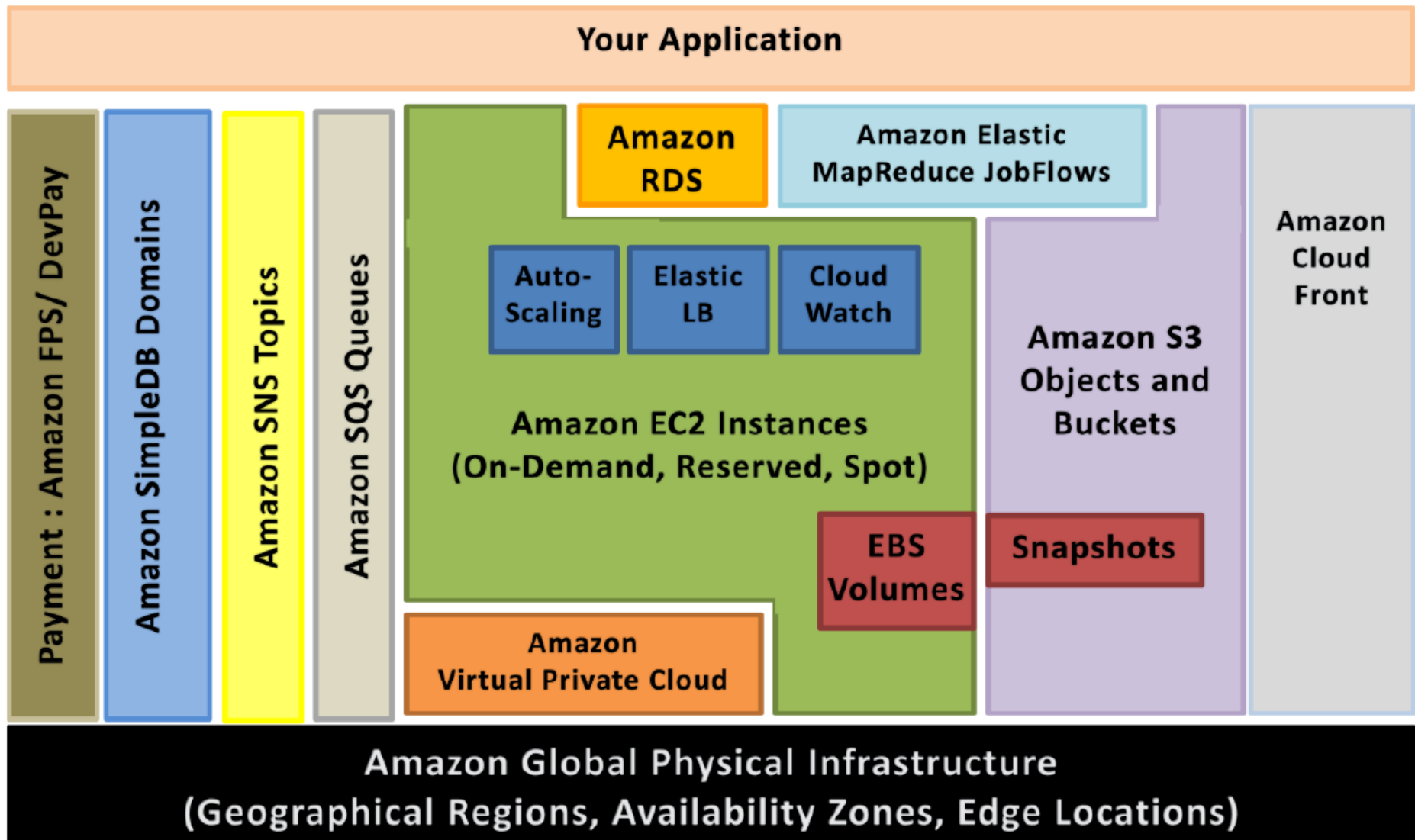
Forrás: <http://cloud.dzone.com/articles/introduction-cloud-computing>

Amazon EC2

- Infrastructure as a Service
 - EC2: sokáig „A” Cloud Computing (IaaS-re)
- Nem csak csupasz OS lehet
 - DB2, WebSphere, InfoSphere, Lotus Forms, Windows Server 2003/2008, MS SQL, ...
- Szoros integráció a többi Amazon Web Service-szel



Amazon Web Services



Forrás: <http://www.zdnet.com/blog/>

Amazon Web Services

Compute & Networking

-  **Direct Connect**
Dedicated Network Connection to AWS
-  **EC2**
Virtual Servers in the Cloud
-  **Route 53**
Scalable Domain Name System
-  **VPC**
Isolated Cloud Resources

Storage & Content Delivery

-  **CloudFront**
Global Content Delivery Network
-  **Glacier**
Archive Storage in the Cloud
-  **S3**
Scalable Storage in the Cloud
-  **Storage Gateway**
Integrates On-Premises IT Environments with AWS

Analytics

-  **Data Pipeline**
Orchestration for Data-Driven Workflows
-  **Elastic MapReduce**
Managed Hadoop Framework
-  **Kinesis**
Real-time Processing of Streaming Big Data

App Services

-  **CloudSearch**
Managed Search Service
-  **Elastic Transcoder**
Easy-to-use Scalable Media Transcoding
-  **SES**
Email Sending Service
-  **SNS**
Push Notification Service
-  **SQS**
Message Queue Service
-  **SWF**
Workflow Service for Coordinating Application Components

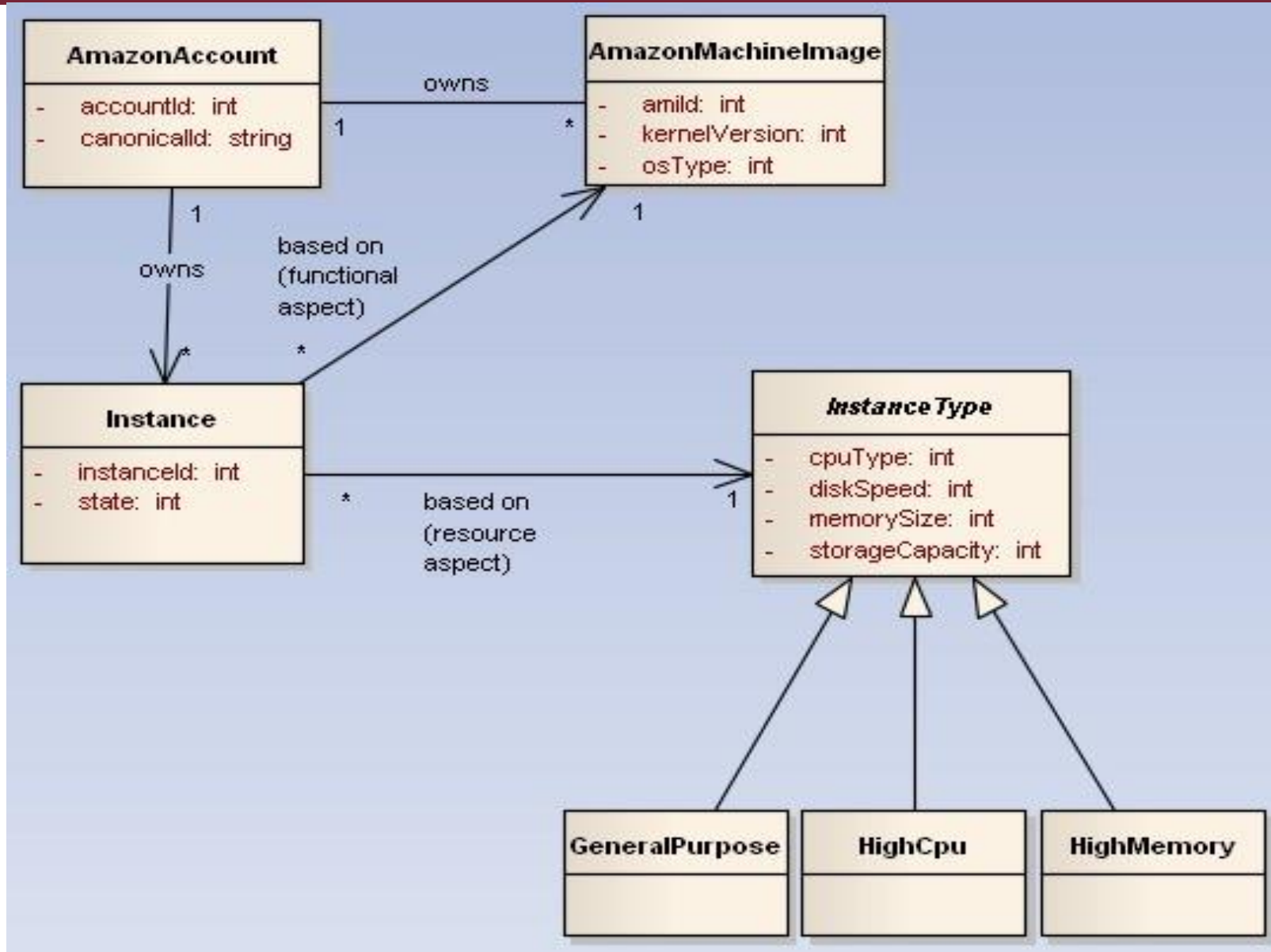
Database

-  **DynamoDB**
Predictable and Scalable NoSQL Data Store
-  **ElastiCache**
In-Memory Cache
-  **RDS**
Managed Relational Database Service
-  **Redshift**
Managed Petabyte-Scale Data Warehouse Service

Deployment & Management

-  **CloudFormation**
Templated AWS Resource Creation
-  **CloudTrail**
User Activity and Change Tracking
-  **CloudWatch**
Resource and Application Monitoring
-  **Elastic Beanstalk**
AWS Application Container
-  **IAM**
Secure AWS Access Control
-  **OpsWorks**
DevOps Application Management Service

Amazon EC2 - alapfogalmak



- Alapvető műveletek
- Példányok létrehozása
- Nagy(ob)ban építkezés: pl. AWS CloudFormation
 - Példa: https://s3-us-west-2.amazonaws.com/cloudformation-templates-us-west-2/WordPress_Multi_AZ.template (jsonviewer.stack.hu)

Az infrastruktúra a szolgáltatás mögött

North America



US East (Northern Virginia) Region

EC2 Availability Zones: 5*

Launched 2006

US West (Oregon) Region

EC2 Availability Zones: 3

Launched 2011

US West (Northern California) Region

EC2 Availability Zones: 3*

Launched 2009

AWS GovCloud (US) Region

EC2 Availability Zones: 2

Launched 2011

Asia Pacific



Asia Pacific (Singapore) Region

EC2 Availability Zones: 2

Launched 2010

Asia Pacific (Sydney) Region

EC2 Availability Zones: 2

Launched 2012

Asia Pacific (Tokyo) Region

EC2 Availability Zones: 3

Launched 2011

China (Beijing) Region

EC2 Availability Zones: 1 Coming Soon

[Learn more at www.amazonaws.cn](http://www.amazonaws.cn)

Az infrastruktúra a szolgáltatás mögött

Europe / Middle East / Africa



EU (Ireland) Region

EC2 Availability Zones: 3

Launched 2007

EU (Frankfurt) Region

EC2 Availability Zones: 2

Launched 2014

AWS Edge Locations: Amsterdam, The Netherlands (2), Dublin, Ireland, Frankfurt, Germany (3), London, England (3), Madrid, Spain, Marseille, France, Milan, Italy, Paris, France (2), Stockholm, Sweden, and Warsaw, Poland

South America



São Paulo Region

EC2 Availability Zones: 2

Launched 2011

AWS Edge Locations: Rio de Janeiro, Brazil, São Paulo, Brazil

Az infrastruktúra a szolgáltatás mögött

- Titkos, de valószínűsíthetően óriási méretek
- Min. 1,4 millió, de inkább 2,8-5,6 kiszolgáló*
- Konténerizáció, saját link adatközpontok között, integrált saját CDN, ...
- Más szolgáltatók: kisebbek
 - De gondoljunk bele, hogy -1 nagyságrend mi...

*Forrás: <http://www.enterprisetech.com/2014/11/14/rare-peek-massive-scale-aws/>

Gartner IaaS MQ 2014



Jelentősége

- Ha nem építitek/üzemeltetitek, akkor használni fogjátok
 - Adatbiztonsági problémák ellenére rég nem „shadow IT”
 - Privát és hibrid cloudok!
 - Üzleti megfontolások
- A (virtualizált) infrastruktúrát érteni kell
 - Skálázás-hibatűrés-(~)biztonság az alkalmazásban
 - dinamikus huzalozás
- XaaS: MONaaS/MaaS, **NFVlaaS**, UCaaS, NaaS, D(esktop)aaS, ...
- IoT: „the swarm on the edge of the cloud”

Alapvető problémák

- Mikor kell / érdemes / nem érdemes / tilos felhőt alkalmazni?
- Hogyan teszünk különbséget?
- Teljesítmény?

Néhány jellemző használati eset

- Kiegészítő képességek és kompetenciák
- „Core and context”
- „Cash Flow”
- (Dinamikus igények miatt nem megfelelő) kapacitás
- Üzletmenet-folytonosság és katasztr. helyreállítás
- „Sebesség”

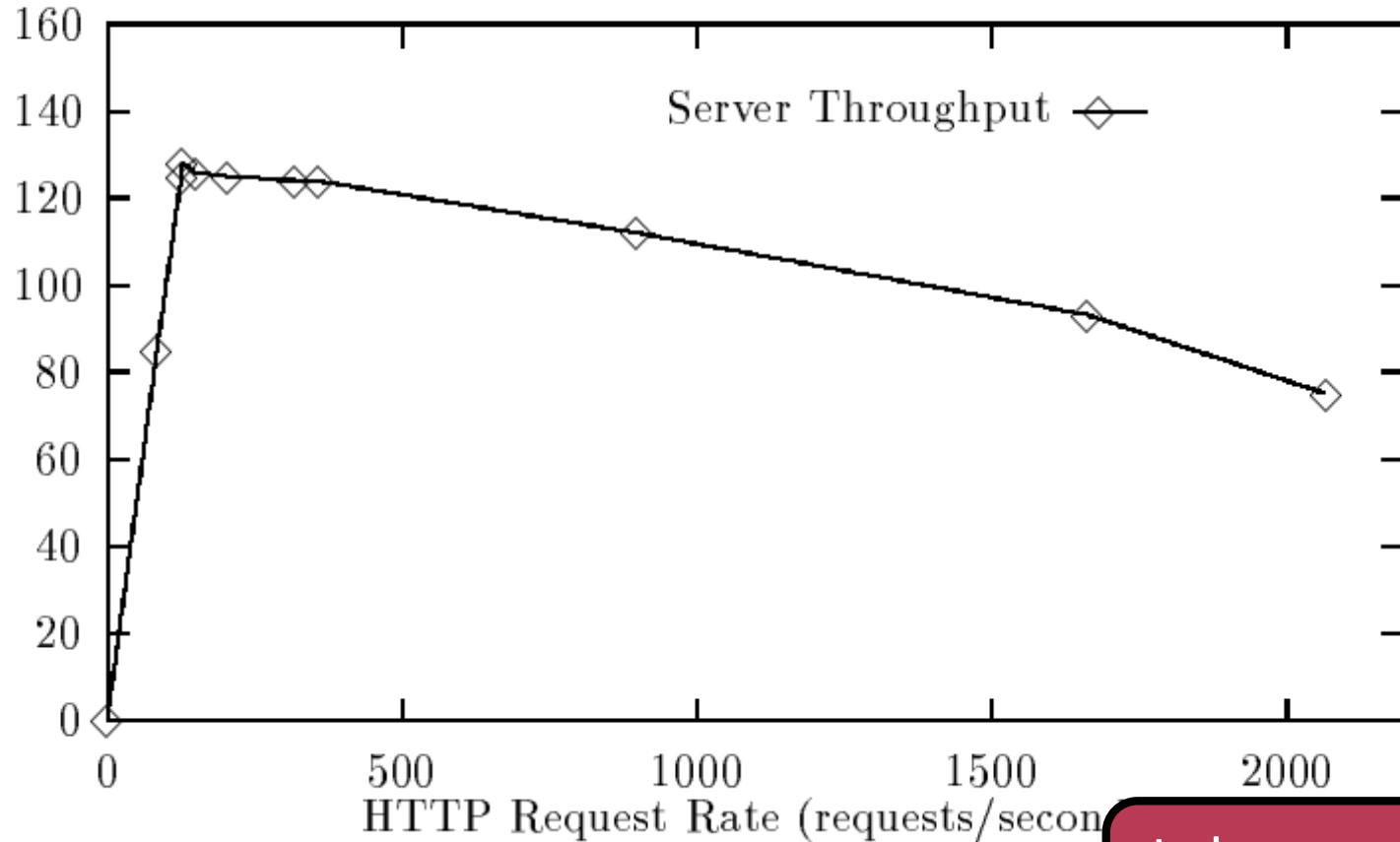
Jellemzően mikor *ne*?

- Konstans terhelés
- Örökölt (*legacy*) rendszerek
- Adatbiztonság, törvényi és szabályozói megfelelés

Igények és kapacitás (*demand and capacity*)

Véges kapacitású erőforrások

HTTP Server Throughput (connections/sec)



Lehet ez ellen
védekezni?

Ábra forrása: [2], p 10

Erőforrások skálázása

■ „Scale up”



■ „Scale out”

- Kiszolgálás párhuzamosíthatósága?
- „webscale” technológiák
- → Fürtözés és replikáció
- (Mongo DB Is Web Scale)



Kapacitástervezés

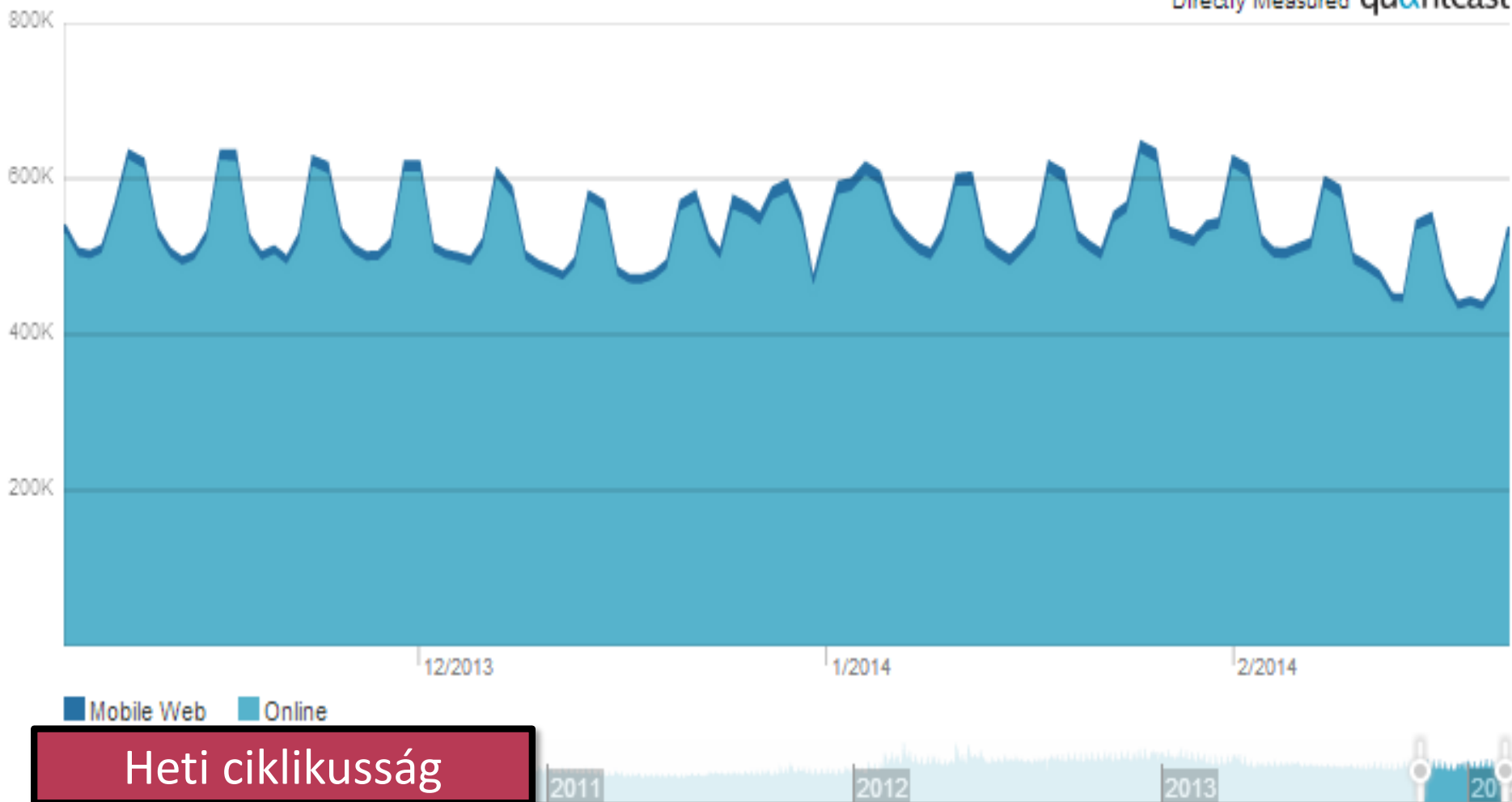
- Túl alacsony kapacitás: nem kiszolgált igény
 - Pl. átlagos terhelésre tervezés
- Túl nagy kapacitás: pazarlás(?)
 - Pl. csúcsterhelésre tervezés
- Helyi („*in-house*”) fizikai kapacitás ált. lassan és drágán növelhető...
 - ... és csökkenthető!
- (Egyszeri) befektetés és *elköteleződés*
- Igények?

Uniques (Global) per Day | Week | Month

Compare Site

More Options ▼

Directly Measured **quantcast**

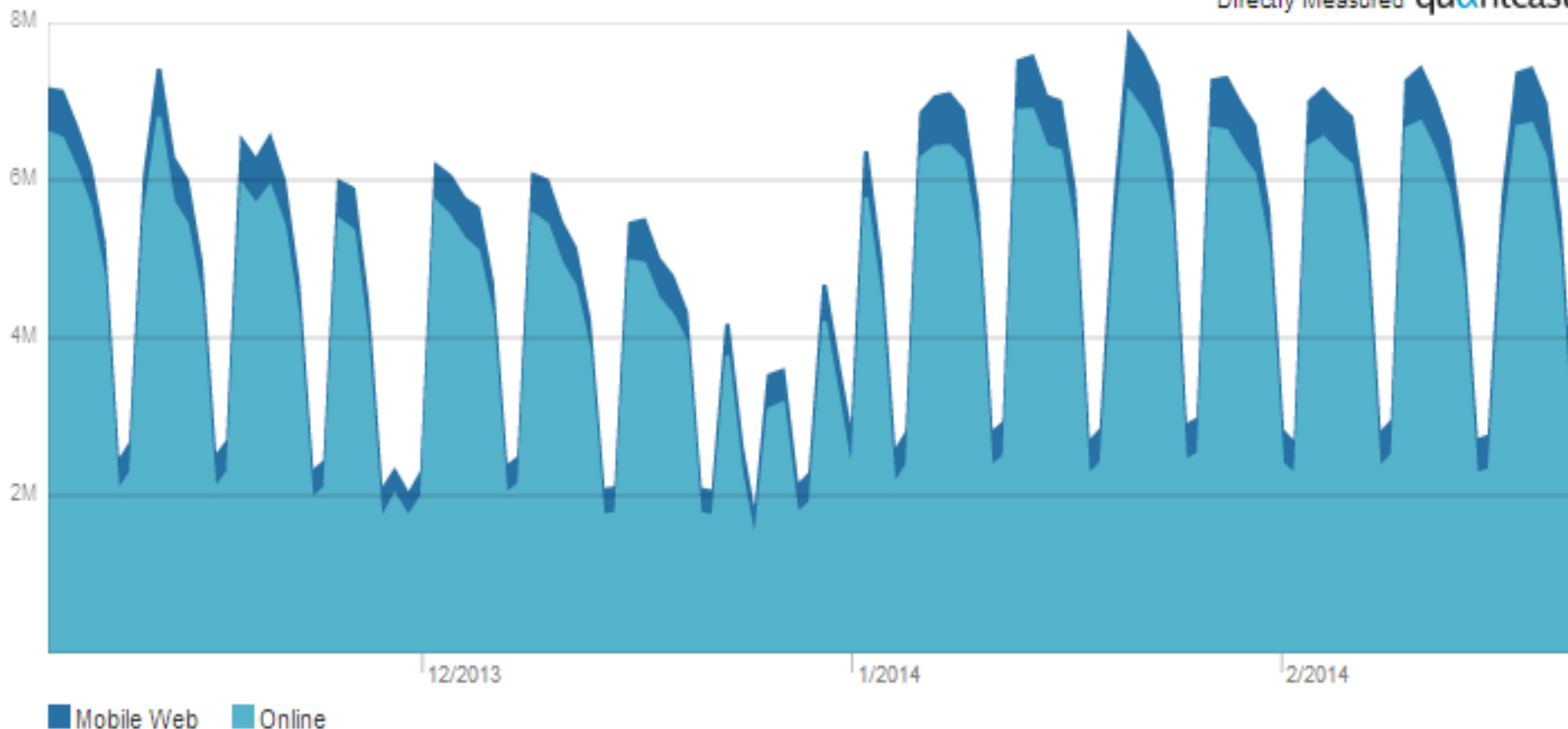


Uniques (United States) per Day | Week | Month

Compare Site

More Options ▾

Directly Measured quantcast



Heti ciklikusság + szezonálitás

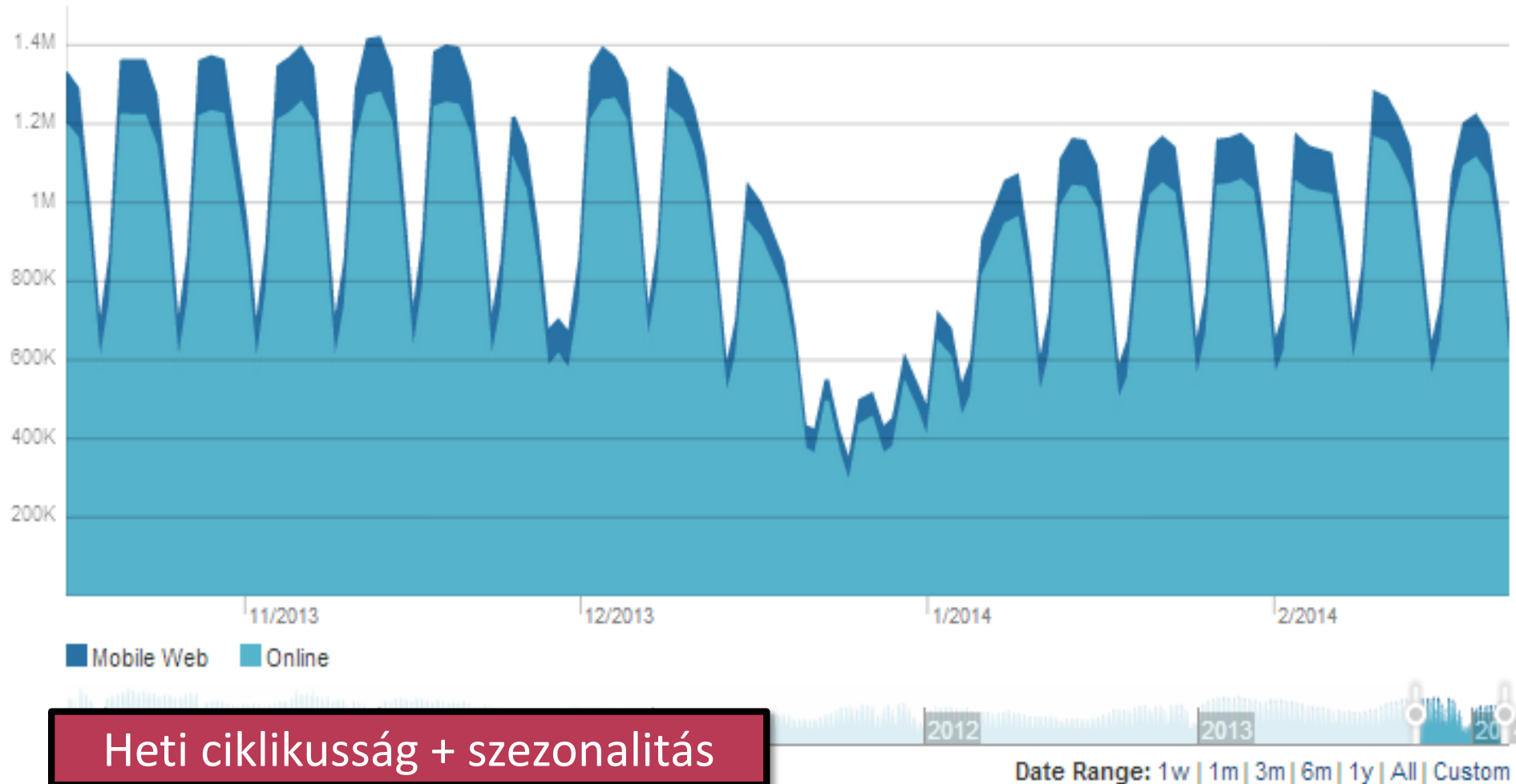
Date Range: 1w | 1m | 3m | 6m | 1y | All | Custom

Uniques (Global) per Day | Week | Month

Compare Site

More Options ▼

Directly Measured **quantcast**



Heti ciklikusság + szezonalitás

Uniques (United States) per Day | Week | Month

Compare Site

More Options ▼

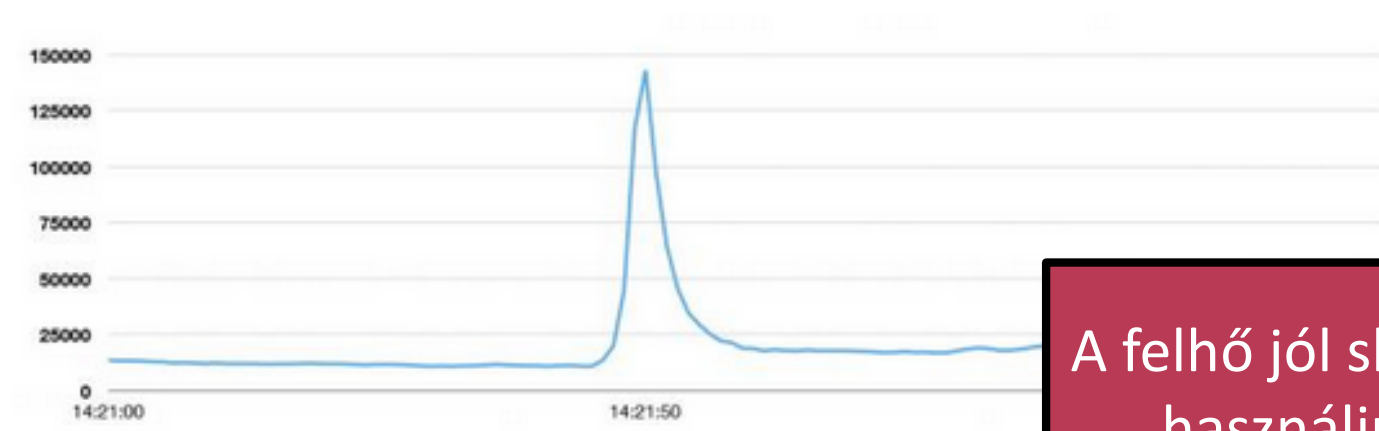
Directly Measured **quantcast**



...előrejelezhetetlen változások

Recently, something remarkable happened on Twitter: On Saturday, August 3 in Japan, people watched an airing of *Castle in the Sky*, and at one moment they took to Twitter so much that we hit a one-second peak of 143,199 Tweets per second. (August 2 at 7:21:50 PDT; August 3 at 11:21:50 JST)

To give you some context of how that compares to typical numbers, we normally take in more than 500 million Tweets a day which means about 5,700 Tweets a second, on average. This particular spike was around 25 times greater than our steady state.



A felhő jól skálázható;
használjuk azt!