facebook

Flash at Facebook

The Current State and Future Potential

Jason Taylor, PhD

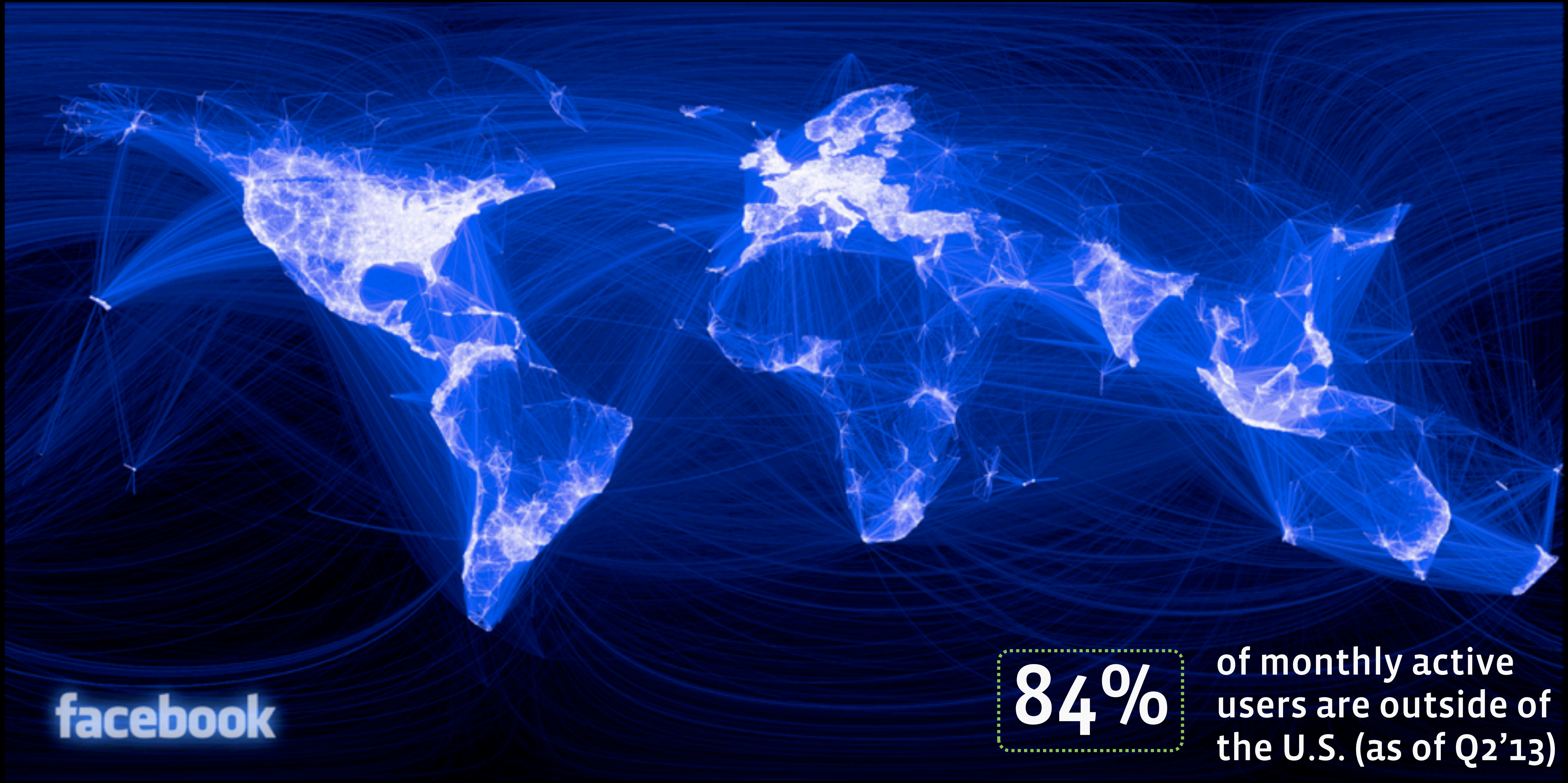
Director, Infrastructure

August 13th, 2013

Agenda

1	**Facebook Scale & Infrastructure**
2	Flash at Facebook
3	Cold Flash

Facebook Scale



84% of monthly active users are outside of the U.S. (as of Q2'13)

Data Centers in 5 regions.

Facebook Stats

- 1.15 billion users (6/2013)
- ~700 million people use facebook daily

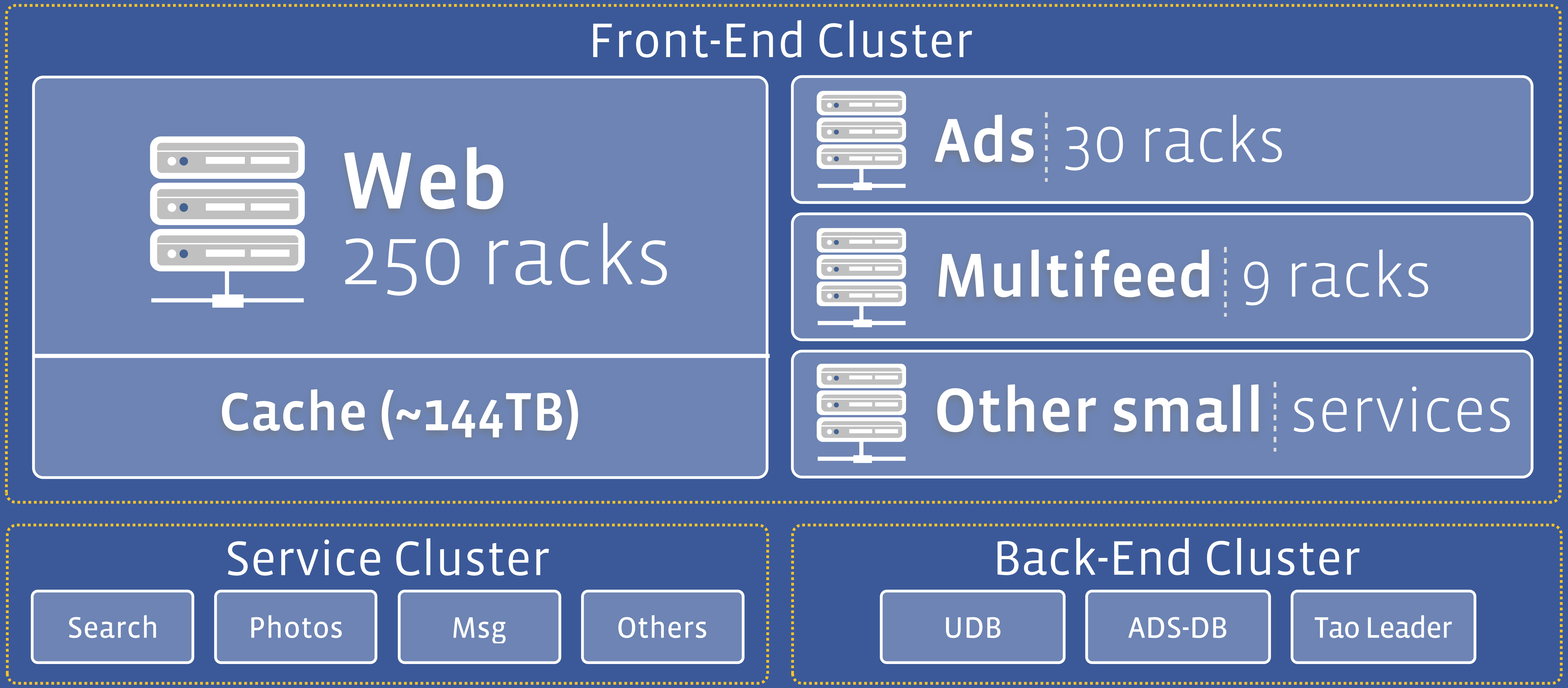
- 350+ million photos added per day (1/2013)
- 240+ billion photos

- 4.5 billion likes, posts and comments per day (5/2013)

Cost and Efficiency

- Infrastructure spend in 2012 (from our 10-K):
  - "…$1.24 billion for capital expenditures related to the purchase of servers, networking equipment, storage infrastructure, and the construction of data centers."
- Efficiency work has been a top priority for several years

Architecture



Front-End Cluster

Web 250 racks

Cache (~144TB)

Ads 30 racks

Multifeed 9 racks

Other small services

Service Cluster

Search

Photos

Msg

Others

Back-End Cluster

UDB

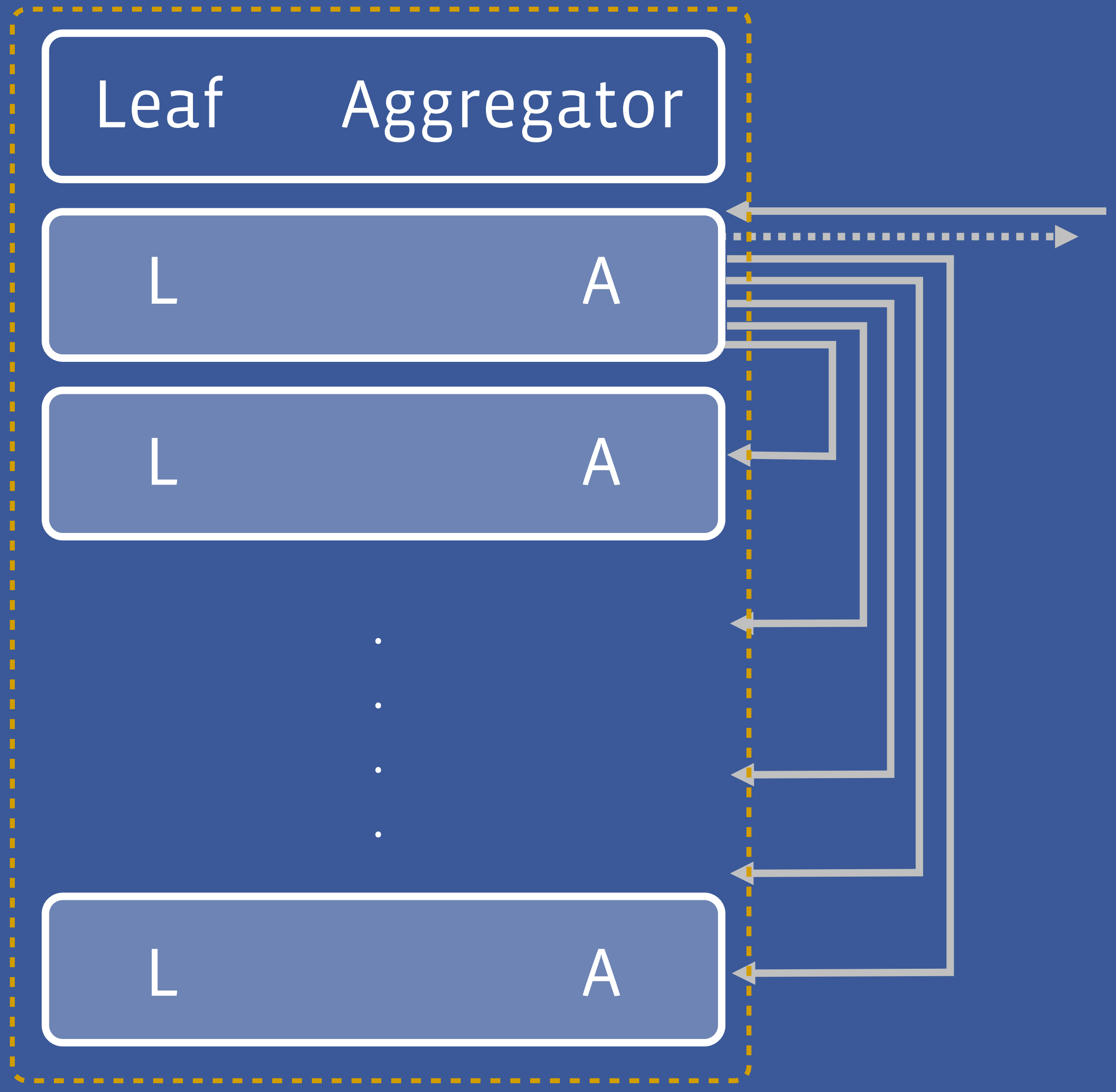
ADS-DB

Tao Leader

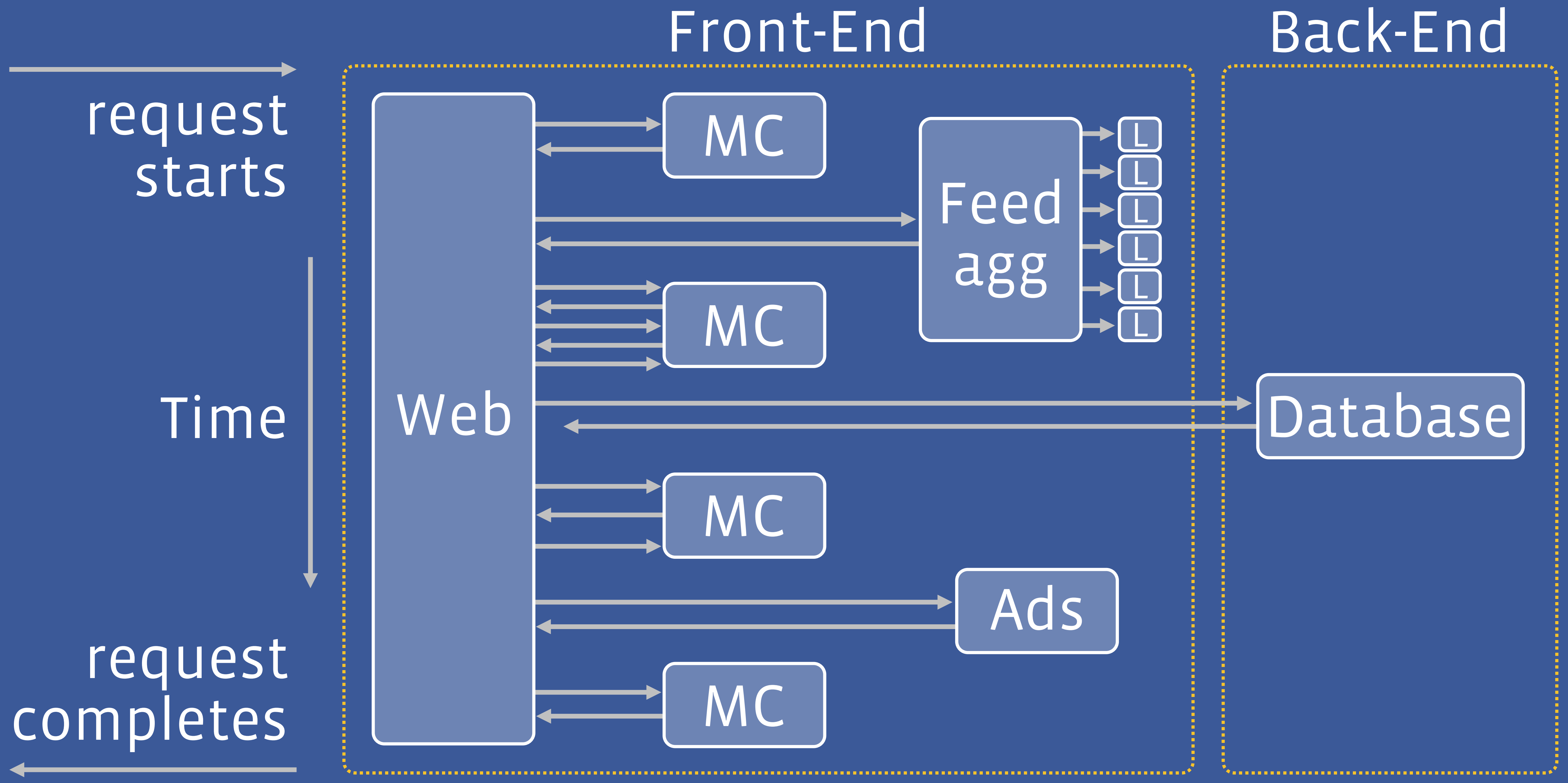
Lots of “vanity free” servers.

News Feed rack

- The rack is our unit of capacity
  - All 40 servers work together
- Leaf + agg code runs on all servers
  - Leaf has most of the RAM
  - Aggregator uses most of the CPU
- Lots of network BW within the rack



Life of a "hit"


Front-End
Back-End
request starts
Time
request completes
Web
MC
Feed agg
L
L
L
L
L
L
MC
Database
MC
Ads
MC

Five Standard Servers

Standard Systems	I Web	III Database	IV Hadoop	V Photos	VI Feed
CPU	High 2 x E5-2670	High 2 x E5-2660	High 2 x E5-2660	Low	High 2 x E5-2660
Memory	Low	High 144GB	Medium 64GB	Low	High 144GB
Disk	Low	High IOPS 3.2 TB Flash	High 15 x 4TB SATA	High 15 x 4TB SATA	Medium
Services	Web, Chat	Database	Hadoop (big data)	Photos, Video	Multifeed, Search, Ads

Five Server Types

Advantages:

- Volume pricing
- Re-purposing
- Easier operations - simpler repairs, drivers, DC headcount
- New servers allocated in hours rather than months

One Exception:

- PCIE Flash cards for Index services: News Feed, Search, etc.

Agenda

1	Facebook Scale & Infrastructure
2	**Flash at Facebook**
3	Cold Flash

Flash at Facebook

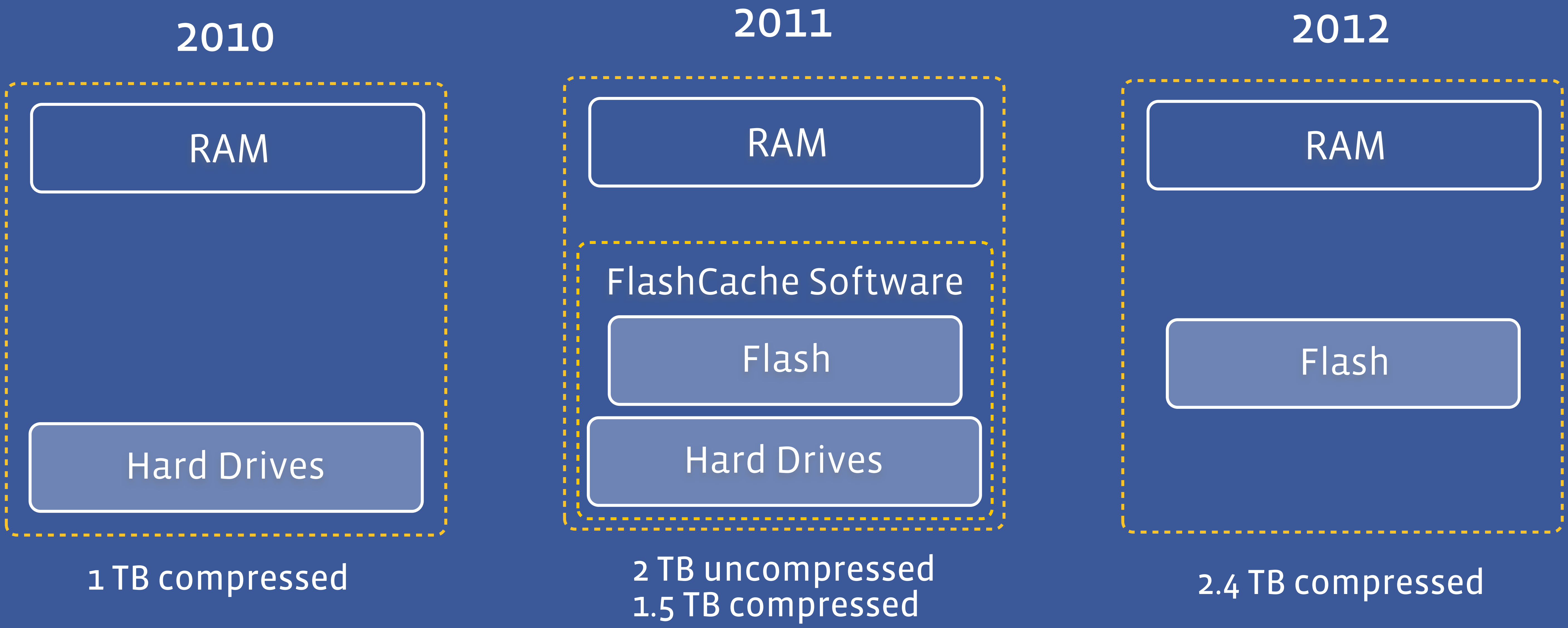
Databases

- FlashCache
- Flash as Disk

Index Servers

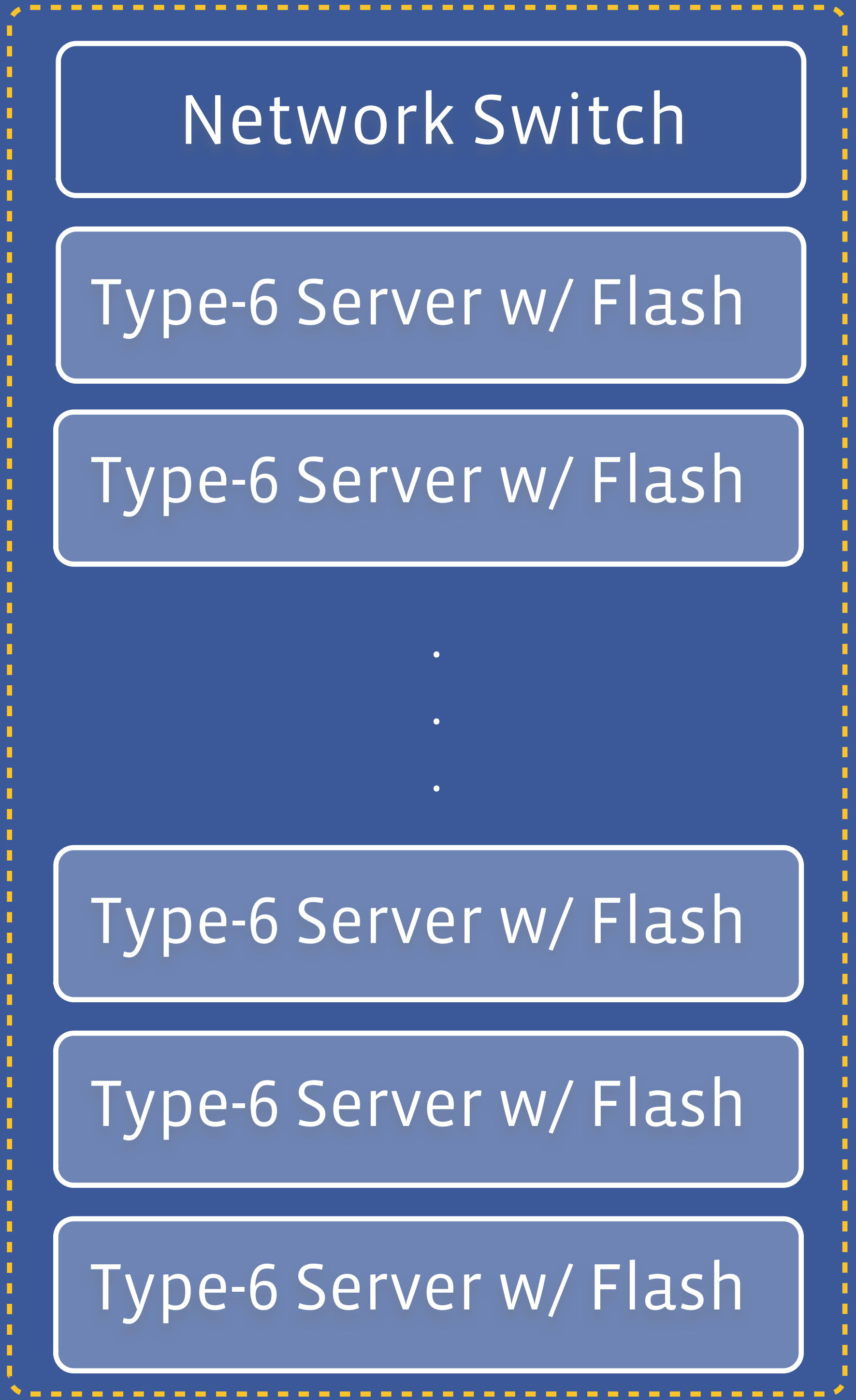
- Flash as a RAM ram replacement.

Flash in User Databases



Each iteration increased capacity and lowered latency outliers.

Flash in Index Servers


Network Switch

Type-6 Server w/ Flash

Type-6 Server w/ Flash

⋮

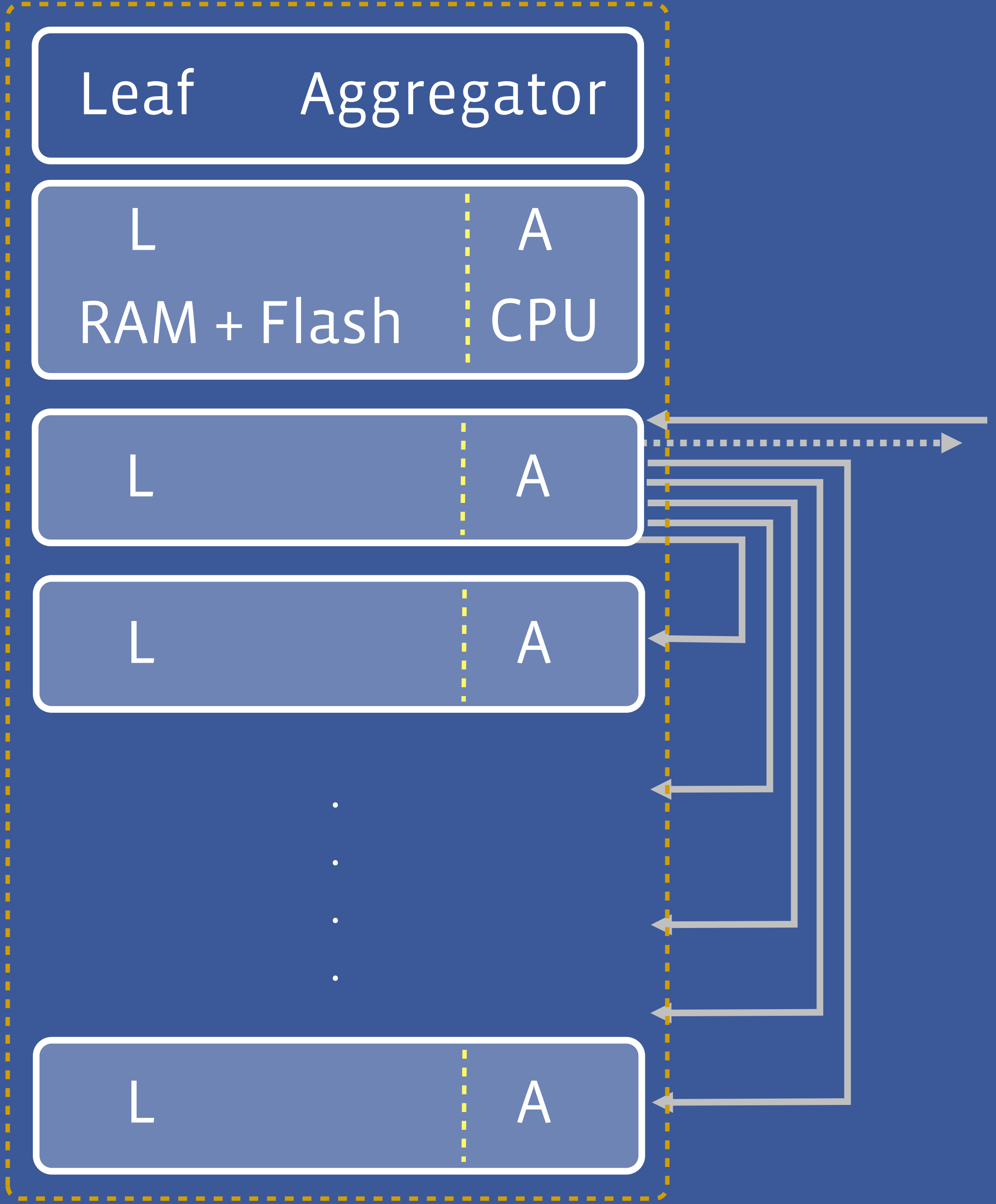
Type-6 Server w/ Flash

Type-6 Server w/ Flash

Type-6 Server w/ Flash

=>

80 processors	COMPUTE
5.8 TB	RAM
80 TB	STORAGE
30 TB	FLASH


Leaf Aggregator

L | A

RAM + Flash | CPU

L | A

L | A

⋮

L | A

For News Feed each leaf holds an index of a couple of weeks of stories.

Agenda

1	Facebook Scale & Infrastructure
2	Flash at Facebook
3	**Cold Flash**

Flash (solid state storage)

Today: Flash SSD drives are commonly used in databases and applications that need low-latency high-throughput storage.

The flash industry has focused on driving higher and higher write-endurance and performance. By looking in the opposite direction--low-endurance and poor-performance--a “cold flash” storage option is possible.

Next: Write Once Read Many (WORM) flash or an alternative solid state technology could provide high density storage at a reasonable cost.

Flash (solid state storage)

Knox rack of drives

- 1,920 TB of data
- 1.5 kW of power (cold storage)
- 2,500 lbs
- 0.78 Watts/TB

Rack of SSDs

- 3,990 TB of data
- 1.9 kW of power
- 2,050 lbs
- 0.47 Watts/TB

A focused effort in WORM solid state options should yield much higher densities & longer hardware lifetime at a reasonable cost.

Cold Flash*

The Facebook Ask?

Make the worst flash possible--just make it dense and cheap.

Long writes, low endurance and lower IOPS/TB are all ok.

* Other solid-state technologies may also work for “cold flash.”

Questions?

facebook