Scaling Up Quickly on the Cloud

EDMOND LAU
edmond@quora.com

5th Extremely Large Databases Conference
10.19.2011
“Quora is a continually improving collection of questions and answers created, edited, and organized by everyone who uses it.”
At what point does it make sense for a company to migrate off of cloud hosting (like EC2 or Rackspace) to their own datacenters?

When it saves money while not impacting reliability or quality of service.

EC2 and other dynamically provisioned virtualized instance services are great for "peak and valley" scenarios, product launches, time-limited computationally intensive compute jobs that require lots of CPU cores, etc. (Basically anything short-term or with an unpredictable component that drives resource utilization). They provide a lot of flexibility and significant cost savings in these scenarios over more traditional hosting/data-center deployments.

For predictable, steady-state traffic/CPU load use cases, these services are less ideal. If you're getting a definite amount of steady-state baseline traffic/server-load each month, EC2 may not be the most efficient usage of your hosting dollars.

Hybrid deployment scenarios (combining dedicated hosting/data-center resources with EC2-style cloud services) can offer the best of both worlds in many cases. This enables you to reduce cost basis via leveraging of dedicated servers/data-center resources for "steady-state" traffic, while retaining the peak load handling capability of EC2 and flexibility it provides to IT resource provisioning and deployment.

Other cases where virtualized cloud deployments don't make sense are cases where customized hardware is absolutely necessary — solid state storage arrays, FPGA cards, large amounts of local storage, etc. EC2 does offer machines with Nvidia GPUs as a hardware option for those engaging in HPC tasks, but hardware options in the cloud are still fairly limited.
The Team

• Founders include ex-CTO of Facebook and ex-head of Facebook’s platform team
• Small team of 30, based in Palo Alto
  • 14 engineers, 4 designers, 2 data analysts
• Founded mid-2009.
• Launched invite-only beta in Jan 2010, with full launch in June 2010.
How do we use the cloud?

- Entire product built on top of Amazon Web Services (AWS)
  - EC2, EBS, S3, Hive via Elastic MapReduce over S3
- Services: MySQL, memcached, search, home feed, web servers, dependency trackers, scribe, development boxes
- Data logging
  - Fixed-schema data stored in MySQL, horizontally partitioned by time.
  - Unstructured data and structured data with flexible schemas (json) logged to scribe, periodically uploaded to S3
  - Large blobs written directly to S3.
What things do we compute?

REAL-TIME
• trending topics & users to follow
• home page and topic feeds
• suggested users to ask to answer
• error aggregation

BATCHED CRON
• PeopleRank reputation scores
• internal dashboards/metrics (actives, signups)
• fake name detection
• answer classification
• weekly digest emails

• For many known and structured metrics, computation over MySQL with aggressive use of memcached is fast.
• Easy to leverage application-level logic in computation.
What ad-hoc queries do we compute?

- Sample analyses:
  - How often do users click the help link in the footer?
  - What % of users who clicked on a home page feed stayed on the site for at least 3 minutes?
  - What’s the average time spent on the site given a particular landing page?
- Run SQL-like Hive queries on both persistent and temporary AWS clusters
  - user-defined schemas and functions in Python and Scala to extract data
Why does it make sense to use the cloud?

- Quick to spin up machines to increase CPU capacity (~5 min)
  - Makes it easy to evaluate experimental services
  - Used for ad-hoc queries, computation of weekly digest emails for all users, accommodate traffic spikes
  - Provides flexibility to trade off engineering time to optimize software vs pay for more or faster machines
- For growth startups, AWS allows us to amortize cost of servers
  - Valuation of early-stage consumer companies are based more on growth and engagement trends and less on expenditure
- Optimize for product iteration and development speed
Additional Follow-up Links

• Why does Quora use MySQL as the data store instead of NoSQLs such as Cassandra, MongoDB, CouchDB etc?

• How will the April 2011 AWS outage impact the future of the cloud?

• Database Systems topic on Quora
  http://www.quora.com/Database-Systems