

Towards Understanding the Workload of a IaaS Cloud

Loïc Perennou

Outscale, ISEP

loic.perennou@outscale.com

September 13, 2018

Outline

Introduction

Data Collection

Comparison of Outscale's and Azure's Workloads

Relationship Between Tags and CPU Utilization

Outline

Introduction

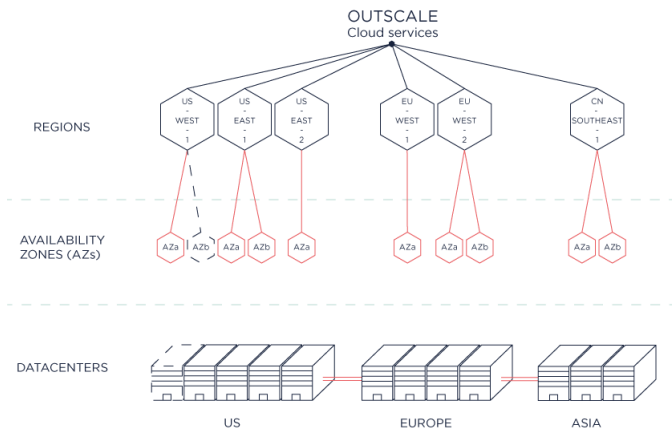
Data Collection

Comparison of Outscale's and Azure's Workloads

Relationship Between Tags and CPU Utilization

Outscale

- ▶ Founded in 2010, acquired by Dassault Systemes in 2017.
- ▶ Provides virtualized hardware like VMs, and services to manage them.
- ▶ Develops its own orchestrator, TINA OS, compatible with Amazon EC2.



Motivations

- ▶ We need to make resource allocation fit utilization.
- ▶ Utilization is unknown when a VM starts, but could be predicted by ML.
- ▶ Data must be available to propose and test models.

Related Cloud Workload Traces

Organization	Google	Eucalyptus Sys.	Bitbrains	Azure
year	2011	2014	2015	2017
# jobs/VMs	0.7M jobs	9,173 VMs	1,750 VMs	2M VMs
resource usage	no	no	yes	yes
starts/stops	yes	yes	no	yes
reference	[1, 2]	[5]	[3]	[4]

- ▶ Problem : We are not sure if Outscale's workload is similar to Azure's.

Outline

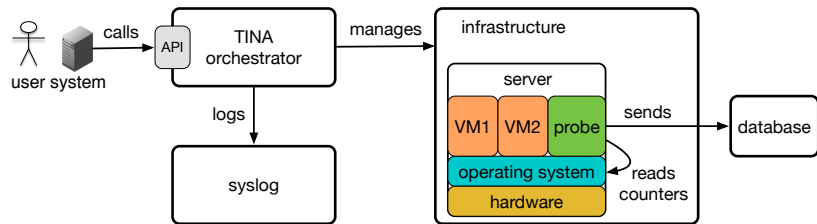
Introduction

Data Collection

Comparison of Outscale's and Azure's Workloads

Relationship Between Tags and CPU Utilization

Overview



2 data sources:

- ▶ Logs of user actions from TINA OS
- ▶ Measurements of hardware utilization of Virtual Machines

Descriptive Statistics

- ▶ 4 months
- ▶ 700 000 VMs in total
- ▶ 10 000 VMs running simultaneously

Outline

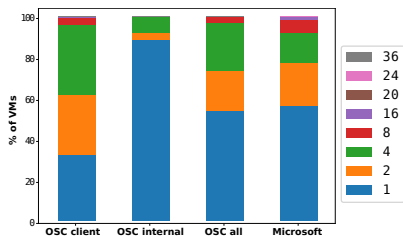
Introduction

Data Collection

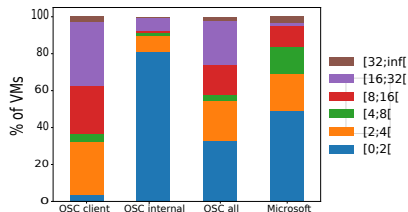
Comparison of Outscale's and Azure's Workloads

Relationship Between Tags and CPU Utilization

Distribution of Resources Requested by VMs



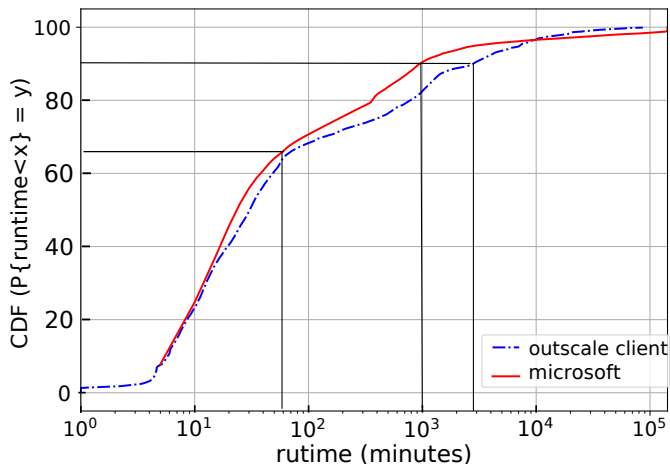
(a) cores requested



(b) ram requested

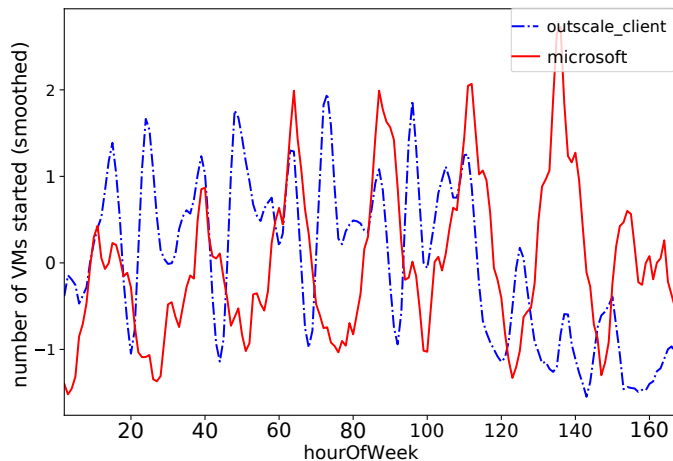
- ▶ Internal accounts at Outscale launch small VMs (test).
- ▶ Clients create bigger VMs than at Microsoft.

Distribution of Runtime



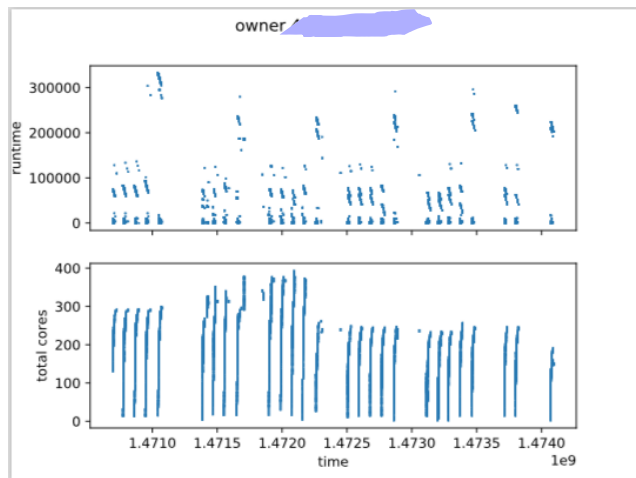
- ▶ The runtime of 65% VMs is \leq 1h.
- ▶ Clients create slightly longer VMs than at Microsoft.

VM Start Rate



- ▶ 2 peaks/day at Outscale, 1 at Microsoft.
- ▶ Less activity at Outscale in the weekend.

Relationship Between Start Time and Runtime



- ▶ Daily creation of VMs from Monday to Friday.
- ▶ VMs created on Friday run during the whole weekend.

Conclusion on Workload Comparison

- ▶ Bigger requests, longer runtimes at Outscale.
- ▶ Relatively more activity during the week, less in weekends.
- ▶ Activity patterns exists, at least for some users.

Outline

Introduction

Data Collection

Comparison of Outscale's and Azure's Workloads

Relationship Between Tags and CPU Utilization

Definition of Tags

Freely-typed string that describes a VM.

- ▶ Example (ideal): “Release 2.4 of Kafka used in production”.
- ▶ Example (real):
“EV6MTNDBLU_FUn3xIIATTiOAoDJYIeYGA
MT_Database2_0 420403n2q”.

Methodology

- ▶ Group VMs according to their tags (clustering).
- ▶ Visualize the CPU utilization of VMs within each cluster.

Convert Text Tags to Vectors for Clustering

VM 1 : "Kafka for test"

VM 2 : "Kafka for production"

VM 3 : "Visualization Engine"

	kafka	test	production	visualization	Engine
VM 1	YES	YES			
VM 2	YES		YES		
VM 3				YES	YES

Figure: Dictionary Vectorization

Hierarchical Clustering



- ▶ At the beginning, there is 1 group per vector.
- ▶ The two closest groups are merged (based on the distance between their elements).

Visualization of the CPU utilization of tag groups

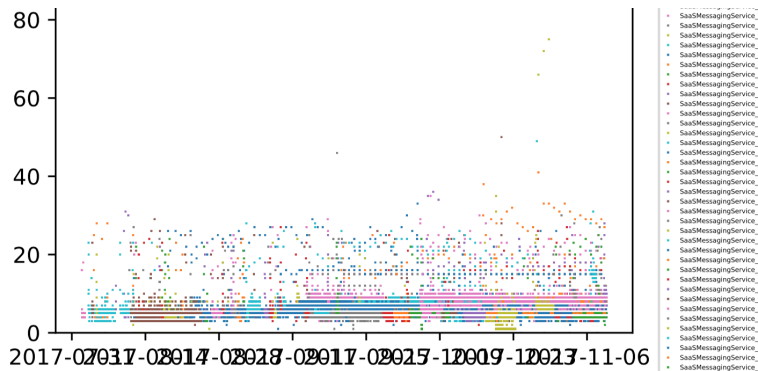


Figure: group A

Low utilization for every VM

Visualization of the CPU utilization of tag groups

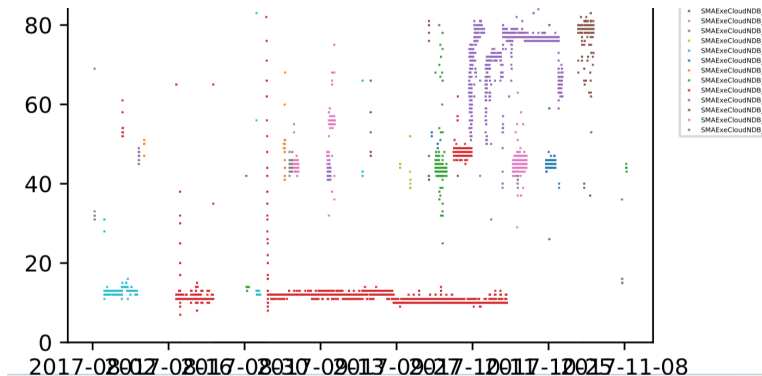



Figure: group B

Tags alone fail to explain the variance.



Conclusion

- ▶ Resource allocation of VMs needs to be based on predicted utilization.
- ▶ Predictive models need data to be trained and tested.
- ▶ Outscale's data is different from Azure's and justifies that we look for our own models.
- ▶ Tex information (tags) could provide interesting features (ongoing work).

References I

-  J. Wilkes, “More Google cluster data.” Google research blog, Nov. 2011.
Posted at <http://googleresearch.blogspot.com/2011/11/more-google-cluster-data.html>.
-  C. Reiss, A. Tumanov, G. R. Ganger, R. H. Katz, and M. A. Kozuch, “Heterogeneity and dynamicity of clouds at scale: Google trace analysis,” in *Proceedings of the Third ACM Symposium on Cloud Computing, SoCC '12*, (New York, NY, USA), pp. 7:1–7:13, ACM, 2012.
-  S. Shen, V. v. Beek, and A. Iosup, “Statistical characterization of business-critical workloads hosted in cloud datacenters,” in *2015 15th IEEE/ACM International Symposium on Cluster, Cloud and Grid Computing*, pp. 465–474, May 2015.

References II

-  E. Cortez, A. Bonde, A. Muzio, M. Russinovich, M. Fontoura, and R. Bianchini, “Resource central: Understanding and predicting workloads for improved resource management in large cloud platforms,” in *Proceedings of the 26th Symposium on Operating Systems Principles*, SOSP '17, (New York, NY, USA), pp. 153–167, ACM, 2017.
-  R. Wolski and J. Brevik, “Using parametric models to represent private cloud workloads,” *IEEE Transactions on Services Computing*, vol. 7, pp. 714–725, Oct 2014.