

## Problem: how can you study the cloud if you don't know what is being asked of the cloud? (i.e., what are users doing?)



## We will try to solve this problem with real production data.

### Azure SQL Database (ASD)



- A dominant cloud relational database service.
- Clusters in datacenters in 17 regions around the world.
- Initial data sets from N.A. and E.U.
- provisioning decisions.

Microsoft Azure SQL Database Telemetry. Willis Lang, Frank Bertsch, David J. DeWitt, Nigel Ellis. SoCC 2015

# **Get Your Cloud Data Here!**



Currently, this type of telemetry data is being used in Azure as input to load balancers, placement strategies, and for

• Researchers may be able to take advantage of this data for: multi-tenancy, cluster/server configuration, cloud user experience, and cloud modeling. (Academic collaborations currently underway studying multi-tenancy & modeling.)

• To find out more & request the data, please visit: http://gsl.azurewebsites.net/Projects/SQLAzure/AzureTelemetry.aspx





What we are providing in our data sets are back-end telemetry from the production ASD clusters, NOT queries.

	machine	db	timestamp	CPU time	Logical pages read	Logical pages written	•••	
Iviacnine 88	88	1	12:00-12:05	752	375	82	•••	
SQL instance	88	5	02:15-02:20	83	7	0	•••	
db 1	88	3	17:55-18:00	1458	98	3759	•••	
db 5	88	3	18:00-18:05	657	64	1857	•••	
db 3	88	9	00:30-00:35	2876	9928	2855	•••	
db 9	88	•••	•••	•••	•••	•••	•••	
CPU seconds/5 min interval over a month Logical pg reads/5 min interval over a month								
_ 10.0 -	ال مغلبين من ا			10,000 1,000				
	MMM	144	di				M.4	

0.1	Marine Marine Marine and a large selection		
<u> </u>	Hander fragenskiller fra hander og for an	han an a	10,000 1,000 9 100
ප 0.1			db2 pead 100 100 1



