

Static System Properties & Dynamic Optimization

HotTopiCS 2014 BreakOutSession – WrapUP

Topics 1 – Static Aspects

- Metrics for Scalability
 - Measure scalability aspects of networks, different topologies
 - Scalability of Networks
 - Best-Practice Topology: Fat-Trees
 - Limiting Factor: Infrastructure Provider
 - Robustness for different resource types / deployments
 - Dependencies between resource types, abstraction layer for resources
 - Weighting of different aspects
- Ways to implement Scalability – avoid bottlenecks
- Different Perspectives:
 - Cloud User vs. Cloud Provider
 - Business Perspective:
 - Costs (cost models are not always elastic (charge per hour), different granularity)
 - How much money is saved
 - Technical Perspective

Basics:

- Scalability is a prerequisite for elasticity
- Dimensions – Basic Resource Types:
 - Active Resources
 - Processing
 - Storage
 - Networking
 - Passive Resources:
 - Software / Service
 - Locks
- Bottlenecks can hinder scalability
 - Analysis of bottlenecks
 - Contention tracking

Perspectives

- Business Perspective

- Currently no SLOs for performance requirements in practice
 - No response time guarantees
 - No throughput guarantees

- Provider View

- Loss of customers is costly
- Reputation
- Market share

- Customer View

- Has its own customers -> Cloud Usage Pattern (CUP) formalism by SPEC RG Cloud
- How to quantify the impact of breaking SLAs?

Main Challenges for benchmarking clouds

- Repeatability & Consistency of Results:
 - Instances are located at different places
 - Other customers influence measurement
 - Result is valid one during measurement
 - E.g. benchmark run as DoS attack

→ Continuously monitoring collected metrics yields evolving interpretation

Topics 2 – Dynamic Aspects

- Appropriate Workloads
- Self-aware elasticity mechanisms
 - Updating auto-scaling rules
- Aspects
 - Latency
- Energy efficiency of cloud systems
- Implications of metrics
 - Best practices to tackle bad scalability/elasticity

Appropriate Workloads

- Goal: Assess the degree of determinism of system behavior
- Deterministic factors should be on the safe side
 - Provider don't want to violate SLOs
- Stress system with configurable mix of patterns
 - Patterns stress specific resources