

Implications of Service Level Agreements for Orchestration

Future Internet, Stockholm Nov 23 2009

Joe Butler
Intel IT R&D, Intel Labs Europe.

In the real world ..



Multiple service types hosted

- Communications
- Social networking, media, entertainment
- Enterprise

On a heterogeneous infrastructure :

- Data centres, grid,
- Wired, wireless,

With associated diversity of :

- Margin and profitability,
- Customer expectations and satisfaction requirements,
- SLA 'levels' – platinum ... bronze ... etc,
- Workload patterns and scheduling / provisioning requirements

Additionally ..



Known additional constraints

- Internal efficiency requirements –
 - ◇ Managed resource consumption
 - ◇ Utility cost minimisation
 - ◇ Operational cost minimisation
 - ◇ And third party resources and services
- Internal governance requirements
 - ◇ Logging for billing and auditability – potential complexity issues

And the unknown :

- Slashdot effect
- Resource failure
- Unforeseen SLA adjustments

Constraints expressed in Service Level Objectives

- 'Make composite resource available between 09:00 and 11:30.'
- 'Complete processing of a daily data transformation by 14:00.'
- 'Sustain average throughout of 3000 transactions/hour.'
- 'Peak throughput of 150 transactions/minute for up to 10 minutes.'

Statistical constraints

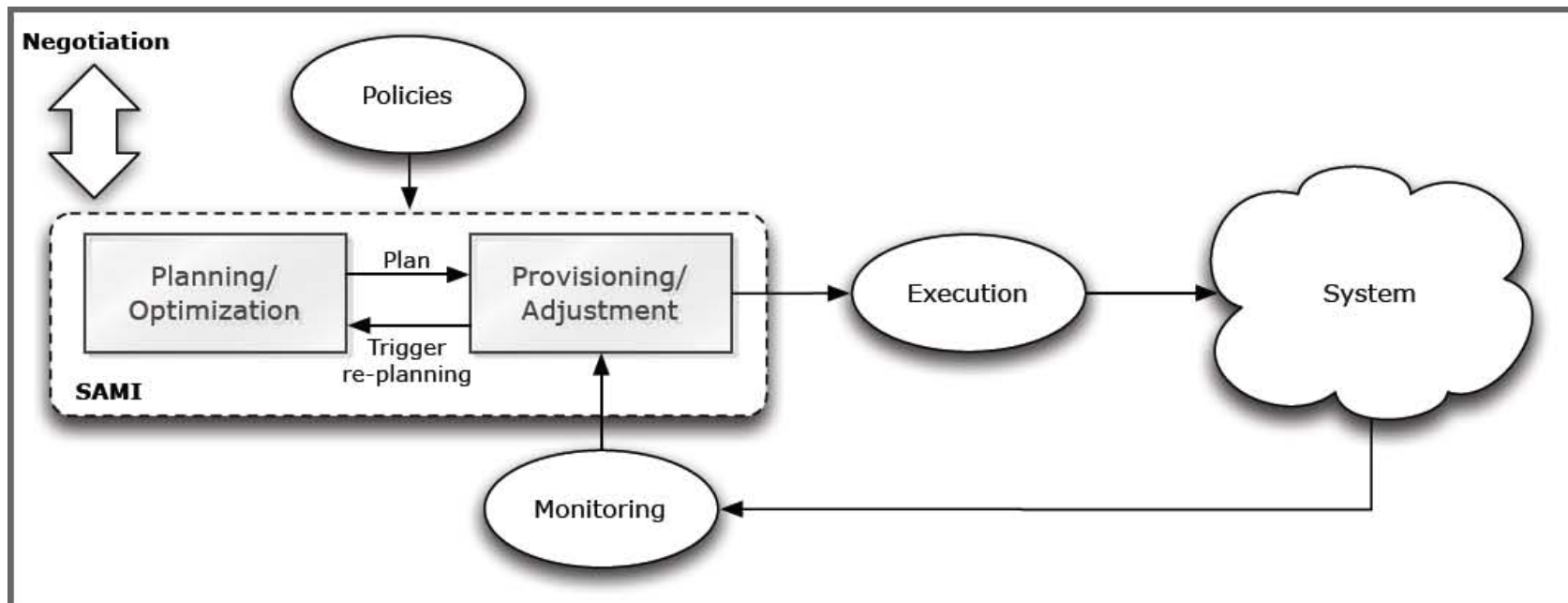
- '99.98% availability'
- '<1% transaction timeouts'
- '90% transactions complete within 1sec.'

Complex provisioning challenge

Simple view - provisioning

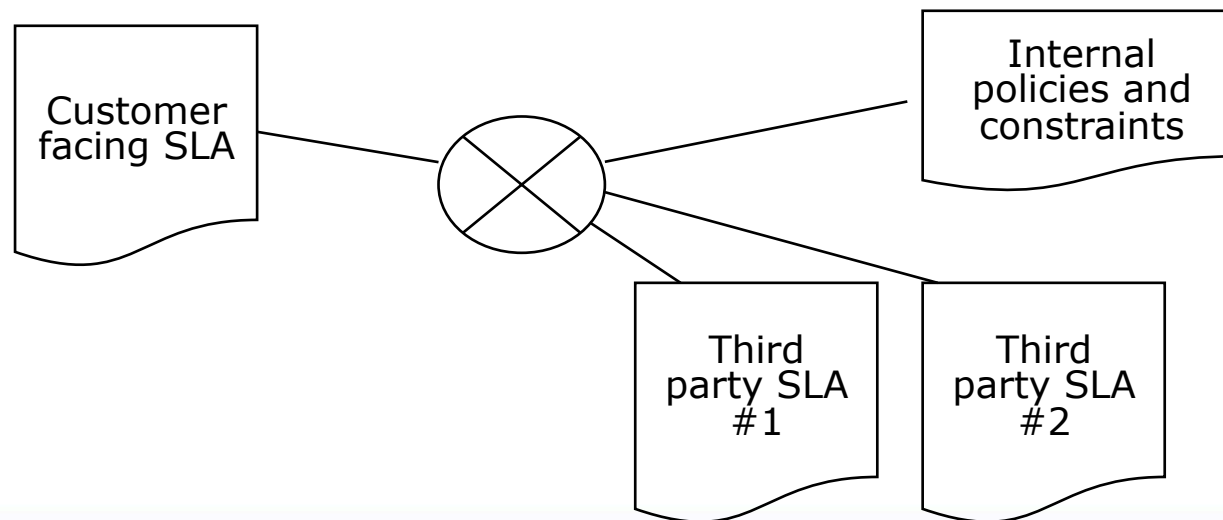
Provisioning / adjustment

- Normal lifecycle :
negotiation → planning/optimisation → provisioning → execution
- Monitoring provides a feedback loop for adjustment at run time.



Implications from service level objectives –

- Selection and configuration of individual resources based on performance and cost.
- Statistical likelihood of SLA compliance with end customer.
- Calculation and expression of SLOs, negotiation of SLAs with third party providers.
- Final orchestration and provision of service to end customer.



Proactive adjustment

- Normal lifecycle :
negotiation → planning/optimisation → provisioning → execution
- Monitoring provides a feedback loop for adjustment at run time.
- Prediction methods can infer an adjustment requirement based on monitored resource and workload trends.
- Adjustment –
 - ◇ Dynamic re-configuration of resource allocation to any part of the service chain.
 - ◇ Full re-provisioning of any part of the service chain followed by re-orchestration.

Reactive adjustment

- Recovery from failure.
- Adjustment –
 - ◇ Full re-provisioning of any part of the service chain followed by re-orchestration.

In summary



Negotiation and Planning

- Complexities introduced by variations in SLO parameters, and mapping these to contracted SLOs with consumed services.
- Statistical reasoning and optimisation problem.

Run time

- Workload patterns and variations, combined with policy objectives, may dictate a re-composition of consumed services.
- Possibly enabled through elasticity rules within these SLAs,
- SLA adjustment for these resources,
- or a re-negotiation with these or other services.

**Thank
you!**