

Service Level Agreement: what to know

Strategies for Success in the Network Economy



Natale “Nani” Prampolini
Senior Consultant



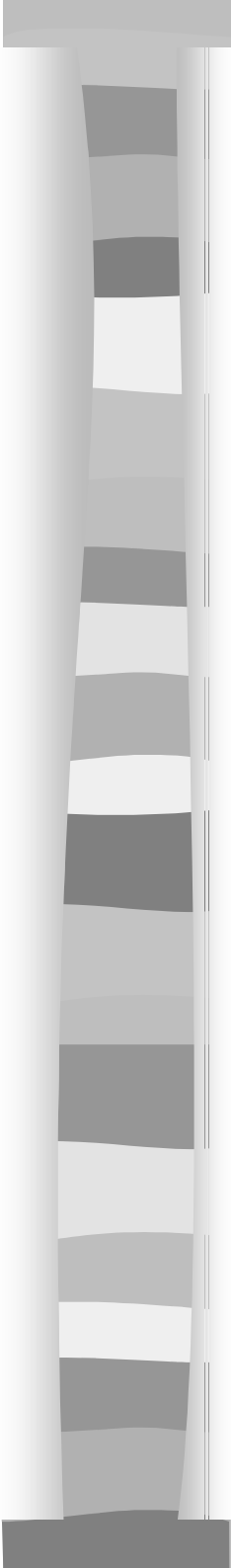
Agenda

- SLA definition and content
- A practical view
- Technical Requirement for SLA



Agenda

- SLA definition and content
- A practical view
- Technical Requirement for SLA



What are Business Services and Service Levels ?

- ✂ Business Service is the functionality (or service) the System/Application provides to do business.
- ✂ Service Level is the level at which the end-user or customer expects a business service to:
 - perform,
 - be available for use,
 - scale with growth, and
 - be secure from attack or mistakes



Contents of a SLA

The type and nature of service to be provided

The expected performance level of the service

The process for reporting problems with the service

r response and problem resolution

The process for monitoring and reporting the service level

The credits, charges, or other consequences for the service provider in not meeting its obligation

Escape clause and constraints, including the consequences if the customer does not meet his or her obligation



Business Requirement for SLA

Purpose of the SLA

Description of Service provided

Start date and duration of service

Payments terms

Penalties for non performance

Termination Conditions

Warranties, indemnities and limitation of liability

Account management contacts

Customer Service & Support

Response time to resolve unplanned outages

Projections of planned outages

Service upgrades and revisions

Reporting Procedures



SLA Chapters - Core

- Detailed description of services offered
- Implementation of Security: details
- Implementation of availability / scalability: details
- Provisioning of new apps, new users
- QoS standards and measurement in place
- Backup and Restore
- Support, escalation paths, committed resolution times
- User Reporting
- Customizations and change requests: the process
- Patches / Maintenance



SLA Chapters - Additional

- Levels of service
- Pricing and payment
- Penalties
- Privacy statement
- Content statement (no obscene or illegal, no viruses)
- Changes to contract (new apps, +uptime, diff. Backups)
- Legal items (disclaimers, litigations, courts, etc.)



Agenda

- SLA definition and content
- A practical view
- Technical Requirement for SLA



Service Level is...(practical perspective)

- Availability

- Can I get to applications and data?

- Can I always get access when I want it?

- Security

- Can I trust it?

- Capacity = Scalability

- Is there enough?

- How fast can I grow?

- Response = performance, provisioning, troubleshooting

- How slow/fast is it?

- How quickly I will have the service in place?

- How long it will take to fix it?

- Checking = reporting

- How can I check respect of contract clauses?

Availability

End_to_End availability =
 $\text{Availability_of}(\text{piece1}) * \dots * \text{Availability_of}(\text{pieceN})$

It is always less (or at best equal) than the availability of any piece

EXAMPLE

Network	99.90%	└─→	99.10 %
Datacenter, systems	99.70%		
Applications	99.50%		

Either commit on the single pieces, or provide an overall guarantee which is the combination of all.
No silver bullet.



Availability - Considerations

Per month or per year?

99.99 implies 50 min. a year, 4 min. a month.

If a month you have 6 min. Unavailability

monthly SLA is broken, yearly is still ok.

Monthly availability is 12 times harder than yearly one.

Per working hours or 24x7?

250 hours per month acceptable in some (SME) cases

15 min. downtime every night (2/2.15am) acceptable in many cases

Planned or unplanned?

PLANNED unavailability with 2 months notice is acceptable.

NUMBER OF FAULTS

How is the average split?

If same amount of data/time is lost at each unavailability

a single 4 min. stop is better than 240 1sec. ones.

Hint: Commit on *number of faults* as well.



Backup and Restore

Backup granularity to be defined

- Do I backup disks or mailboxes ?

Backup frequency

- Depends on SLA level

RESTORE:

- How long does it take?
- Granularity: Can I restore just the x subfolder?
- Can it be performed by the user, once the copy is on line?

Archiving / Media Storage



Security: Protecting customer's environment

Protection INSIDE the Datacenter

Protect from other hosted companies:

- Denial of Svc attacks

- Unauthorized access

Protect from ASP personnel

- Access

- Damage: much more critical!

- Hint: prove that only a minimum set of persons at the ASP is accessing sensitive data (and possibly name them).

- Issue a privacy statement

Protection of customer equipment from the Datacenter

- Issue and require from all customers a Written commitment

- Configure and offer CPE firewalls/filters



Scalability

Capacity Planning Policy

Monitoring the growth
Setting the alarms

How to scale

Flexibility of the architecture
Vertical vs Horizontal scalability



Problems detection / resolution

ACTION which activates problem resolution:

- Customer call (bad)
- Automatic alarm inside the DC

Better, proactive

Typical times to detect failure: 3-6 min.

Typical times to start resolution: 10-15 min.

- Troubleshooting times:

Different levels in SLA

Typical: 24/48 hours resolution

for major failures

TOOLS used to test
pinger machines.

OS alerts

Ad-hoc scripts

snmp agents + Enterprise Mgmt framework



Provisioning

Detail procedures and time required to add

Companies

Machines

Users

Applications

Include modifications: Change user X into
Y

and any other possible change.

Include removal (of applications especially)
Example: Remove Excel from all users desks.



Upgrade / Maintenance

Example: SP1 is issued.

Example: New major version of XXX is available

Should the ASP implement it:

- immediately

- on customer request

- within a max delay (typical: 2 months)

- depending on customers (mgmt nightmare)

All ways are viable, just select one.



Reporting

Which kind of report I have ?
real time, paper, etc

How can I trust the reports?
Certificate, digital signature



Agenda

- SLA definition and content
- A practical view
- Technical Requirement for SLA



Technical Requirement for SLA

Network availability

Application availability

Network throughput

Network bandwidth provisioned

Application performance

Redundancy

Service delivery platform & architecture

Network scalability

Application Scalability

Peering Arrangements

Security Architecture

Data availability

Application administration

Network delay

Monitoring



Categories of SLA Technical Requirement

Technical Requirement: Network availability

Category: Availability

What is the guaranteed network up-time, expressed as a percentage? (e.g. 99,5% uptime to the customer's Local Area Network)

Technical Requirement: Application availability

Category: Availability

What is the guaranteed application availability, expressed as a percentage of each component (e.g. database, directory, datastore) of the system as well as for the entire system?



Categories of SLA Technical Requirement

Technical Requirement: Redundancy

Category: Availability

How is redundancy provided? Which specific components are redundant? Are there any single point of failure in the application architecture?

Technical Requirement: Data availability

Category: Availability, Security

Where is data stored, how is it backed up and how quickly can be restored? How secure is the archiving of data? How are service configuration data stored and recovered? Is there a backup datacenter for catastrophic failure? Is all user, configuration and security information available in backup facilities?



Categories of SLA Technical Requirement

Technical Requirement: Network data loss

Category: Performance

What percentage of data loss is acceptable for the provisioned application? (Real-time interactive applications such as voice require less than 1% loss rates whereas Web-surfing is acceptable with packet loss rates approaching 5%)

Technical Requirement: Network delay

Category: Performance

What response time is required for the application and therefore how much network delay is acceptable? (Real-time interactive applications such as voice require response times of less than 100 milliseconds whereas Web-surfing is bearable with a response time of 250ms.)



Categories of SLA Technical Requirement

Technical Requirement: Application performance

Category: Performance

What is the capacity for the application in terms of transactions per second, keystroke response time, database response time, etc.? At what capacity does the application operate routinely and what capacity is guaranteed? If multiple customers reside on the same servers, how are system resources allocated to specific customers?

Technical Requirement: Network throughput

Category: Performance, Scalability

What is the capacity of the network? How many other customers reside on the same network? At what percentage capacity does the network operate routinely?



Categories of SLA Technical Requirement

Technical Requirement: Network bandwidth provisioned

Category: Performance , Scalability

For what bandwidth is the customer subscribing? How granular is the allocation per user, per service? How is bandwidth allocated and managed in the system?

Technical Requirement: Network Scalability

Category: Scalability

What is the network's reach? How is the network extended to reach new locations such as a new customer site or a new branch office location? How does the xSP match server capacity to customer traffic patterns?



Categories of SLA Technical Requirement

Technical Requirement: Application Scalability

Category: Scalability

As more users are added, can the applications architecture Scale without degrading the service level below agreed-upon levels? Can you define specific system upgrades that will allow the number of users or application performance to scale to specific levels?

Technical Requirement: Peering arrangements

Category: Scalability

What happens if the customers is located outside the network's footprints?



Categories of SLA Technical Requirement

Technical Requirement: Security Architecture

Category: Security

What specific security tools and procedures are in place? Are there back doors to the network or applications architecture that make it vulnerable to hackers?

Technical Requirement: Application administration

Category: Security

Are there capabilities for customers to self-manage their users and customized application elements? How long does it take for changes to propagate through the production system?



Categories of SLA Technical Requirement

Technical Requirement: Monitoring

Category: Availability, Performance, Scalability, Security

How is the application and system monitored?
What specific software tools will be used for monitoring? What metric can be viewed in real time and what visibility is offered to the customer in real time?

Technical Requirement: Service delivery platform and architecture

Category: Availability, Performance, Scalability, Security

What specific equipment is being used for both the network and the computer hardware?
How is the network designed? How is the storage provided?



Questions?

- Natale “Nani” Prampolini
- Senior Consultant
- nprampo@tin.it