

Execució de proves de rendiment sobre sistemes d'informació de Salut

**Àrea d'Arquitectura de Solucions i
Qualitat del Departament de Salut**

Març de 2025



**Generalitat
de Catalunya**

Direcció o coordinació: Joan Esteve Riasol, Josep Antoni Mira Palacios

Autors o redactors: Àrea d'Arquitectura de Solucions i Qualitat del Departament de Salut

Alguns drets reservats

© 2025, Generalitat de Catalunya. Departament de Salut.



Els continguts d'aquesta obra estan subjectes a una llicència de Reconeixement-NoComercial-SenseObresDerivades 4.0 Internacional.

La llicència es pot consultar a la pàgina web de Creative Commons.

Unitat promotora:

Àrea d'Arquitectura de Solucions i Qualitat de Salut

Versió 1.0.1, 1a edició:

Barcelona, març de 2025.

DOI:

10.62727/DSalut/13028

Assessorament lingüístic:

Servei de Planificació Lingüística del Departament de Salut.

Disseny de plantilla accessible 1.09.

Oficina de Comunicació. Identitat Corporativa.

Sumari

1	Introducció	4
1.1	Propòsit del document	4
1.2	Motivació de l'informe de proves de rendiment	4
2	Característiques desitjables de les proves	5
2.1	Monitoratge i mesura	5
3	Requeriments no funcionals per les aplicacions	7
3.1	Requeriments relacionats amb la gestió de la capacitat	7
3.2	Requeriments relacionats amb el temps de resposta	8
4	Especificació de les proves de rendiment a Salut	10
4.1	Introducció	10
4.2	Prova de línia base	10
4.3	Prova d'estabilitat	10
4.4	Prova de càrrega	11
4.5	Prova d'estrès	11
5	Presentació de l'informe amb els resultats de les proves	13
5.1	Introducció	13
5.2	Definició del cas d'ús	13
5.3	Execució de les proves	14
5.4	Conclusions	15
6	Annexos	16
6.1	Annex 1. Procediment de valoració de les proves de rendiment	16
6.2	Annex 2. Responsabilitats dels equips implicats	16
6.3	Annex 3. Contingut de la <i>checklist</i>	18

1 Introducció

1.1 Propòsit del document

El propòsit d'aquest document és ressaltar la importància de les proves de rendiment dins el cicle de vida dels sistemes informàtics, i explicar les característiques essencials i tipus de proves requerides al Departament de Salut.

Està dirigit tant als proveïdors d'aplicacions (encarregats de la construcció del sistema segons les característiques d'arquitectura demanades), com als gestors d'aquestes solucions (que han de garantir-ne la implantació amb èxit a producció), proporcionant una guia clara i comprensible sobre com dissenyar, executar i revisar aquestes proves.

El CTTI disposa d'informació referent als requeriments d'execució de proves de rendiment i a tot el seu cicle de vida, recollides en el següent enllaç:

https://qualitat.solucions.gencat.cat/procediments/realitzar_proves/

Aquest document detalla i amplia aquests continguts, adaptats a la importància dels sistemes d'informació del Departament de Salut.

1.2 Motivació de l'informe de proves de rendiment

Les proves de rendiment són crucials per assegurar que un sistema informàtic funcioni de manera eficient i fiable sota diferents condicions de càrrega ja que permeten identificar i corregir colls d'ampolla o punts de fallida únics, validar l'escalabilitat dels components, i assegurar l'estabilitat i resiliència del sistema. Tot això es tradueix en una millor experiència per a l'usuari final i una optimització dels recursos i costos totals de l'aplicació.

Un document de proves de rendiment ben elaborat i estructurat facilita la col·laboració entre els diferents equips tècnics i de negoci: desenvolupadors, administradors de sistemes, grups de suport funcional o tècnic, i altres parts interessades. Un dels principals beneficis és la capacitat de detectar i solucionar problemes de rendiment de manera proactiva, disminuint el risc que afecti els usuaris finals.

Així mateix, un document clar i concís permet una millor planificació i execució de les millores necessàries, contribuir a un sistema més saludable i robust, i també objectivar possibles carències en sistemes transversals fora de l'abast de les aplicacions.

La definició, preparació, i execució de les proves de rendiment, i també la interpretació i anàlisi de resultats d'aquestes, han de complir els nivells d'exigència necessaris per garantir la correcta implantació del sistema a entorns productius.

2 Característiques desitjables de les proves

Reproductibilitat

Les proves de rendiment han de produir resultats consistents quan s'executen sota les mateixes condicions. Això assegura que els problemes detectats siguin reals i no fluctuacions aleatòries, i permet una identificació precisa i efectiva de possibles afectacions en el rendiment sota una determinada càrrega d'operacions.

Representativitat

Les proves han de reflectir escenaris d'ús realistes, per tant, les condicions han de ser representatives del comportament dels usuaris en l'entorn de producció. Per exemple, si l'aplicació s'utilitza majoritàriament durant les hores de feina, les proves han de simular aquesta càrrega. Això assegura que els resultats siguin rellevants i aplicables a l'ús real del sistema.

Automatització

Automatitzar les proves de rendiment permet realitzar-les de manera eficient i repetitiva, sense intervenció manual, assegurant consistència i estalviant temps. Les eines d'automatització ajuden a executar una gran quantitat de proves de forma ràpida i faciliten la detecció primerenca de problemes. Per exemple, un script automatitzat pot simular centenars d'usuaris realitzant transaccions simultàniament.

L'eina per executar les proves de rendiment fixada per CTTI és **Load Runner**, amb els corresponents injectors de càrrega que simulen els usuaris virtuals (*VUser*):

<https://qualitat.solucions.gencat.cat/eines/alm-loadrunner/>

2.1 Monitoratge i mesura

Durant les proves de rendiment és essencial monitorar i mesurar diferents mètriques de l'aplicació, com són el consum de CPU i la memòria de la infraestructura que hostatja els diferents components, el nombre de crides, els temps de resposta, les taxes d'error, el nombre d'usuaris concurrents, el tràfic de xarxa, etc.

Aquest monitoratge permet analitzar el comportament del sistema sota càrrega i detectar possibles riscos de fallida, colls d'ampolla en forma de saturació de recursos o lentitud en el temps de resposta, o també oportunitats d'optimització.

És igualment important disposar de mètriques dels temps de resposta de les interaccions amb serveis externs o altres sistemes de suport de tercers com bases de dades. Aquestes mètriques proporcionen una visió completa del rendiment del sistema, ja que les interaccions amb components externs sovint tenen un impacte significatiu en el rendiment global de l'aplicació.

És responsabilitat del lot d'aplicacions garantir la generació i la disponibilitat d'aquestes mètriques durant l'execució de les proves de rendiment, així com la seva recopilació i interpretació a l'hora d'avaluar-ne els resultats.

En els casos en què el propi lot gestioni també components o infraestructures, com contenidors o serveis desplegats en el núvol públic, haurà de proveir directament aquestes dades. En canvi, per mètriques relatives a infraestructures *on-premise* o *cloud privat* administrades per altres lots (com el CPD) o a components transversals monitoritzats per sistemes comuns (com Talaia), el lot d'aplicacions haurà de planificar conjuntament amb els responsables corresponents la seva recopilació o obtenir-les de les fonts externes que corresponguin.

3 Requeriments no funcionals per les aplicacions

3.1 Requeriments relacionats amb la gestió de la capacitat

Entenem per capacitat el consum de l'aplicació sobre diversos aspectes. Com a norma general, es té en compte la capacitat màxima utilitzada per l'aplicació en relació amb la RAM, l'emmagatzematge persistent (si aplica) i l'ús de CPU.

En previsió d'anomalies o fallides derivades d'una saturació del sistema, s'estableixen els següents llindars màxims per cadascuna d'aquestes mètriques, segons instal·lació, per les quals caldrà donar una justificació de quina ha estat la causa del comportament observat durant la prova de rendiment:

Instal·lacions tipus PaaS, IaaS i Hosting:

RAM: 65% CPU: 65% Emmagatzematge: 65%

Instal·lacions basades en contenidors:

RAM: 85% CPU: 85% Emmagatzematge: 85%

Cal destacar que parlem de valors màxims o pics, no de valors mitjans. Per exemple, si un servei d'una aplicació basada en contenidors té un consum pic de CPU del 90%, encara que la mitjana sigui del 40%, no complirà amb el llindar i s'haurà de justificar i investigar la causa d'aquest pic. Establint si es consideren necessàries millores per evitar-ho o accions posteriors a tenir en compte per a futurs evolutius.

És important també tenir en compte:

- L'entorn de PRE on s'executen les proves hauria de tenir una arquitectura idèntica a l'entorn de PRO (elasticitat, autoescalabilitat, alta disponibilitat...) però amb una capacitat d'escalabilitat vertical i horitzontal que pot ser menor.
- En execucions d'escenaris asíncrons, el servei sempre ha d'ajustar la configuració del seu *throughput* de producció o lectura de missatges al tòpic, per adaptar-se a la capacitat màxima que no permeti saturar els seus recursos. Per exemple, un sistema produeix X missatges per segon en un tòpic, i el sistema consumidor els va llegint a un ritme de $Y < X$ per segon per evitar el desbordament de la seva memòria.

Per últim, en el cas de components adquirits en la modalitat gestionada com a servei (SaaS), cal recopilar també les mètriques de rendiment de què disposi el producte durant les proves.

D'aquesta manera, es pretén treure profit de les opcions d'elasticitat i escalabilitat automàtica que ofereixen aquestes modalitats d'aprovisionament per optimitzar el rendiment del component davant pics puntuals d'una prova d'estrès.

3.2 Requeriments relacionats amb el temps de resposta

S'entén que el temps de resposta mesurat comprèn:

- Interoperabilitat **síncrona**: des de la invocació del client a un servei de l'API de negoci fins la recepció de la resposta esperada o error per part d'aquest.
- Interoperabilitat **asíncrona**: des de la consumició d'un missatge d'un tòpic fins la finalització de les operacions relacionades en el tractament d'aquest.

Per avaluar els temps de resposta resultants d'una prova, s'han de prendre com a referència els llindars que haurien d'estar definits en el document de Descripció d'Arquitectura (DA) de la solució, concretament, en el capítol "Rendiment i escalabilitat".

https://canigo.ctti.gencat.cat/arquitectura/ajuda_da/#RendimentEscalabilitat

Tanmateix, com a norma general en els casos que aquests llindars no estiguin especificats explícitament en el DA, s'estableixen aquests requeriments respecte operacions sobre el sistema transaccional de l'aplicació:

Operacions d'alta, baixa i modificació de registres únics

Màxim acceptat de **60ms (mil·lisegons)**, dins d'un percentil 95.

Operacions de consulta d'un registre concret per identificador o clau primària

Màxim acceptat de **40ms (mil·lisegons)**, dins d'un percentil 95.

Agregació d'operacions en un mateix cas d'ús

Per calcular la durada prevista d'una interacció que refereix diferents accessos únics, s'ha de mesurar el temps des de la capa més exterior, per exemple: del frontal que invoca una crida REST, o des de la recepció d'un missatge per part d'un consumidor d'un tòpic.

Com exemple, es pot calcular la durada estimada d'una reserva de cita a un metge com l'agregació dels serveis:

- operació síncrona de cerca de pacient per identificador (40ms)
- operació síncrona de cerca de metge per identificador (40ms)
- operació síncrona de consulta d'un despatx per disponibilitat (80ms)
- operació síncrona de registre d'una cita nova sobre BD transaccional (60ms)
- publicació asíncrona del corresponent esdeveniment d'auditoria (10ms)

Donaria un temps de resposta estimat de 230ms en aquest cas d'ús.

Latència en el temps de resposta en núvol públic

Tant si treballem en una arquitectura purament en núvol públic, com una arquitectura híbrida amb recursos aprovisionats en *on-premise* o privat i d'altres en públic, cal tenir en compte que per la part de núvol públic treballem en diferents regions sobre el territori europeu, i és un requeriment indispensable identificar i mesurar la latència de xarxa entre l'origen i el destí, a l'hora d'interpretar els resultats de les proves de rendiment i justificar els temps de resposta obtinguts.

Per exemple, en una prova de rendiment, l'origen seria l'injector de Load Runner que genera els *VUsers* per començar a disparar operacions contra el sistema a provar.

Aquesta mesura és especialment rellevant atesa la localització actual de les zones de desplegament del CTTI en els diferents proveïdors de núvol públic:

- AWS a Irlanda
- Azure a Alemanya o Holanda
- Google Cloud a Alemanya

Cada un d'aquests proveïdors disposa d'una *Landing Zone* de *networking* gestionada per NUS en l'àmbit de núvol públic, que es coneix com a **NET0**.

Per cada regió, tenim una infraestructura de línia dedicada i privada de comunicacions dins la NET0, per connectar el núvol amb els diferents recursos *on-premise* dels CPDs de CTTI:

- DirectConnect per AWS.
- ExpressRoute per Azure.
- *InterConnect* per Google Cloud.

És responsabilitat del lot d'aplicacions entendre aquesta arquitectura de xarxa, en solucions que requereixen interoperabilitat amb components *on-premise* i implementar els fluxos de comunicacions de la manera més òptima possible al respecte. Per exemple: resoldre a nivell DNS dominis que són *.intranet.gencat.cat dins components desplegats al núvol, tramitar les corresponents regles de *Firewall* entre comunicacions *on-premise* i núvol, o gestionar correctament característiques com el *keep-alive* de les peticions HTTP entre recursos locals i recursos al núvol.

La següent guia de CTTI dona informació sobre els diferents fluxos de comunicacions de la NET0 de NUS, segons si són peticions de domini intranet, internet o ambdues:

https://canigo.ctti.gencat.cat/arquitectura/ajuda_da/#FluxosComunicacions

4 Especificació de les proves de rendiment a Salut

4.1 Introducció

Existeixen diverses tipologies de proves de rendiment amb objectius específics: proves de línia base, proves d'estrès, proves de càrrega i proves d'estabilitat, entre d'altres.

Cada tipus de prova es centra a analitzar aspectes diferents del rendiment de l'aplicació, com la capacitat de resposta, la resistència a càrregues elevades, la gestió de càrregues esperades i la sostenibilitat del rendiment a llarg termini.

És important destacar que no hi ha un "aprobat" o "suspès" fixat en aquestes proves. En canvi, els resultats proporcionen valors específics que han de ser considerats per mesurar el risc associat a l'aplicació. Aquests ajuden a determinar si pot complir amb els requisits de rendiment sota diverses condicions, i quines àrees necessiten millores per garantir una experiència d'usuari òptima i una operació eficient.

Al Departament de Salut ens basem en el marc de proves de rendiment definit per CTTI en el seu model de Qualitat de Solucions, que ha de ser de coneixement del proveïdor d'aplicació abans de definir-ne i dissenyar-ne els escenaris d'execució:

<https://qualitat.solucions.gencat.cat/guies/proves-rendiment/>

S'han identificat les següents proves com les més representatives i útils per assegurar que les aplicacions poden donar suport a les càrregues esperades i mesurar fins a quin punt un pic inesperat pot ser assolit per aquestes sense tenir una caiguda total del servei.

4.2 Prova de línia base

La prova de línia base de rendiment és una prova inicial amb un únic *VUser*, amb una duració curta (5 minuts aproximadament), que ens serveix per verificar que l'script simulador de l'escenari és correcte, i tenir un primer punt de referència sobre l'estat del sistema, comparatiu per la resta de proves a executar.

4.3 Prova d'estabilitat

La finalitat d'aquesta prova és validar l'estabilitat de l'aplicació amb la **càrrega esperada contínua** de Producció durant un temps prolongat (entre 4 i 6 hores). És la prova més descriptiva del rendiment de l'aplicació en condicions normals i en la qual cal fer focus per trobar punts de millora.

Ens serveix per observar el comportament del sistema en una jornada completa, identificant possibles problemes com *memory leaks*, connexions sense tancar o altres degradacions que només afloren amb l'aplicació en hores en funcionament.

4.4 Prova de càrrega

La prova de càrrega analitza el comportament del sistema **en condicions de pics** superiors a les condicions normals mitjanes de Producció.

Primer, s'arriba de manera controlada a les condicions normals i, després de mantenir una fase estable, es continua augmentant aquesta càrrega de forma progressiva fins assolir les condicions de pics màxim. L'objectiu és identificar com evoluciona el rendiment a mesura que la càrrega creix, observant els temps de resposta, el consum de recursos i l'estabilitat del sistema, i detectant possibles degradacions quan ens anem acostant als moments màxims de l'aplicació.

El valor principal d'aquesta prova rau a validar quins punts comencen a degradar-se primer quan se supera la càrrega esperada del sistema.

Duració d'aquesta prova: hauria de ser suficient amb 1 hora.

4.5 Prova d'estrès

La prova d'estrès, en canvi, verifica el comportament de l'aplicació en situacions **per sobre de les condicions de pic**, buscant intencionadament provocar la fallida del sistema degut a un desbordament de peticions, per determinar quins són els components que poden saturar.

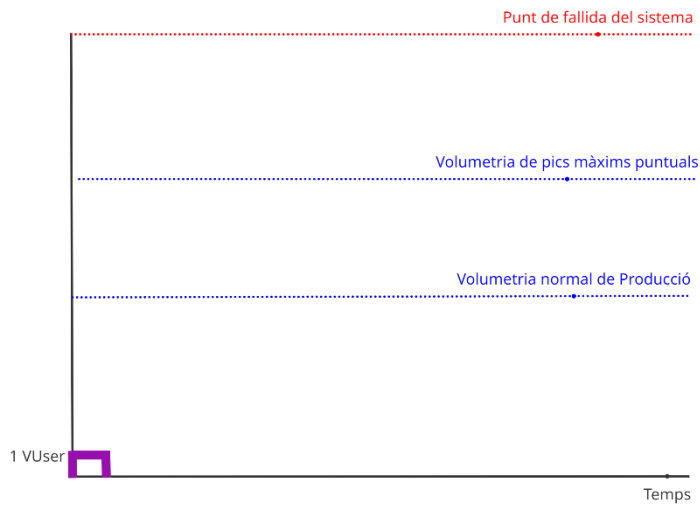
Consisteix en un augment controlat de la càrrega fins a assolir la capacitat prevista d'una prova de càrrega i, a continuació, una pujada pronunciada i agressiva d'usuaris virtuals, fins acabar forçant el sistema al seu punt de trencament.

La seva durada sol ser breu, ja que el propòsit és identificar ràpidament el límit real de capacitat del sistema, observant primer la degradació progressiva d'aquesta (com en la prova de càrrega), fins detectar en quin punt i component es produeix la fallida: bé per saturació de recursos, temps de resposta disparats o increment notable de les taxes d'error de peticions.

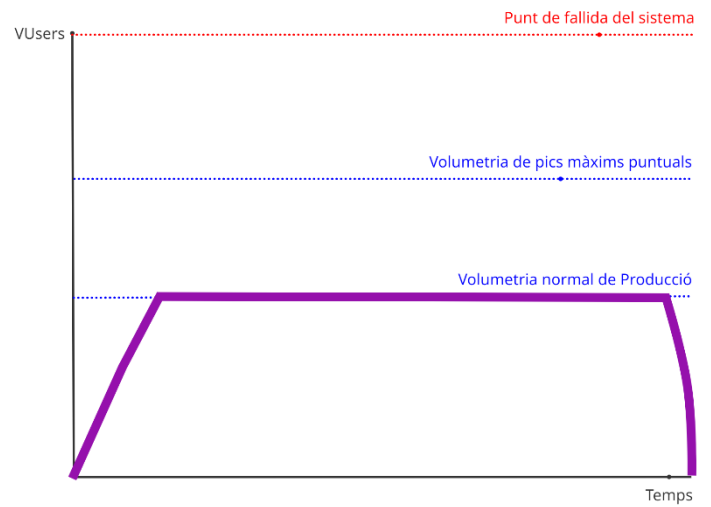
Això ens permetrà anticipar la reacció del sistema davant de situacions inesperades o extremes, o escalar de manera més eficient els recursos per aconseguir que el punt de trencament de l'aplicació sigui superior.

Duració d'aquesta prova: hauria de ser suficient amb menys d'1 hora.

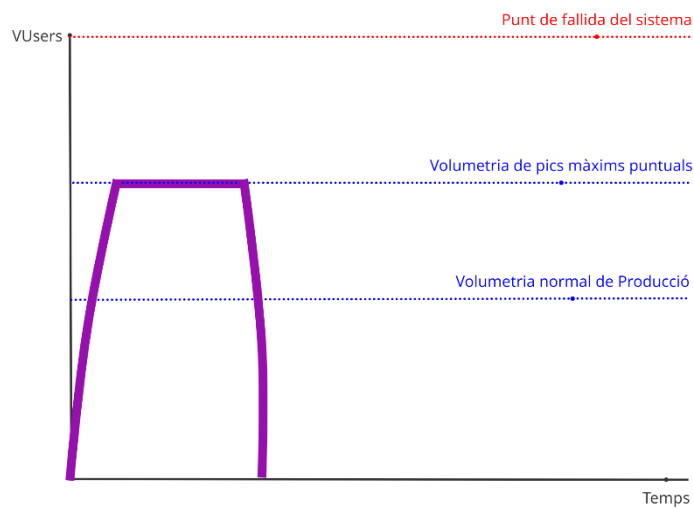
Taula comparativa del quatre tipus de proves



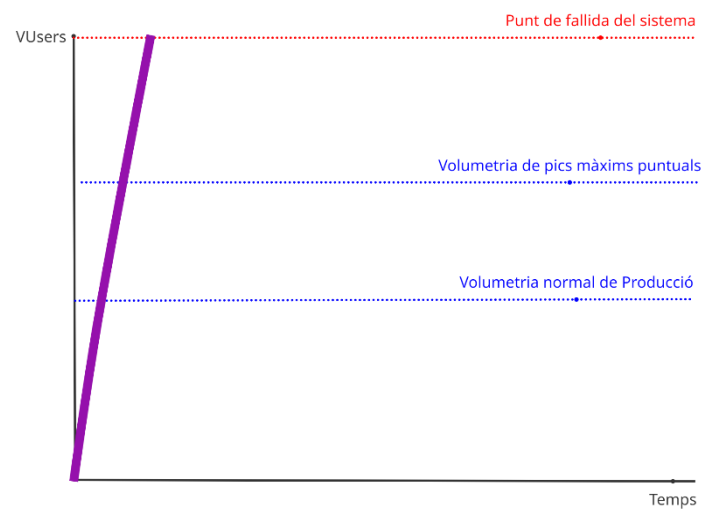
Prova de línia base



Prova d'estabilitat



Prova de càrrega



Prova d'estrès

5 Presentació de l'informe amb els resultats de les proves

5.1 Introducció

En aquest punt, incloem una petita guia per completar l'informe de proves de rendiment, amb l'objectiu de facilitar-ne la revisió, tant per als responsables de la solució, de l'Oficina d'Arquitectura i Qualitat de Solucions de Salut, i d'altres equips que el necessitin.

Intentarem definir una estructura comuna que, posteriorment, podrà ser ampliada o complementada per altres seccions, les quals depenen tant de l'estat de maduresa de l'aplicació com de les casuístiques pròpies de cadascuna d'elles.

Les plantilles per presentar aquest informe es troben a la documentació del CTTI:

https://qualitat.solucions.gencat.cat/procediments/proves/informe_resultat_proves/

5.2 Definició del cas d'ús

Les proves de rendiment s'han de fer sobre escenaris definits per l'equip de desenvolupament, i validats i aprovats per l'equip funcional o responsables de solució, que són els que millor coneixen les necessitats d'aquesta.

El escenaris hauran de simular les càrregues d'usuaris virtuals (VUsers) de cada prova, tal com s'ha explicat en el capítol anterior: línia base, càrrega, estrès i estabilitat. Per tant, cal indicar clarament quina serà la configuració de la volumetria i concurrència de cadascuna:

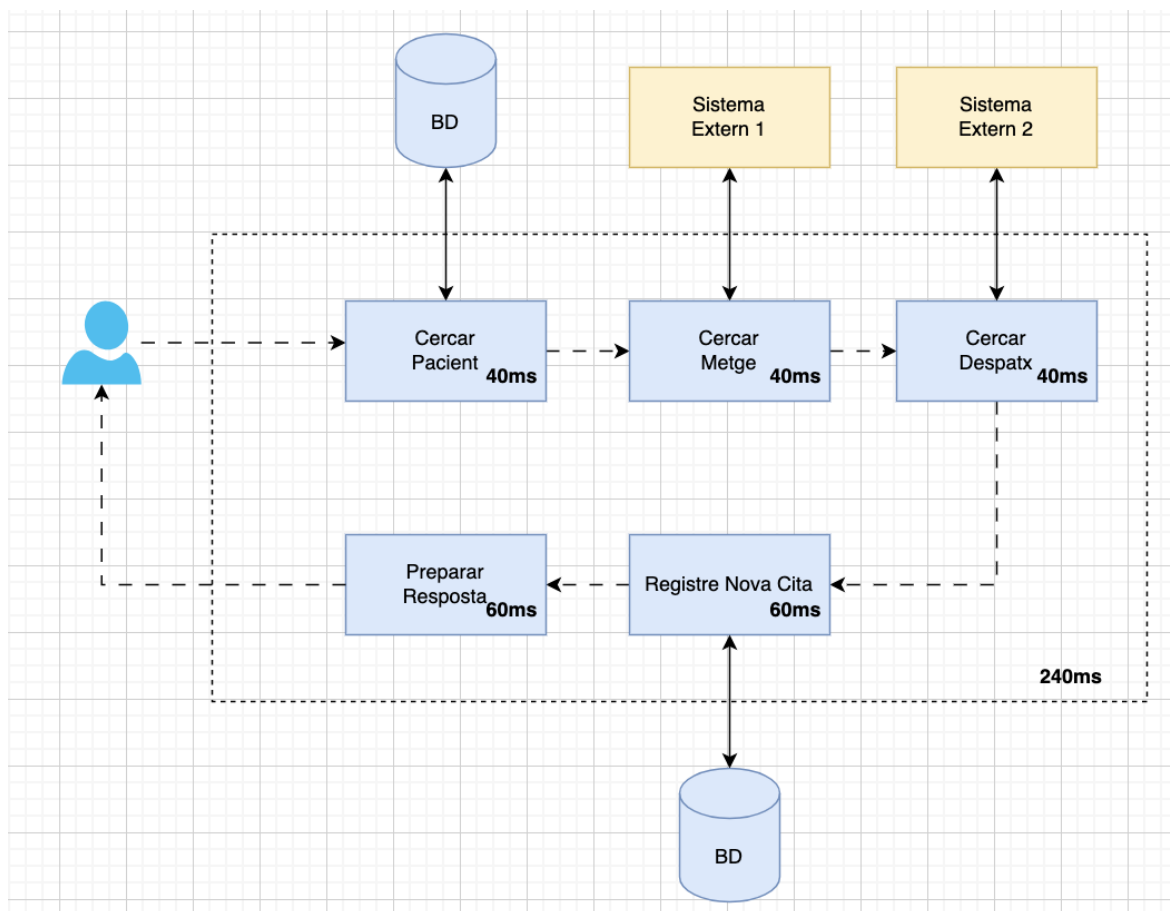
- Crides / Transaccions de Load Runner esperades per segon
- Nombre d'usuaris virtuals inicials, i ritme d'increment d'usuaris fins assolir el topall

Per definir cada cas d'ús, caldrà tenir en compte que el públic objectiu d'aquest document no necessàriament coneix el funcionalment de l'aplicació. La definició ha de ser prou clara i concisa per permetre el lector obtenir una visió suficient que li permeti entendre les integracions i fluxos que s'estan validant.

Un escenari podrà estar format per un o més casos d'ús executats seqüencialment o de forma concurrent, segons el comportament que es vulgui simular en cada escenari.

Considerem que la millor forma de presentar aquesta informació és amb un diagrama de seqüència en el qual s'incloguin tots els components involucrats en el cas d'ús, juntament amb una descripció textual de la seva funcionalitat.

A més, aquest diagrama de seqüència ens permetrà estimar i justificar els temps de resposta esperats, atenent als temps especificats en l'arquitectura de referència.



Idealment, el temps de resposta objectiu pel cas d'ús hauria de ser quantitatiu, i l'aplicació hauria de disposar de les mètriques o logs necessaris per calcular totes les interaccions descrites. En cas que alguna interacció no es pugui mesurar, es podrà explicar quina funcionalitat s'està utilitzant en aquest punt, i sempre s'haurà d'incloure com a proposta de millora l'actualització de l'observabilitat per tal de permetre un càlcul objectiu del temps.

5.3 Execució de les proves

Per cada prova executada, cal identificar les mètriques que ens permetran valorar el grau de compliment dels requisits no funcionals de l'arquitectura de referència. Amb aquest objectiu, i per facilitar tant la definició dels resultats esperats com l'obtenció dels resultats obtinguts, s'hauran d'indicar els valors per cada cas d'ús dins de l'escenari.

Caldrà presentar els resultats de l'execució de les diferents proves (línia base, càrrega, estrès i estabilitat) en els escenaris. Per cada cas d'ús o línia de transacció de Load Runner:

- Total de crides i % de respostes OK, fallides, *timeouts*...
- Temps de resposta mínim, mitjà, p=95% i màxim, en mil·lisegons

- Gràfic de volum de *VUsers*, transaccions executades i temps de resposta,... de l'inici fins al final de l'execució de la prova. Per exemple: en la d'estabilitat, aquesta línia hauria de ser bastant estable, i en la prova de càrrega, anar observant la degradació a mesura que augmenten els usuaris, i en la d'estrès, observar el punt de fallida de l'aplicació, etc.

Els resultats relatius a la gestió de la capacitat es mostren per cada escenari de prova en conjunt. Com s'ha esmentat anteriorment, un escenari pot estar compost per més d'un cas d'ús, però ens interessa mesurar a nivell de components i amb execucions concurrents dels cas d'ús (si així es planifica).

A part de les gràfiques de Load Runner, també cal recopilar les relatives a la infraestructura (com a mínim, CPU i memòria):

- Infraestructura núvol públic: gràfiques sobre les mètriques dels contenidors, bases de dades i altres possibles components.
- Infraestructura núvol privat o *on-premise*: aquestes gràfiques s'hauran de sol·licitar al corresponent CPD, per activar-ne el monitoratge durant la prova.
- Mètriques de negoci o pròpies de l'aplicació: aquestes s'hauran de recopilar dels corresponents *stacks* Prometheus+Grafana de l'aplicació, si escau.
- Components transversals, per exemple, Eventhub: gràfiques de *throughput* dels tòpics utilitzats, gràfics de consum dels nodes broker, o dels nodes *Kafka connect* o *Schema Registry*, si escau, etc...

5.4 Conclusions

Cal presentar unes conclusions respecte al comportament del sistema en les diferents proves de rendiment, justificant els casos on no s'hagin pogut assolir els temps de resposta exigits en l'arquitectura de la solució. També cal presentar possibles línies d'actuació per corregir els casos on el sistema no hagi estat capaç d'assolir el rendiment esperat: optimització de consultes, utilització de mecanismes de caché, reordenació de crides entre microserveis, reconfiguració de l'escalabilitat dels components o, fins i tot, evolutius o correctius respecte la implementació.

Aquest apartat de conclusions és el que permetrà al responsable de la solució decidir si el sistema està preparat per desplegar a Producció, en condicions òptimes.

6 Annexos

6.1 Annex 1. Procediment de valoració de les proves de rendiment

En cas que un evolutiu inclogui la necessitat de realitzar proves de rendiment, caldrà seguir el procediment següent com requisit previ a la posada en producció de la solució:

L'equip de projecte ha de planificar la realització de les proves de rendiment com a pas previ a la posada en producció, assegurant-se que es disposa del temps i dels recursos necessaris dins del calendari de l'evolutiu.

L'equip de projecte ha de revisar prèviament l'informe de definició de les proves de rendiment per assegurar-se que s'inclouen totes les dades requerides i que es compleixen els criteris establerts en la documentació corresponent.

Cal utilitzar les eines i infraestructures fixades per l'Oficina de Qualitat a l'hora d'executar aquestes proves.

Un cop finalitzades les proves de rendiment i completat el document amb els resultats obtinguts, el gestor de solucions ha de sol·licitar la valoració per part de l'Oficina d'Arquitectura i Qualitat. Aquesta sol·licitud s'ha de formalitzar mitjançant l'enviament d'un correu electrònic a l'adreça: **arquitectura.salut@gencat.cat**, adjuntant el document definitiu que inclogui els escenaris i els resultats.

L'equip de l'Oficina d'Arquitectura i Qualitat ha de revisar la documentació i completar una llista de control (**checklist**) establerta a tal efecte, amb els principals punts de control a verificar. Aquesta *checklist* ha d'incloure observacions i comentaris sobre qualsevol aspecte rellevant detectat durant la revisió d'aquest informe.

Un cop completada la revisió i la *checklist*, l'equip d'Arquitectura i Qualitat ha de remetre un document en format PDF amb la valoració al gestor de solucions.

Aquest, a partir dels resultats i comentaris reflectits, haurà de decidir si es considera adequat continuar amb el procés de posada en producció o si, per contra, és necessari dur a terme adaptacions o repetir les proves de rendiment.

Independentment de la decisió presa respecte al pas a producció, el document PDF amb la valoració de les proves de rendiment haurà d'incorporar-se a la documentació de qualitat del projecte. Aquest document ha d'estar disponible per a revisar-lo en futures auditories.

6.2 Annex 2. Responsabilitats dels equips implicats

Amb l'objectiu de garantir una execució i valoració correctes de les proves de rendiment, s'estableixen les responsabilitats següents per als equips participants:

Gestor de Solucions

- Vetllar perquè les proves de rendiment es tinguin en compte a l'hora d'estimar el termini de realització de l'evolutiú.
- Assegurar-se que els casos d'ús i escenaris proposats es corresponguin amb escenaris representatius de l'ús previst en producció de l'aplicació.
- Revisar la *checklist* i el document de proves per determinar la conveniència o no de posar l'evolutiú en producció.
- Valorar la inclusió, en el *backlog* del projecte, de les millores proposades per l'equip de desenvolupament com a resultat de les proves de rendiment.

Equip de Desenvolupament

- Definir els casos d'usos crítics que s'han d'incloure en els escenaris de prova, assegurant-se que siguin representatius en termes de càrrega i rellevants pel que fa a l'ús real en producció.
- Preparar i subministrar les dades necessàries per a l'execució de les proves, garantint-ne la validesa i representativitat. Això inclou, per exemple, preparar una col·lecció de dades significativa que permeti assolir la volumetria prevista per cada tipus de prova. Aquesta col·lecció de dades haurà d'estar degudament generada amb aleatorietat, per evitar ocurrences repetides de crides al sistema amb els mateixos paràmetres d'entrada i de resposta.
- Atendre i resoldre les incidències identificades durant l'execució de les proves.
- Aportar conclusions en relació amb els punts de sensibilitat detectats (colls d'ampolla, lentituds, fallides de components) ja siguin interns, com ara processaments pesats o sobrecàrrega de base de dades, com externs, com ara consum de serveis externs o plataformes transversals.
- Proporcionar context addicional a les proves, identificant millores:
 - A nivell d'observabilitat: proposant mètriques addicionals que permetin objectivar iterativament el temps de resposta esperat de cada cas d'ús.
 - A nivell d'aplicació: suggerint accions per millorar els temps que no compleixin els requeriments establerts, així com per reduir el consum de recursos (punts crítics de memòria, CPU, etc.).

Equip d'Arquitectura de solucions

- Revisar que la documentació de les proves inclogui la informació necessària per justificar els temps de resposta estimats per a cada cas d'ús, definits per l'equip de desenvolupament.
- Validar que es compleixen els llindars de rendiment establerts per als diferents casos d'ús, segons la descripció d'arquitectura de la solució.
- Completar la *checklist* amb la informació i observacions necessàries per a facilitar que el gestor de solucions pugui prioritzar les millores identificades per a l'optimització de l'aplicació en futures iteracions. Això inclou llindars de temps de resposta que no s'han complert durant les proves.
- Identificar mancances o nous requeriments a nivell de l'Arquitectura de Referència de Salut, i retroalimentar o revisar llurs principis i decisions en base a comportaments observats en proves, si són de caire general.

Equip de Qualitat

- Identificar les entregues que han d'incloure proves de rendiment, d'acord amb les polítiques definides, assegurant-se que aquesta necessitat quedi clarament establerta als equips implicats.
- Validar que la *checklist* d'arquitectura de proves de rendiment estigui inclosa en la documentació proporcionada, i realitzar les auditories corresponents quan es considerin necessàries.

6.3 Annex 3. Contingut de la *checklist*

La *checklist* es completa després de la revisió de l'informe de proves. Està estructurada en blocs diferenciats amb els criteris de validació corresponents. Per cada ítem, s'ha d'indicar el compliment o no compliment i, si escau, s'han d'afegir comentaris rellevants.

Bloc 1. Definició dels casos d'ús

- Inclou una descripció funcional de cada cas d'ús.
- Inclou el diagrama de seqüència de tots els casos d'ús.
- Els temps de resposta esperats per a cada cas d'ús respecten els llindars establerts per transacció segons els principis d'arquitectura.
- Si el cas d'ús interactua amb sistemes externs.
- El temps d'aquesta interacció es pot obtenir de les mètriques de l'aplicació.

- Si no es pot obtenir de les mètriques, es disposa d'una estimació justificada.

Bloc 2. Proves de línia base

- S'han executat proves de línia base per a tots els casos d'ús afectats per la *release*.
- S'inclou en el document el resum de resultats de l'eina de proves.
- S'inclou en el document l'histograma de resultats de l'eina d'observabilitat.

Bloc 3. Proves d'estrès

- S'han executat proves d'estrès per a tots els casos d'ús afectats per la *release*.
- El document inclou la identificació dels colls d'ampolla o punts de trencament detectats.
- S'inclou en el document el resum de resultats de l'eina de proves.
- S'inclou en el document l'histograma de resultats de l'eina d'observabilitat.

Bloc 4. Proves de càrrega

- S'han executat proves de càrrega per a tots els casos d'ús afectats per la *release*.
- S'inclou en el document el resum de resultats de l'eina de proves.
- S'inclou en el document l'histograma de resultats de l'eina d'observabilitat.
- Es compleixen els requeriments relacionats amb la gestió de la capacitat.
- Si no es compleixen els requeriments de gestió de la capacitat, s'inclou una justificació en el document.
- Es compleixen els requeriments relacionats amb el temps de resposta.
- Si no es compleixen els requeriments de temps de resposta, s'inclou una justificació en el document.

Bloc 5. Proves d'estabilitat

- S'han executat proves d'estabilitat per tots els casos d'ús afectats per la *release*.

- S'inclou en el document el resum de resultats de l'eina de proves.
- S'inclou en el document l'histograma de resultats de l'eina d'observabilitat.
- Es compleixen els requeriments relacionats amb la gestió de la capacitat.
- Si no es compleixen els requeriments de gestió de la capacitat, s'inclou una justificació en el document.
- Es compleixen els requeriments relacionats amb el temps de resposta.
- Si no es compleixen els requeriments de temps de resposta, s'inclou una justificació en el document.