PROVÁDĚCÍ KONCEPT SW ŘEŠENÍ (PK)

projektu
Národní informační systém integrovaného záchranného systému (NIS IZS)

část

G. Spolehlivost, disaster recovery a krizové scénáře

Dokument obsahuje:	Scénáře disaster recovery a mechanismy pro zajištění vysoké spolehlivosti systému NIS IZS
Verze:	6.1
Schválil za Dodavatele:	RNDr. Vladimír Příbramský
Aktualizace:	10/10/2014

Registrační číslo projektu CZ.1.06/3.4.00/11.07071

Obsah

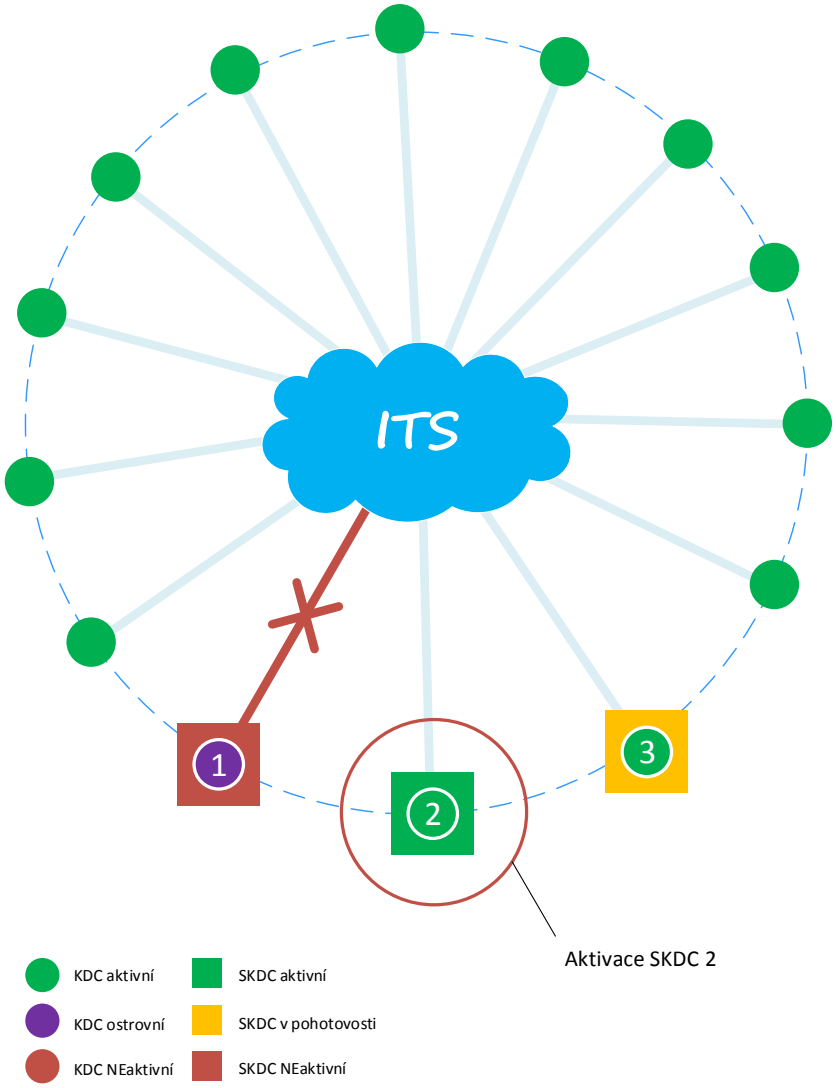
1 Spolehlivost navrženého řešení

Dodavatel návrhem řešení NIS IZS zajišťuje požadovanou vysokou dostupnost a 200% zálohu všech služeb provozovaných v režimu 24x7x365 a definovanou dostupnost dle smlouvy. Následující kapitoly popisují zajištění spolehlivosti jednotlivých subsystémů, krizové scénáře a mechanismy disaster recovery pro jednotlivé komponenty systému.

1.1 Spolehlivost na úrovni datových center

1.1.1 Výpadek SKDC

V případě výpadku obou CE routerů v SKDC dojde k automatickému přesměrování provozu na jiné SKDC.



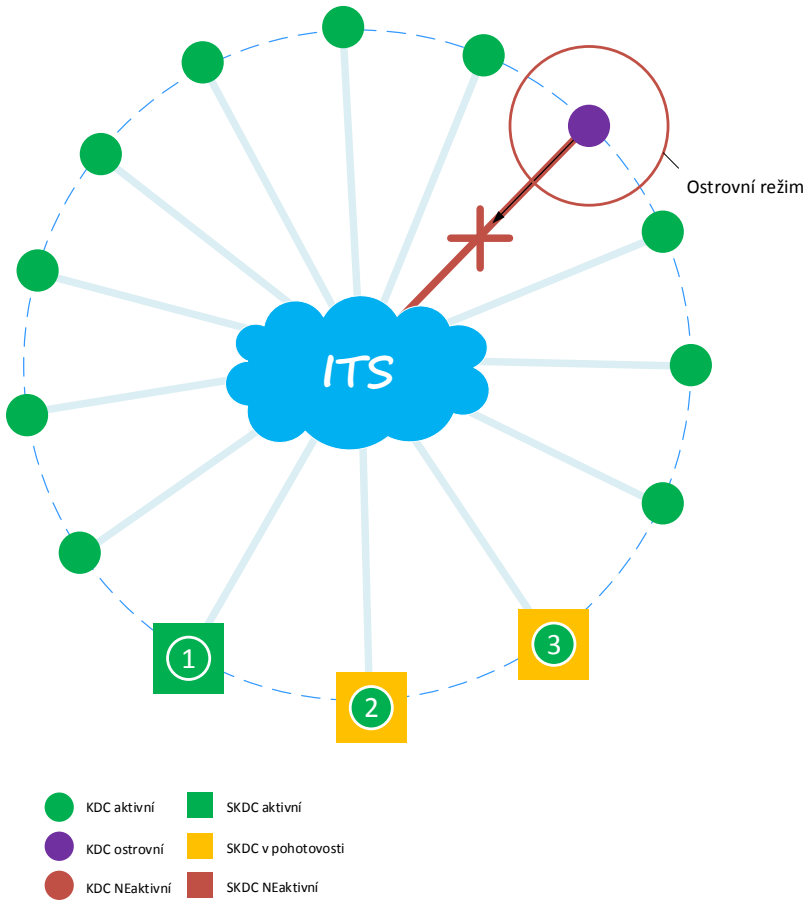
Obrázek 1 Výpadek SKDC

1.1.2 Výpadek KDC

V případě výpadku obou CE routerů v KDC (resp. v SKDC s funkcí KDC) dojde k automatickému přesměrování provozu ze všech systémů OŘ na jiné funkční KDC.

1.1.3 Odpojení kraje od ITS

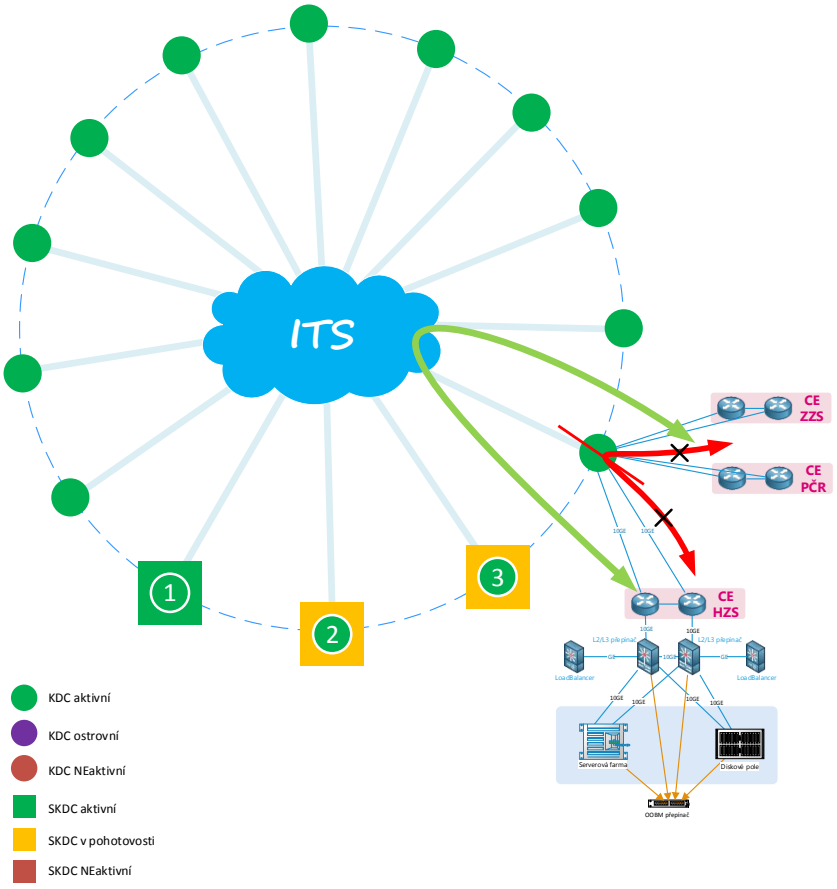
V případě odpojení celého kraje od ITS, přejde síťová infrastruktura kraje na ostrovní režim, kdy dojde k zachování plné datové komunikace v rámci kraje. V rámci kraje bude zajištěno plnohodnotné směrování, tak aby zůstal veškerý krajský provoz zachován.



Obrázek 2 Odpojení od ITS - ostrovní režim

1.1.4 Rozdělení kraje

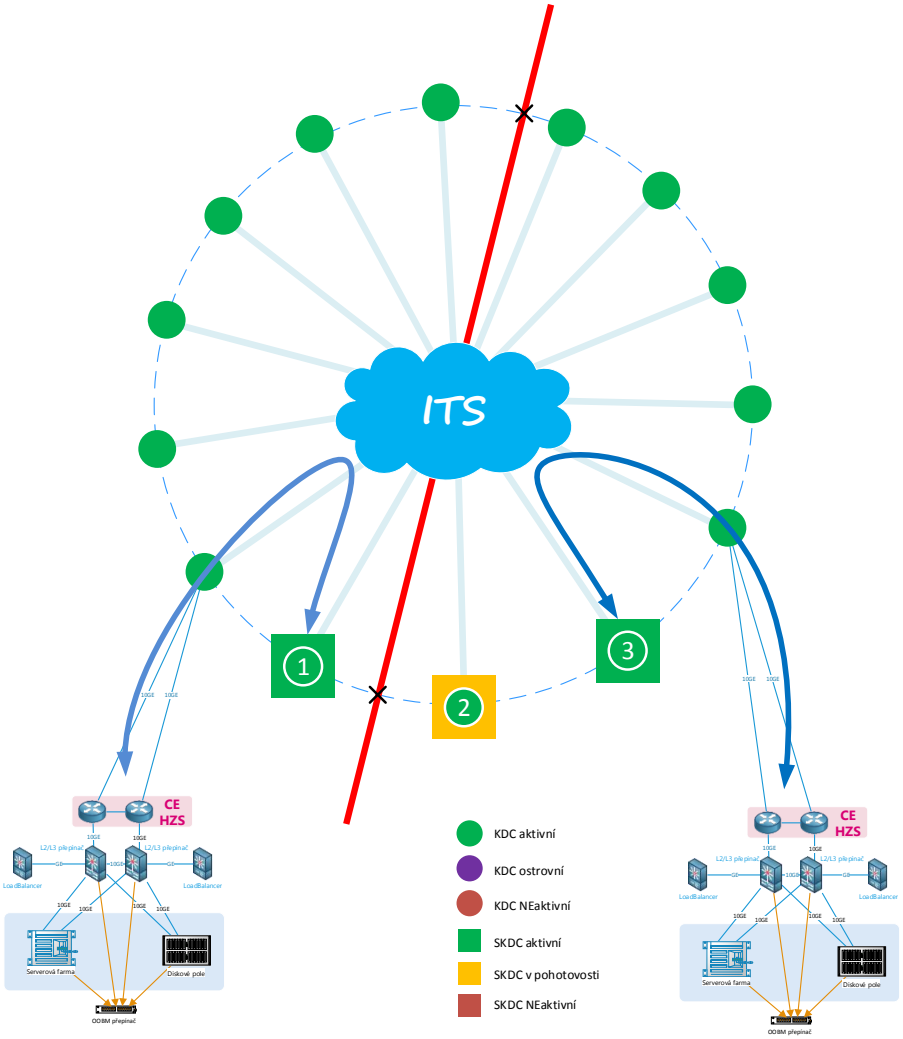
V případě rozdělení kraje, kdy nebude přímá komunikace mezi jednotlivými složkami, dojde k přesměrování datových toků mezi složkami a KDC pomocí sítě ITS přes republikový okruh. Toto přesměrování toků probíhá do 150ms, dojde však ke zvýšení latence, která v rámci jednoho ITS okruhu nepřesáhne 30ms.



Obrázek 3 Rozdělení kraje

1.1.5 Rozpůlení ITS

V případě rozdělení/„rozpůlení“ ITS bude komunikace v rozdělených částech plně zachována, jednotlivá OŘ budou mít plný přístup ke KDC/SKDC které zůstanou v dané části ITS. Komunikace mezi rozdělenými SKDC/KDC nebude možná.



Obrázek 4 Rozpůlení ITS

1.1.6 Ztrátovost na lince

V případě ztrátovosti linky dojde k automatickému odpojení problémové linky a konvergence zachová plnou funkcionalitu všech propojení. V případě ztrátovosti na síťových zařízeních dojde k vyřazení chybujícího zařízení z provozu, konvergence ITS se opět postará o obnovení datových spojení. Konvergence ITS je definována na 150ms.

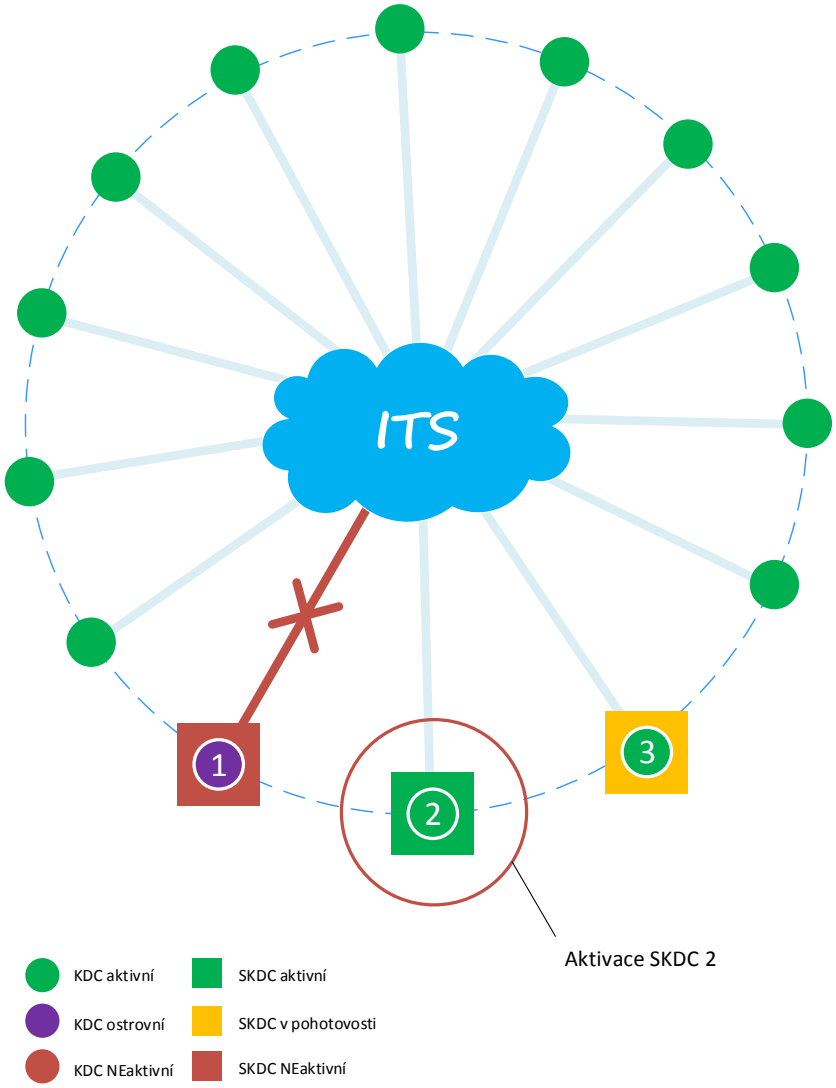
1.1.7 Pád ITS

V případě celkového pádu ITS nebudou funkční žádná propojení. Všechna KDC budou pracovat v ostrovním provozu.

1.1.8 Výpadek jednoho SKDC

Po výpadku jednoho SKDC budou v ostatních aktivních SKDC dostupné veškeré služby NIS IZS. Po obnově vypadlého DC dojde k jeho uvedení do stavu shodného s aktivními DC (automaticky - dosynchronizace DB, přenesení změn na souborovém systému, manuálně – synchronizace

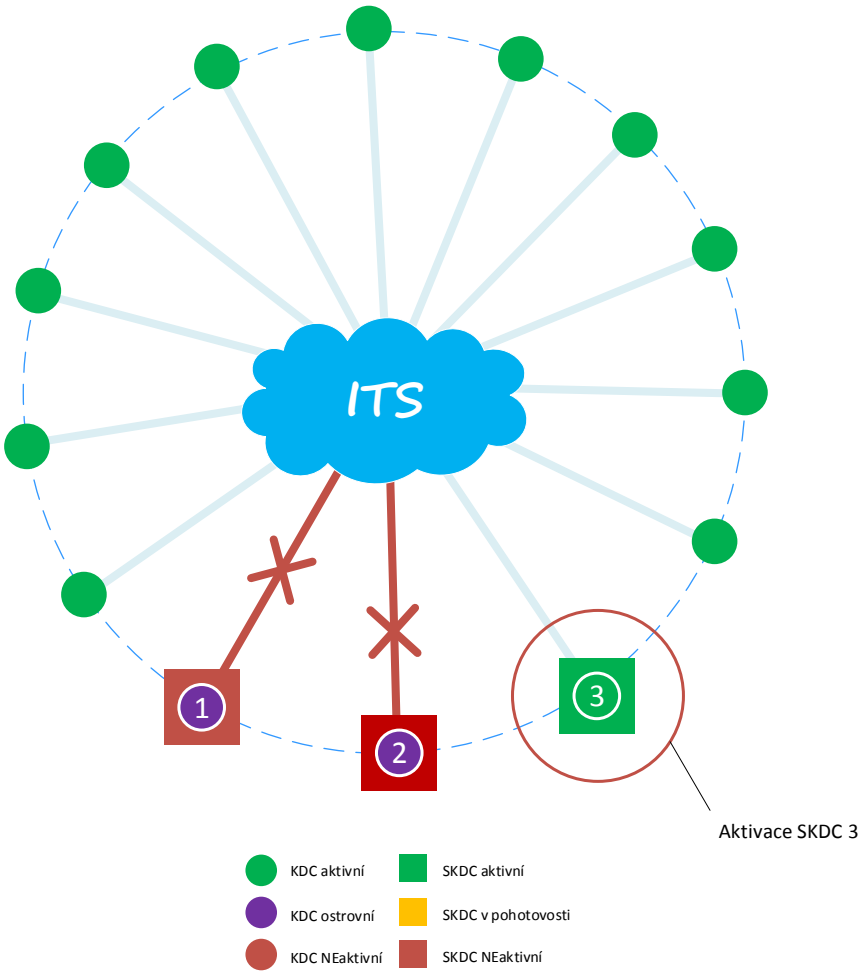
systémových a aplikačních updatů, resp. konfigurací v případě), následně bude ručně obnovena synchronizace dat opět na všechna tři SKDC.



Obrázek 5 Výpadek jednoho SKDC

1.1.9 Výpadek dvou SKDC

Krizový scénář počítá s výpadkem dvou DC (najednou nebo postupně), kdy je veškerý provoz směrován na poslední zbývající SKDC. Běžící SKDC bude mít nastavenou automatickou politiku v Hypervizoru, který pomocí nastavených pravidel pro rozdělování systémových zdrojů zajistí přednost běhu produkčního prostředí na úkor testovacího a preprodukčního, aby zajistilo dostatečný výkon na odbavení operátorských požadavků. Scénář obnovy počítá s postupnou obnovou nejdříve jednoho SKDC a po jeho uvedení do provozu bude započata obnova zbývajícího SKDC (postup bude shodný jako v případě scénáře výpadku jednoho DC).



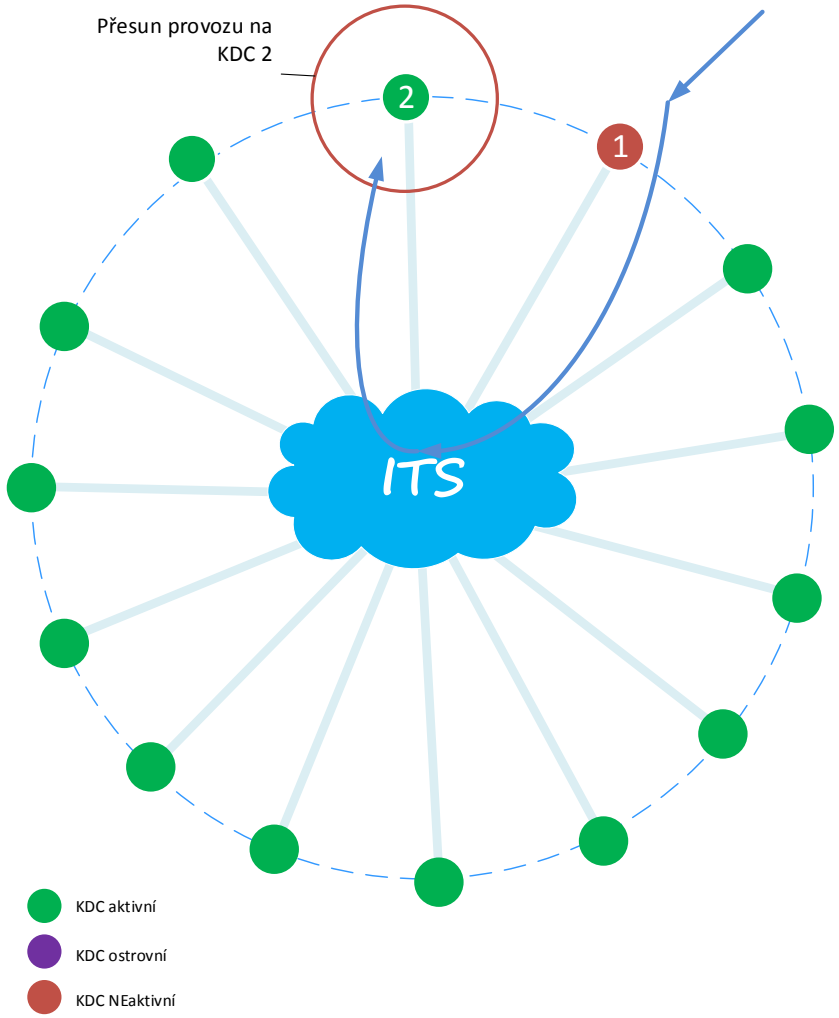
Obrázek 6 Výpadek dvou SKDC

1.1.10 Výpadek tří SKDC

Krizový scénář počítá s výpadkem tří SKDC (najednou nebo postupně). Z pohledu softwarových komponent KDC se jedná o stejnou situaci, jako při ostrovním režimu.

1.1.11 Výpadek KDC

Tento scénář předpokládá nedostupnost KDC z pohledu serverové infrastruktury. Zálohu pro KDC zajišťuje primárně další KDC.



Obrázek 7 Výpadek KDC

1.1.12 Výpadek elektrické energie

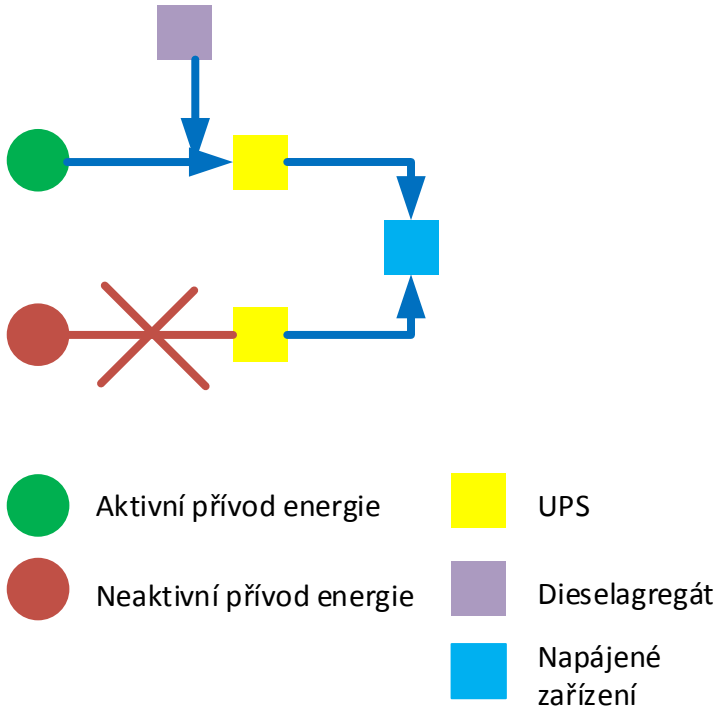
Krizový scénář počítá s možným výpadkem napájení el. energií tak, že se předpokládá, že přívod el. energie pro potřeby napájení zařízení v serverovně DC bude rozdělen do dvou nezávislých větví, minimálně na úrovni zdvojených rozvaděčů. Do každého racku budou přivedeny dva nezávislé přívody a připojeny k PDU (Power Distribution Unit) racku, které jsou taktéž zdvojeny.

Dojde-li k výpadku jedné napájecí větve, jsou všechna zařízení v racku napájena přes zbývající PDU a odpovídající DC zdroje (které jsou v každém konkrétním zařízení taktéž zdvojeny).

Pro případ celkového výpadku dodávky el. energie od dodavatele budou v každé napájecí větvi zapojeny zdvojené UPS odpovídajícího výkonu, které umožní bezpečné uložení a zachování konzistentních stavů dat aplikací běžících v rámci NIS IZS a také případné vypnutí všech zařízení, která jsou součástí řešení NIS IZS. UPS bude schopna dodávat el. energii odpovídajícího

výkonu po dobu nejméně 20 minut. K překlenutí doby přerušení dodávky el. energie bude zároveň do el. sítě automaticky zapojen nezávislý zdroj el. energie v podobě dieselagregátu, který bude schopen dodávat minimálně takový výkon, aby byl schopen udržet provoz všech zařízení NIS IZS po libovolně dlouhou dobu, a to pouze doplňováním potřebných provozních zdrojů.

Pro případ výpadku chladicího systému DC je počítáno s jeho kompletním zdvojením v rámci celého DC v kapacitě potřebné pro udržení provozního prostředí pro všechna zařízení NIS IZS.



Obrázek 8 záloha napájení

1.2 Spolehlivost dodávané infrastruktury

1.2.1 Síťová infrastruktura

1.2.1.1 Výpadek jednoho síťového prvku v redundanci

V případě výpadku jednoho jakéhokoliv prvku v redundanci dojde k okamžitému převzetí plné funkcionality na druhý prvek. Tato konvergence bude v řádech jednotek ms a neovlivní běžný provoz sítě. Pro příklad vypadne jeden z CE směrovačů, kontrolní mechanizmy například HSRP zareaguje okamžitě změnou směrovací tabulky na druhém CE a dojde k obnovení datového okruhu. Podobně se zachová přepínačová, firewallová a load-balancerová vrstva sítě. V tomto scénáři nedojde k rozpojení datových spojení.

1.2.2 Serverová infrastruktura

Architektura řešení vychází z koncepce čtrnácti aktivních nezávislých KDC s rozložením zátěže, vybrané tři kraje jsou navíc rozšířené o funkce SKDC (krajské datové super-centrum). SKDC poskytuje (navíc oproti KDC) centrální databázi událostí, administrátorskou konzoli, centrální správu a aktualizaci GIS dat z externích systémů (RUIAN, JSDI, ČHMU) a složkový GIS pro PČR. Rovnoměrná distribuce výkonu přes všechna KDC/SKDC a využití/zátěže datového provozu je monitorována a spravována síťovými load-balancery v HA zapojení.

V architektuře nezávislých KDC s rozložením zátěže bude nastaven nadřazený monitoring v Hypervizoru, který bude rozkládat zátěž na všechna KDC.

Disaster recovery proces má popsat proces obnovy normálního provozu NIS-IZS při neočekávaném výpadku/havárii, při němž dojde k přerušení nebo omezení činnosti některé ze systémových součástí nebo výpadku více takových součástí. Příčiny výpadku mohou být následující:

- Výpadek na komunikační síti;
- Selhání HW;
- Lidská chyba (manipulace s HW, implementace nových verzí SW);
- Selhání DC (el. energie, povodně, bouřky, požáry, apod.);

V takovýchto případech je potřeba zajistit detekci výpadku, kontinuitu provozu, krizovou komunikaci, ochranu infrastruktury, integritu informačních systémů a obnovení provozu v době, která je definována v provozní smlouvě.

Pro detekci výpadku je nutné zařadit do infrastruktury NIS-IZS monitorovací prvek ve formě dohledového pracoviště, které bude umístěno mimo datová centra NIS IZS. Toto dohledové pracoviště bude nepřetržitě sledovat a vyhodnocovat dostupnost všech sledovaných systémů v jednotlivých KDC/SKDC a v případě nenadálé události a/nebo výpadku, započne proces k zajištění kontinuity provozu, zahájí krizovou komunikaci a podnikne všechny nezbytné kroky k obnovení plného provozu.

Proces zajištění kontinuity provozu spočívá v následujících krocích:

1. Detekce chyby
2. Přesměrování provozu na běžící DC
3. Stanovení příčiny a určení doby, kdy dojde k obnovení plného provozu
4. Obnova DC s výpadkem:
    A. Obecná

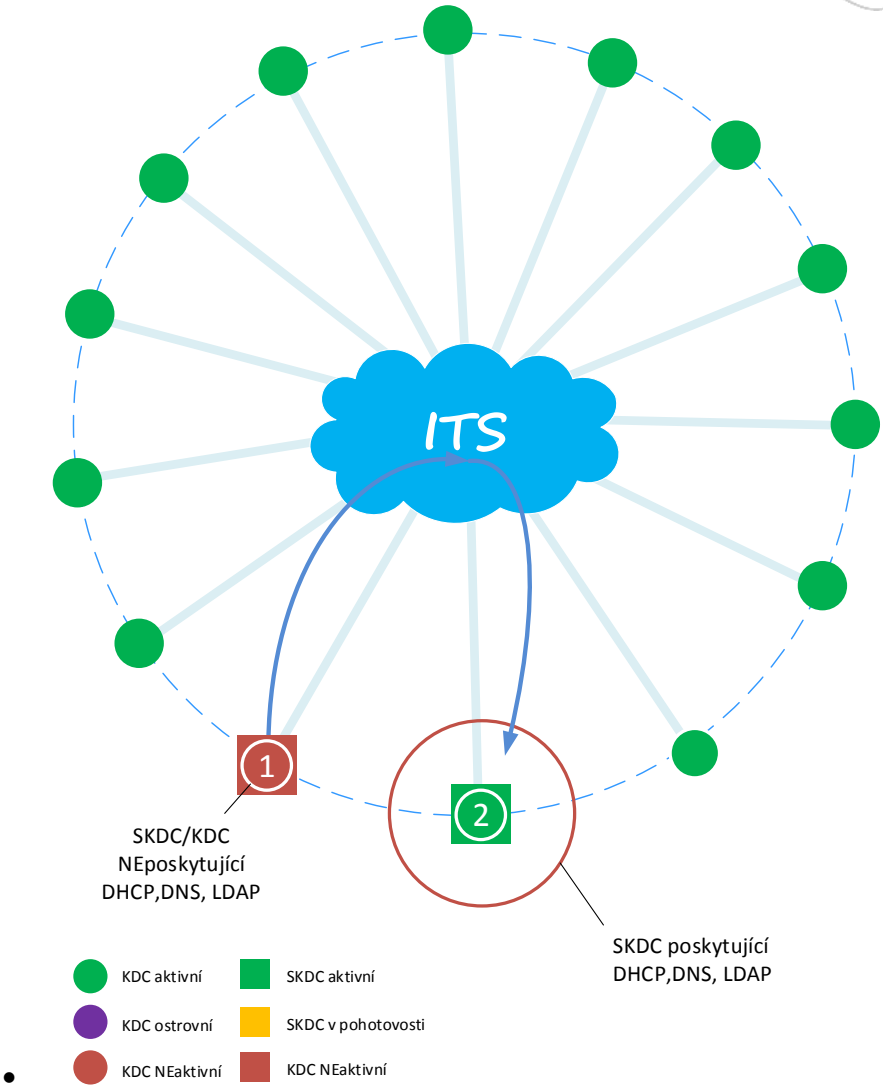
- Nahlášení výpadku s identifikací problému a doby jeho pravděpodobného odstranění.
- Definovaný scénář, který je samostatně uvedený níže (např. selhání HW, viz ad. 5.2.)
- Kontrola konzistentního stavu databázových operací.
- Synchronizace (nebo kontrola) souborových systémů obsahujících telefonní záznamy.
- Připojení DC zpět do provozu;
- Kontrola dosynchronizace dat ze vzniklého časového okna do plně synchronního stavu.

B. Selhání HW
   - Výměna nefunkčního HW

C. Při lidské chybě (např. upgrade SW)
   - Obnovení VM do stavu před zásahem

1.2.3 Výpadek serverů (DHCP, DNS, LDAP)

- V tomto krizovém scénáři se počítá s výpadkem DHCP, DNS, LDAP a to i v případě výpadku všech těchto služeb zároveň.
- V případě výpadku výše zmíněných služeb v SKDC nebo KDC, jsou tyto služby automaticky přístupné z výpadkem nepostiženého SKDC
- Dostupnost výše zmíněných služeb je zajištěna pomocí clusterování služeb a zároveň pomocí loadbalaceru
- Po obnově výše zmíněných služeb dojde k jeho uvedení do stavu shodného s aktivními datovými centra (dosynchronizace nastavení), následně bude obnoveno směrování na výše zmíněné služby.



Obrázek 9 Výpadek DHCP, DNS, LDAP

1.2.4 Selhání HW

Krizový scénář předchází možnému selhání jakéhokoliv HW tak, že veškerý HW je minimálně zdvojen, a to jak na úrovni základních komponent, tak i na úrovni celých DC. V případě selhání HW se počítá se zahájením servisních prací dle sjednaného SLA v provozní smlouvě.

1. Výpadek rack serveru
    a. Výpadek napájecího zdroje
    b. Výpadek ventilátoru
        - Technik vymění v rámci SLA vadný díl za běhu systému. Tyto subsystémy jsou vyměnitelné za běhu (technologie hot-plug)
    c. Závažnější závada (síť, řadič disku, CPU, RAM, základní deska atd.)
        - V tomto scénáři se server odstaví a technik provede potřebnou opravu. Do provozu se RACK server vrátí nainstalováním a nastavením Hypervizoru. Následně se procesně uvede do provozního stavu.
2. Výpadek více jak jednoho rack serveru

- V tomto scénáři se přepokládá odstavení celého datacentra viz výpadek SKDC podle bodu 1.1.8 a viz výpadek KDC podle bodu 0.
- Na serverech provede technik potřebnou opravu. Do provozu se RACK servery vrátí nainstalováním a nastavením Hypervizoru. Následně se procesně uvede do provozního stavu.

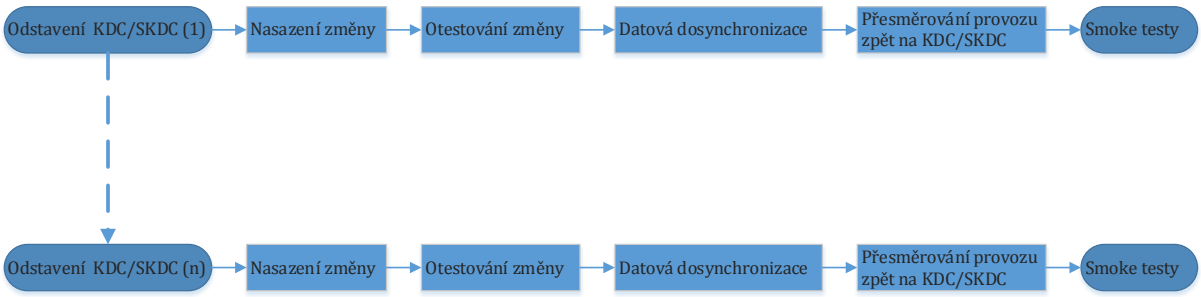
3. Výpadek diskového pole v SKDC/KDC
   a. Havárie disku (redundance dle RAID)
      - Technik vymění disk v rámci SLA za běhu systému. Tyto sub systémy jsou vyměnitelné za běhu (technologie hot-swap).
   b. Havárie zdroje (redundance 2N)
      - Technik zdroj disk v rámci SLA za běhu systému. Tyto sub systémy jsou vyměnitelné za běhu (technologie hot-swap).
   c. Havárie řadiče
      - Přebírá jeho provoz druhý, v případě havárie obou (nebo příslušné police) se odstavuje diskové pole, provoz se pomocí load balanceru směruje na jiné datacentrum (viz výpadek SKDC podle bodu 1.1.8 a viz výpadek KDC podle bodu 0.) a technik provede výměnu dílu, otestuje a obnoví provoz diskového pole a připojí do Hypervizoru. Následně se procesně uvede do provozního stavu.
4. Archivační jednotka LTO (HW)
   - V případě výpadku celé archivační jednotky LTO přebírá automaticky celý provoz pomocí load balanceru archivační jednotka v druhém SKDC. Po zprovoznění archivační jednotky LTO je následně ručně přesměrován provoz zpět a automaticky se zahájí dosynchronizace dat.
   - V případě výpadku všech tří archivačních jednotek se archivovaná data ukládají na provozní diskové pole v každém SKDC. Po zprovoznění jedné archivační jednotky jsou archivovaná data z provozních diskových polí automaticky přesunuta do archivační jednotky pomocí archivačního SW. Všechny 3 archivační jednotky LTO se zprovozňují jedna po druhé a ihned do zprovoznění automaticky pomocí load balanceru a archivačního SW přecházejí do normálního provozního stavu. Každá zprovozněná archivační jednotka se automaticky sesynchronizuje s ostatními běžícími archivačními jednotkami.

1.2.5 Lidská chyba (manipulace s HW, nasazování nových verzí SW)

V případě lidské chyby (zpravidla chyba administrátora), se následky řeší podle typu způsobených škod. V případě nasazování nových SW releasů je vždy popsán mechanismus pro návrat do posledního provozního stavu. V případě zásadních změn aplikace (např. změna databázových objektů) je k nasazování vypracován podrobný scénář s tím, že bude zpravidla postupováno v následujících krocích:

1. Odstavení KDC/SKDC
2. Nasazení změny

3. Otestování změny
4. Datová dosynchronizace
5. Přesměrování provozu zpět na KDC/SKDC
6. Smoke testy
7. Odstavení dalšího KDC/SKD
8. Nasazení změny
9. Otestování změny
10. Datová dosynchronizace
11. Obnovení plného provozu KDC/SKDC
12. Smoke testy
13. Výše zmíněným postupem probíhá nasazování změn na všech KDC/SKDC doku nejsou na všech KDC/SKDC aplikované změny

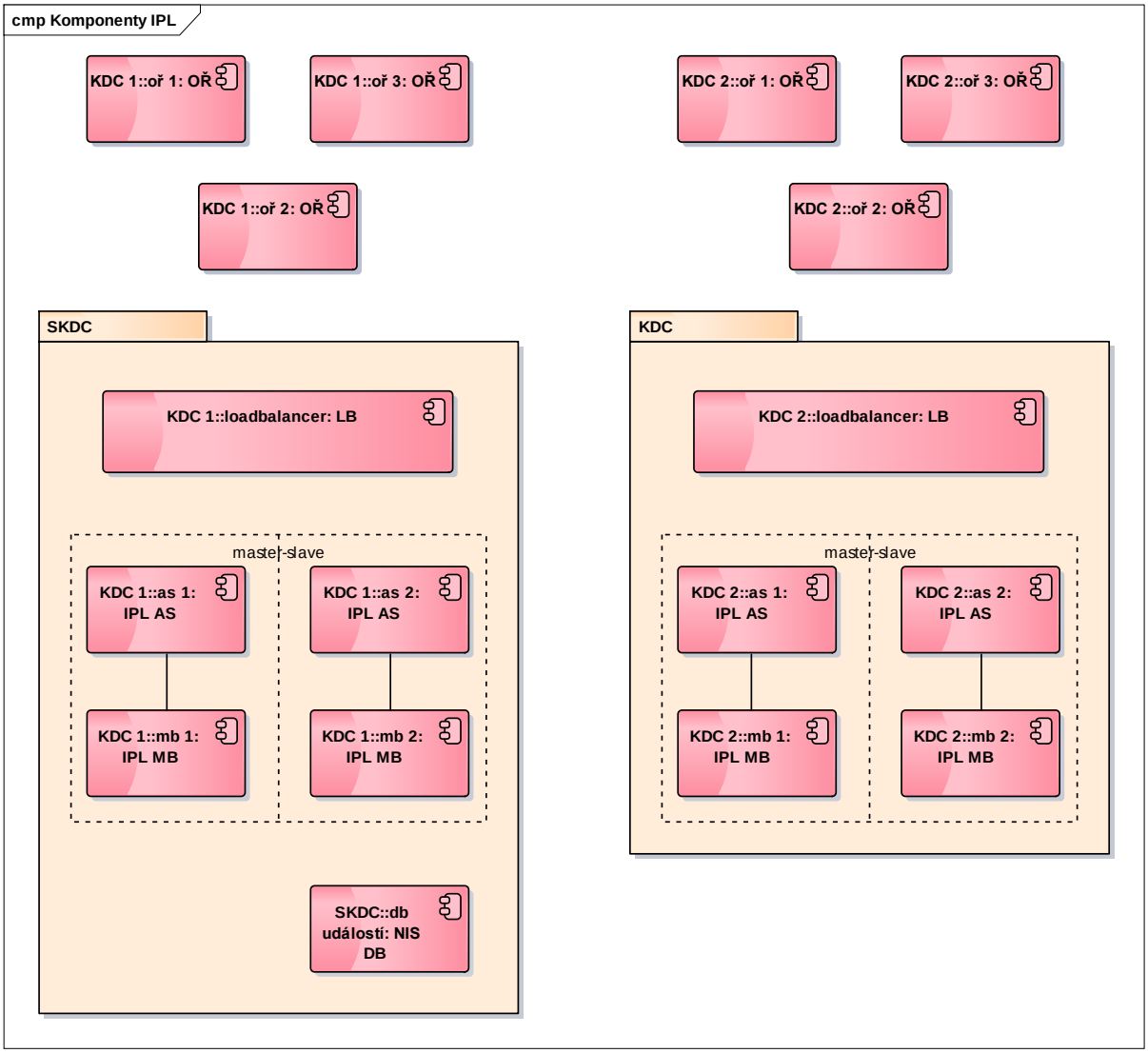


Obrázek 10 Nasazování změn

1.3 Spolehlivost služeb

1.3.1 Spolehlivost IPL

IPL systému NIS se skládá z těchto komponent:

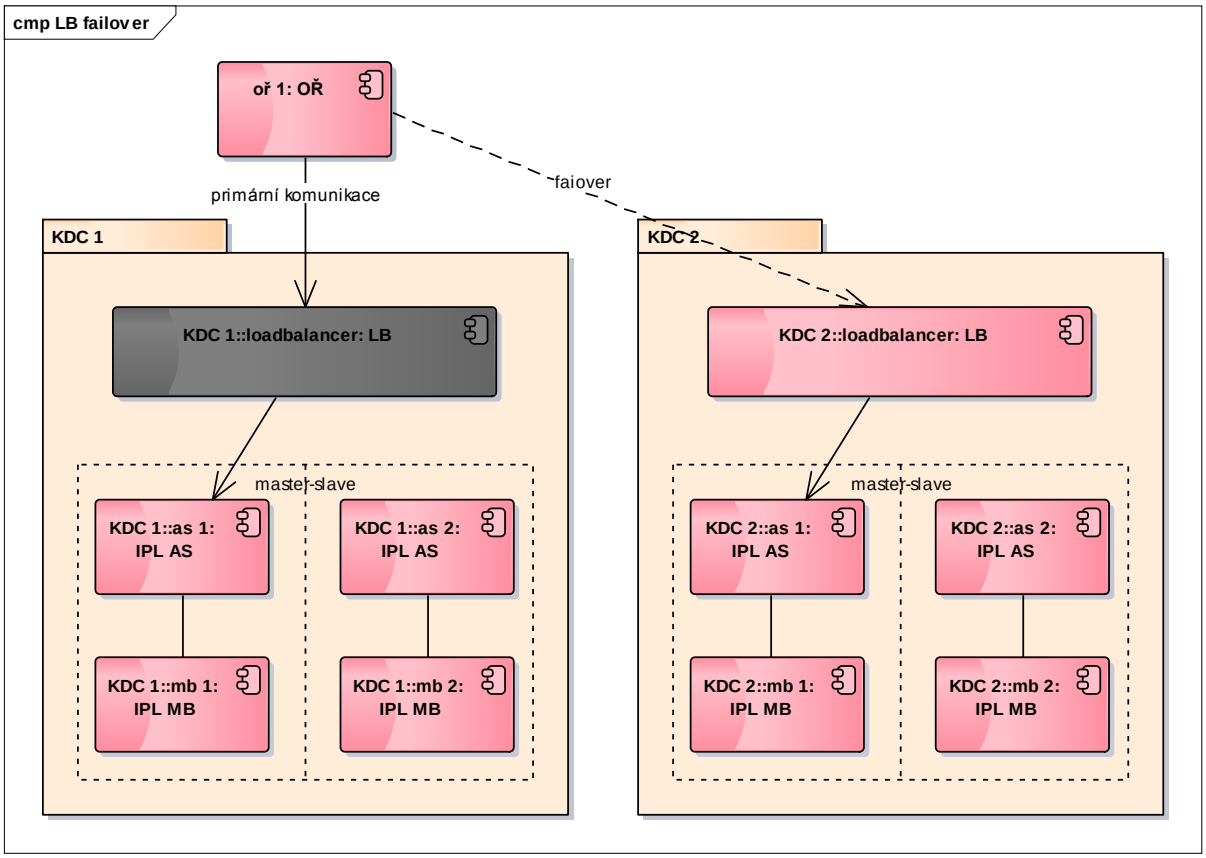


Obrázek 7 Komponenty IPL

- Load balancer – vysoce dostupný load balancer na kraji je společnou komponentou pro funkční blok GIS i IPL, má vnitřně zdvojenou architekturu.
- Aplikační server – aplikační server IPL části softwarového řešení NIS IZS, obsahuje vlastní aplikační logiku. V jeho kontextu navíc běží některé zakázkově vyvíjené služby kombinující data mimořádných událostí i GIS (Služby pro zjištění stavu S&P a operační situace) a administrátorská aplikace (pouze na SKDC).
- Message broker – messaging IPL slouží k frontování příchozích a odchozích zpráv a oddělení jejich příjmu a odeslání od vlastního zpracování. Je logickou komponentou, fyzicky je však součástí aplikačního serveru, a běží v jeho kontextu.
- Databáze událostí – transparentní databázová vrstva pro ukládání dat mimořádných událostí, nastavení, S&P a operačních situací. Je přítomna pouze v SKDC.

Více informací o komponentách IPL viz část B (software) tohoto prováděcího konceptu. Dále jsou popsány pouze výpadku těchto komponent.

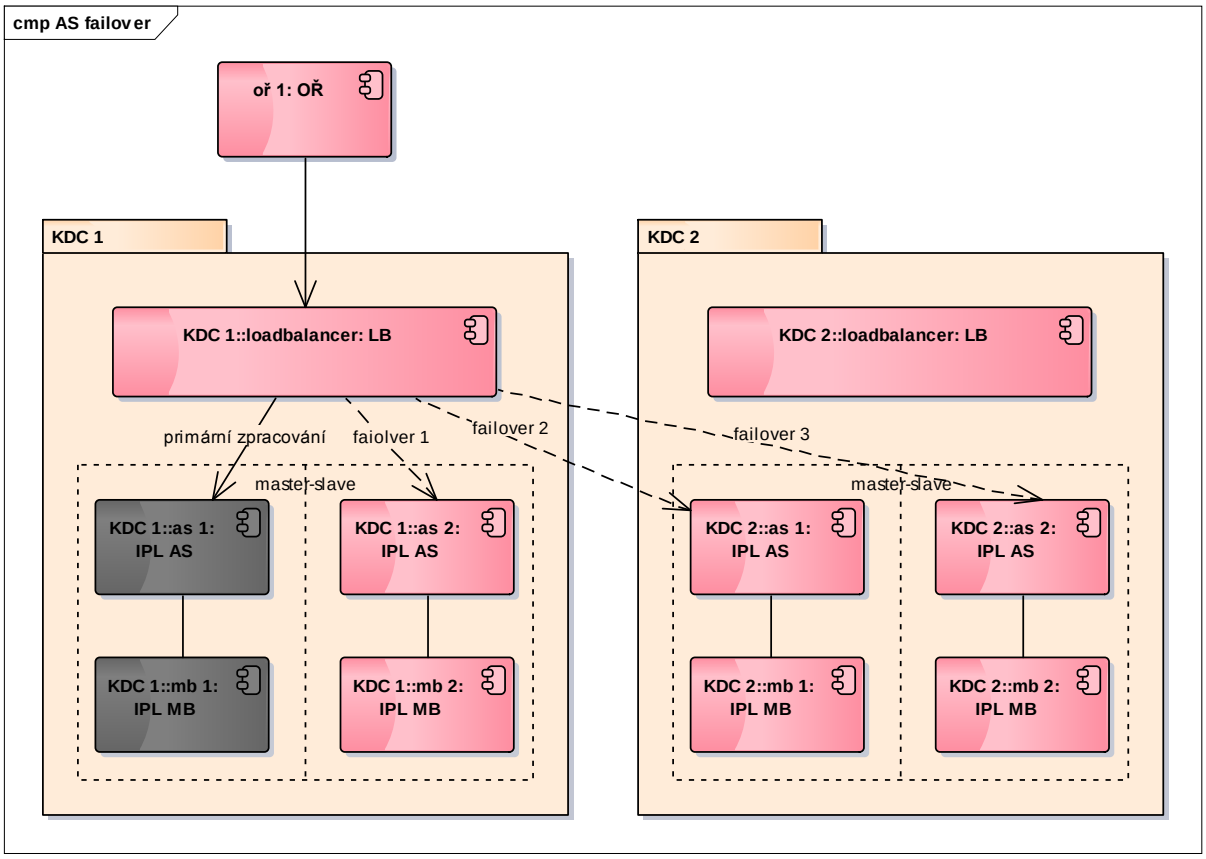
1.3.1.1 Load balancer



Obrázek 8 Výpadek load balanceru

- Load balancer je koncovým bodem pro komunikaci systémů OŘ se systémem NIS IZS. OŘ se primárně obrací na load balancer vlastního kraje.
- Load balancer v KDC/SKDC je vnitřně zdvojen a jako celek dostupný pro okolní systémy pod jednou IP adresou. Vnitřní architekturou load balanceru je tak již zajištěna 100% záloha jeho funkcionality. Při výpadku jednoho vnitřního nodu load balanceru je tento okamžitě nahrazen jeho druhým nodem bez ovlivnění ostatních částí systému. IPL je nadále plně funkční.
- Pro případ výpadku celého load balanceru (obou jeho vnitřních nodů), musí být OŘ schopné přesměrovat své požadavky na load balancer druhého kraje, který přebírá funkcionalitu kraje prvního. Všechny aplikační komponenty IPL (mimo databázi) jsou na úrovni KDC/SKDC bezestavové a rovnocenné, volba druhého kraje tedy může být z hlediska zpracování libovolná. Z důvodu rozložení zátěže bude celkový failover komunikace z OŘ jednotlivé kraje nastaven různě (do kruhu), tedy OŘ příslušející ke KDC 1 budou směřovat failover požadavky na KDC 2, OŘ z KDC 2 na KDC 3 atd. až OŘ z KDC 14 na KDC 1. Load balancer náhradního kraje směřuje požadavek dále obvykle na vlastní aplikační servery, kde dojde k jeho zpracování.
- OŘ může být dále, nad rámec 200% zálohy, konfigurováno pro použití load balancerů třetího a dalších krajů až do celkového počtu všech 14 KDC/SKDC.

1.3.1.2 Aplikační servery IPL



Obrázek 9 Výpadek aplikačního serveru IPL

- Aplikační server IPL je v každém KDC/SKDC přítomen redundantně 2x v konfiguraci master-slave clusteru.
- Vlastního zpracování se za normální situace zúčastní pouze AS master, oba aplikační servery jsou však stále spuštěny a AS slave může v případě pádu master nodu okamžitě převzít plnou funkcionalitu, čímž je zajištěna jeho 100% záloha.
- V případě výpadku obou nodů AS clusteru na vlastním KDC je komunikace přesměrována pomocí load balanceru na další kraj. OŘ nadále komunikuje s load balancerem vlastního kraje a o případném přesměrování se nedozví, systém je pro něj nadále plně funkční. Aplikační servery jsou dimenzovány na dostatečný výkon, aby mohly obsloužit provoz více krajů. Konfigurace failover komunikace load balancerů na aplikační servery dalších krajů je nastavena pro každý kraj různě (do kruhu), tedy LB z KDC 1 směřuje v případě výpadku vlastních AS požadavky na AS v KDC 2, LB z KDC 2 na KDC 2 atd. až LB z KDC 14 na KDC 1.
- LB každého kraje bude, nad rámec 200% zálohy, konfigurován dále pro použití aplikačních serverů třetího a dalších krajů až do celkového počtu všech 14 KDC/SKDC.

1.3.1.3 Messaging IPL

- MB běží v kontextu aplikačního serveru IPL, a nemůže dojít k jeho samostatnému výpadku. Výpadek master nodu MB je zároveň vlastním výpadkem master nodu AS, zpracování v takovém případě přebírá slave nod AS včetně slave MB. Zprávy uložené do front messagingové vrstvy MB master nodem jsou v případě jeho výpadku nadále plynule zpracovány slave MB.

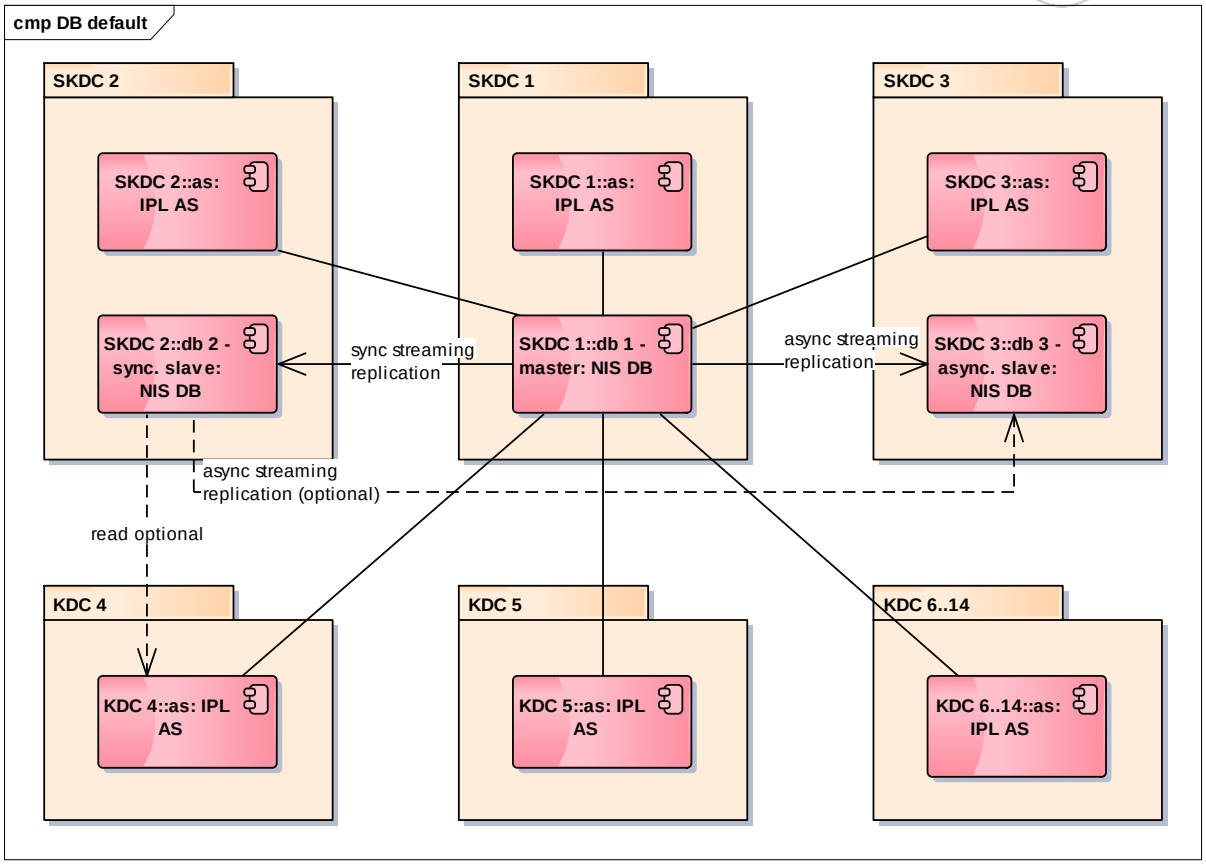
- Synchronizace master/slave je dosaženo pomocí Shared File Systém master-slave konfigurace. Každé datové centrum obsahuje vysoce dostupný SAN disk, který slouží jako sdílené úložiště.
- MB jednotlivých krajů nejsou vzájemně propojené. Výhodou tohoto přístupu je jednoduchá konfigurace a vysoký výkon zejména při zpracování zpráv uvnitř vlastního kraje. Nevýhodou pak to, že v případě výpadku celého KDC/SKDC (nebo konektivity do něj) jsou nedoručené zprávy nedostupné až do obnovení datového centra a v případě výpadku SAN dochází ke ztrátě nedoručených zpráv.

1.3.1.4 Databáze událostí

Databáze událostí je z hlediska ostatních komponent systému konfigurována jako transparentní vrstva s automatickým směrováním na master uzel. Je přítomná pouze v datových centrech SKDC. Jedno SKDC obsahuje master data, na dalších SKDC je udržovaná replika - první v synchronním módu a druhá v asynchronním (master – synchronní slave – asynchronní slave). V případě výpadku master uzlu dochází k automatickému prohlášení synchronního slave za mastera a provoz je směrován na něj, asynchronní slave je změněn na synchronní slave. V případě výpadku druhého uzlu se stává z posledního slave master. V případě rozpůlení ITS (split brain) může dojít i k automatickému prohlášení dvou uzlů jako master. Zpětné připojení databáze do clusteru po výpadku probíhá manuálně dosynchronizováním dat a nastavením konfigurace.

Na každém ze všech databázových serverů (SKDC1, SKDC2, SKDC3) bude instalována vrstva zajišťující pacemaker/heartbeat/health-check a vrstva zajišťující failover v případě výpadku jakéhokoliv nodu. Při detekci chybového stavu bude příslušně upravena konfigurace replikací tak, aby pro aplikační servery byl dostupný jeden master server, přijímající DDL a DML příkazy. Pacemaker vrstva periodicky provádí prostřednictvím resource agenta health-check všech databázových serverů. Heartbeat vrstva zajišťuje vzájemnou komunikaci mezi servery. Při vyhodnocení chybového stavu jsou spuštěny failover aktivity v závislosti na konkrétním stavu běhu databázových serverů a jsou provedeny příslušné rekonfigurace.

Pro případ roztržení sítě ITS, kdy dojde ke split-brain DB serverů a v celé topologii bude více než jeden master server bude zajištěno zamezení komunikace mezi servery v oddělených částech sítě tak, aby po opětovném spojení sítě ITS nedošlo k odstavení žádného z masterů. Řešení takové situace bude prováděno vždy ručně.



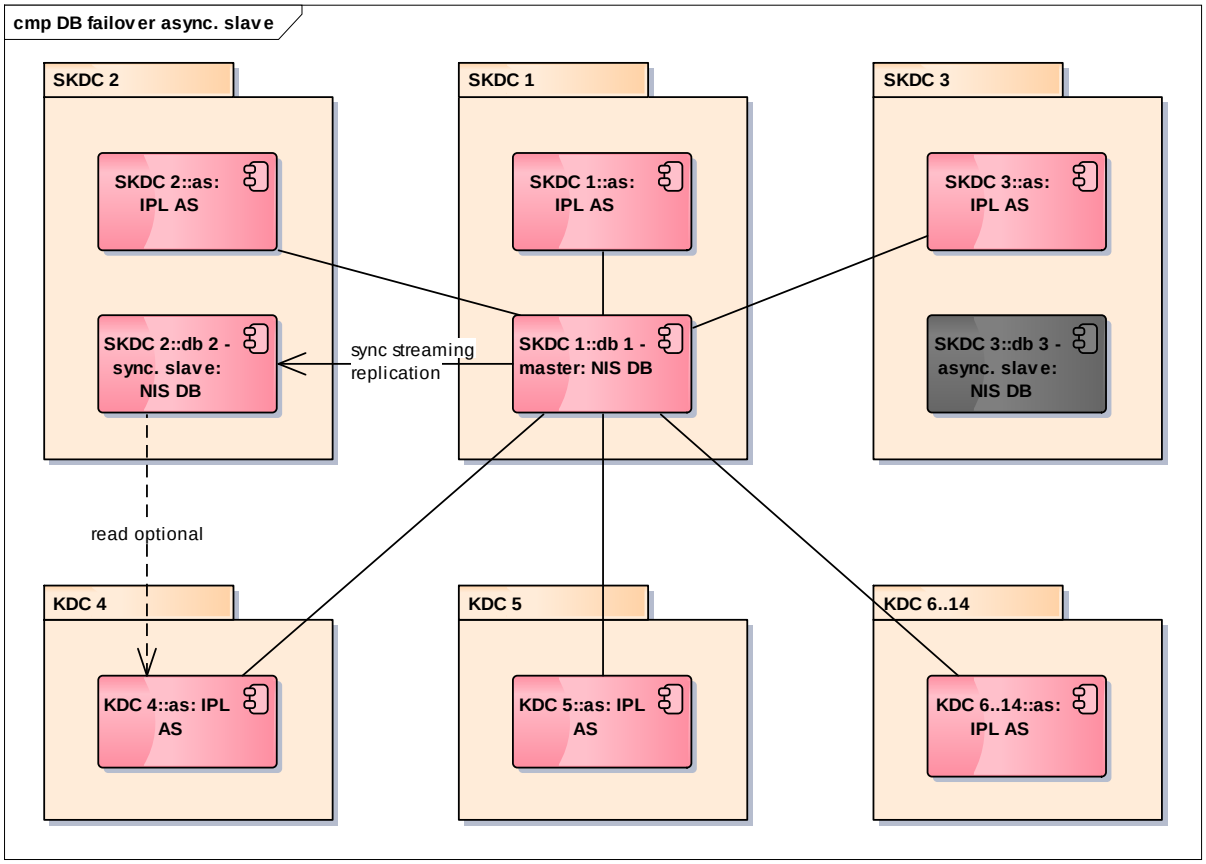
Obrázek 10 Výchozí konfigurace databázové vrstvy IPL

- SKDC1 (dále db1) – master
- SKDC2 (dále db2) – synchronní slave
- SKDC3 (dále db3) – asynchronní slave

Následuje popis všech potencionálních krizových scénářů, které mohou při výpadku databáze na SKDC nastat. Všechny scénáře se sestávají ze dvou fází: První plně automatické, která nastává těsně po výpadku, tak, aby nebyl narušen provoz systému. Druhé návratové, po opravě, která může v některých scénářích vyžadovat manuální zásah administrátora.

1.3.1.4.1 Výpadek jednoho se serverů za běhu všech 3 SKDC

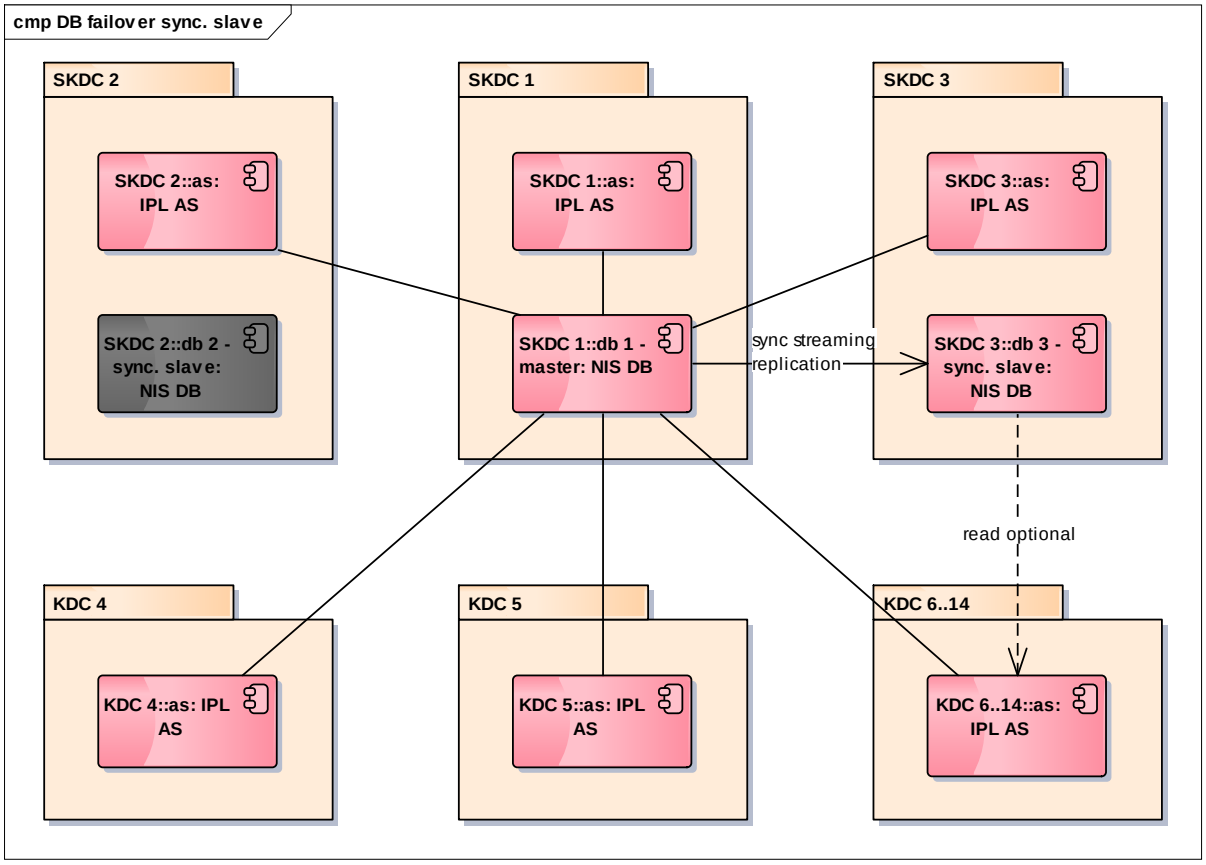
1.3.1.4.1.1 Výpadek db3 – asynchronního slave



Obrázek 11 Výpadek db3 - asynchronního slave

- V případě že resource agent vrstva vyhodnotí výpadek db3, Nejsou prováděny žádné failover aktivity. Na funkčních serverech bude nastavena blokace komunikace s tímto serverem.
- Aplikační servery zapisují nadále přes master node, volitelně mohou pro čtení využívat i synchronní slave nod.
- Databázová vrstva jako celek je nadále plně funkční bez vlivu na zbytek systému.
- Po obnovení databáze db3 je tato zapojena do databázové vrstvy nadále jako asynchronní node, dosynchronizace jejich dat proběhne automaticky.

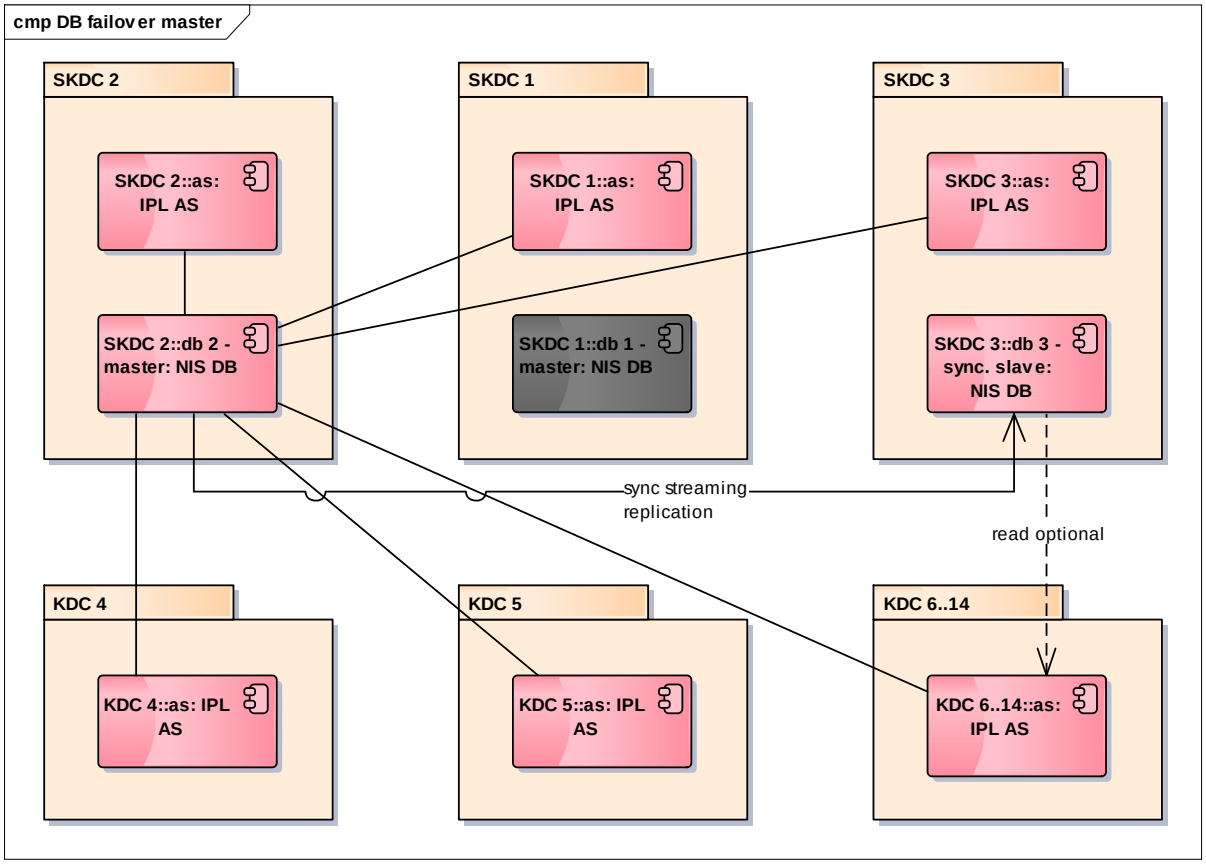
1.3.1.4.1.2 Výpadek db2 – synchronního slave



Obrázek 12 Výpadek db2 - synchronního slave

- V případě že resource agent vyhodnotí výpadek db2, failover aktivita provede změnu replikačního režimu serveru db3 a nastaví režim replikace jako synchronní. Na funkčních serverech bude nastavena blokace komunikace s tímto serverem.
- Aplikační servery zapisují nadále přes master node, volitelně mohou pro čtení využívat po dosynchronizaci i nový synchronní slave node (db3).
- Databázová vrstva jako celek je nadále plně funkční bez vlivu na zbytek systému.
- Po obnovení databáze db2 je tato zapojena do databázové vrstvy nejprve jako asynchronní node. Teprve po jejich dosynchronizaci jejich dat je možné manuálně přepnout konfiguraci do výchozího stavu.

1.3.1.4.1.3 Výpadek db1 – master serveru



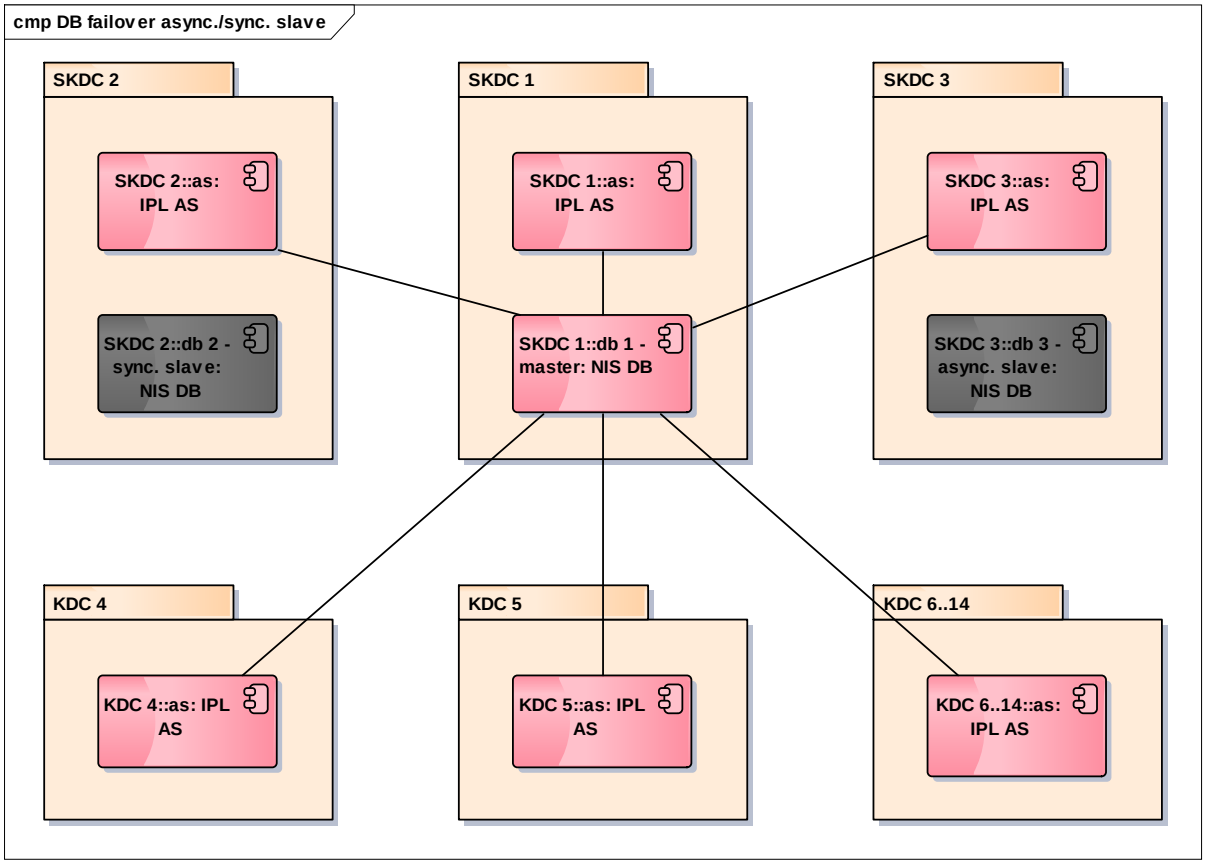
Obrázek 13 Výpadek db1 - master nodu DB vrstvy

- V případě že resource agent vyhodnotí výpadek db1, failover aktivita provede povýšení db2 (synchronního slave) na master. Dále provede test připojení db3 do asynchronní stream replikace s db2. Nebude-li toto možné z důvodu rozdílných timeline serverů, bude provedeno na db3 PIT recovery a následné připojení do stream replikace. Pokud připojení do stream replikace proběhne v pořádku, bude změněn režim replikace na synchronní. Na funkčních serverech bude nastavena blokace komunikace s tímto serverem.
- Aplikační servery zapisují přes nový master node (db2, původní synchronní slave). K přesměrování zápisů přes nový master nod dojde automaticky pomocí využití nástroje Pgpool. Nový synchronní slave (db3) lze po dosynchronizaci využívat pro čtení.
- Databázová vrstva jako celek je nadále plně funkční.
- Po obnovení databáze db1 je tato zapojena do databázové vrstvy nejprve jako asynchronní node. Teprve po dosynchronizaci jejich dat je možné manuálně přepnout konfiguraci do výchozího stavu.

1.3.1.4.2 Výpadek dvou serverů za běhu všech 3 SKDC

Představuje současný výpadek dvou aplikačních serverů okamžitý, nebo v těsném časové sledu (v čase do 100 ms), kdy nemusí v extrémním případě dojít k dosynchronizaci aktuálního asynchronního slave před výpadkem.

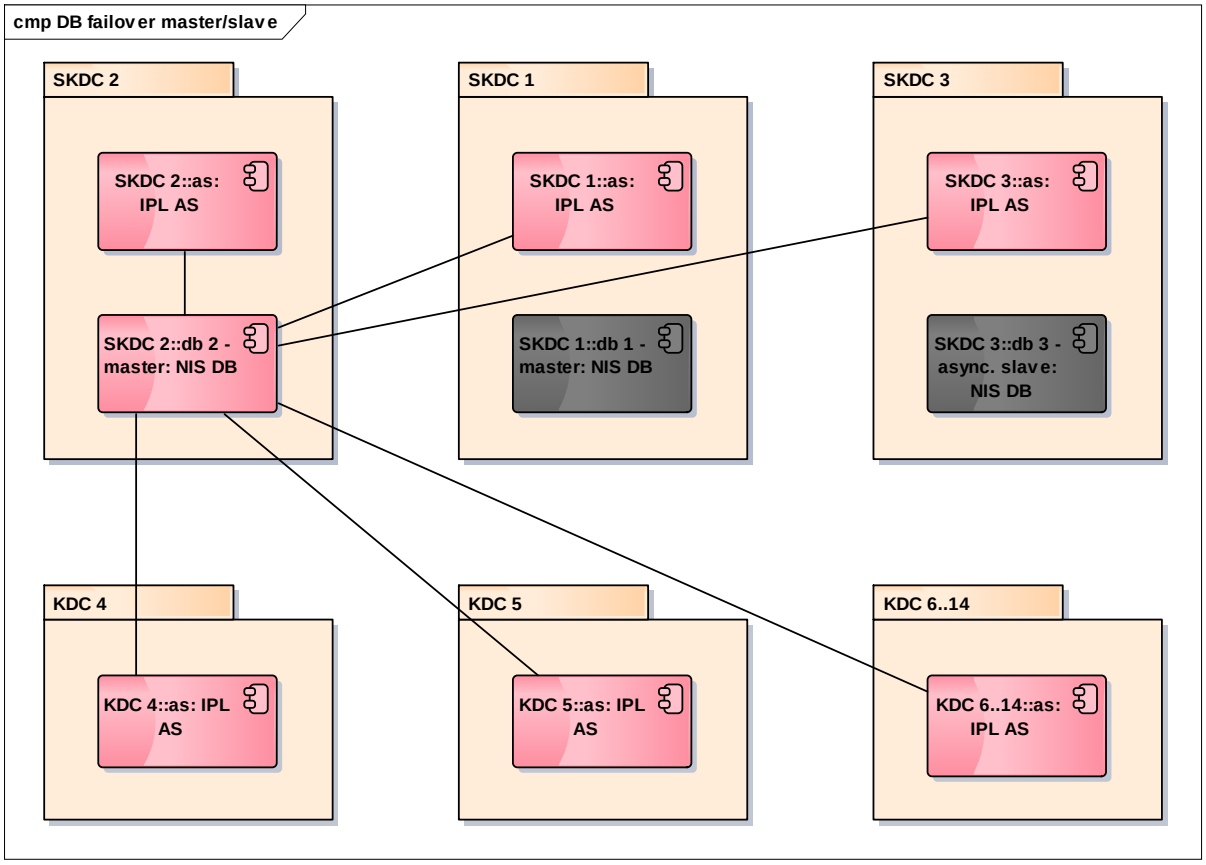
1.3.1.4.2.1 Výpadek 2 serverů, master zůstává funkční



Obrázek 14 Současný výpadek obou slave databází

- V případě že resource agent detekuje výpadek dvou serverů ale master zůstává funkční, failover aktivita provede rekonfiguraci master serveru a zastavení replikací. Na funkčním serveru bude nastavena blokace komunikace s ostatními servery.
- Aplikační servery zapisují a čtou nadále pouze přes master node.
- Databázová vrstva jako celek je nadále plně funkční bez vlivu na zbytek systému.
- Po obnovení databází db2 a db3 jsou tyto zapojeny do databázové vrstvy obě nejprve jako asynchronní node. Teprve po jejich dosynchronizaci jejich dat je možné manuálně přepnout konfiguraci do výchozího stavu.

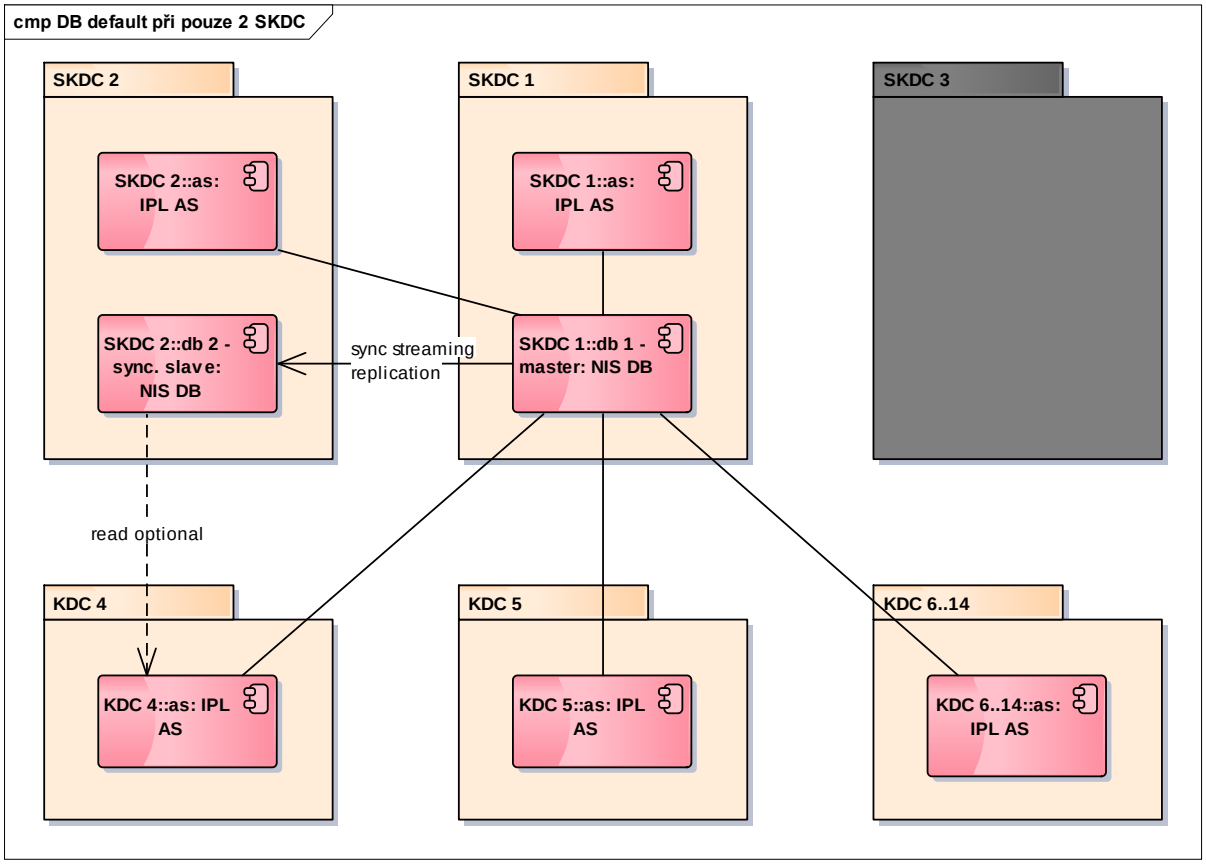
1.3.1.4.2.2 Výpadek 2 serverů, master je nefunkční



Obrázek 15 Současný výpadek jedné master i slave databáze

- V případě že resource agent detekuje výpadek dvou serverů a není dostupný žádný master server, failover aktivita provede povýšení posledního běžícího serveru na master. Dále se neprovádí žádné nastavení replikací. Na funkčním serveru bude nastavena blokace komunikace s ostatními servery.
- Aplikační servery zapisují a čtou nadále pouze přes nový master node. K přesměrování zápisů přes nový master nod dojde automaticky pomocí využití nástroje Pgpool.
- V extrémním případě, kdy se novým master nodem stala původně asynchronní replika (db3) může dojít k situaci, že budou nedosynchronizovaná data (z posledních max. 100 ms) do obnovení ostatních databází nedostupná.
- Databázová vrstva jako celek je nadále plně funkční.
- Po obnovení vypadlých databází je nutné provést před jejich zapojením do systému jejich dosynchronizaci která může v některých případech vyžadovat manuální zásah administrátora.

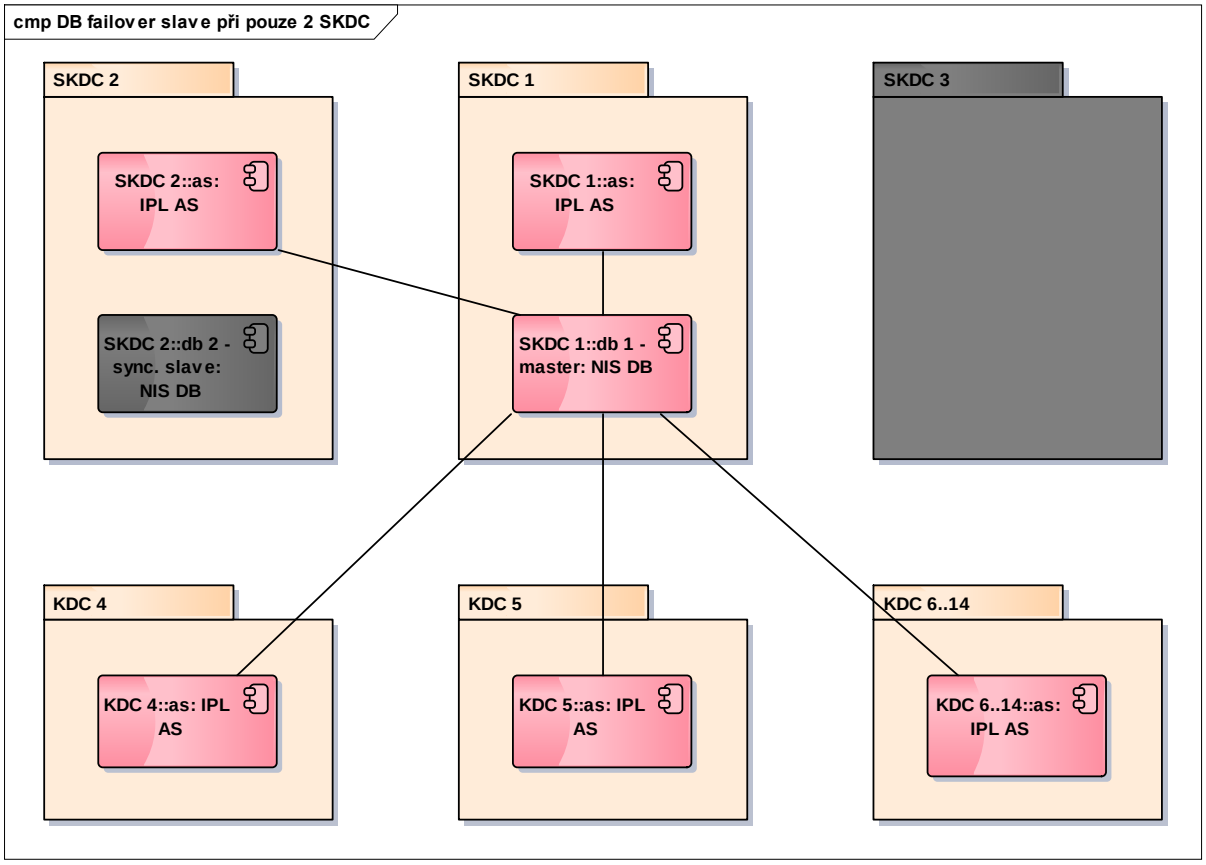
1.3.1.4.3 Výpadek jednoho ze serverů za běhu pouze dvou SKDC



Obrázek 16 Výchozí stav DB při běhu pouze dvou SKDC

Představuje situaci postupného výpadku dvou serverů databázové vrstvy v delším časovém intervalu (po dokončení scénáře výpadku serveru prvního), nebo při nedostupnosti jednoho SKDC, např. po rozdělení sítě. Při běhu pouze dvou SKDC je vždy režim replikace mezi těmito servery synchronní.

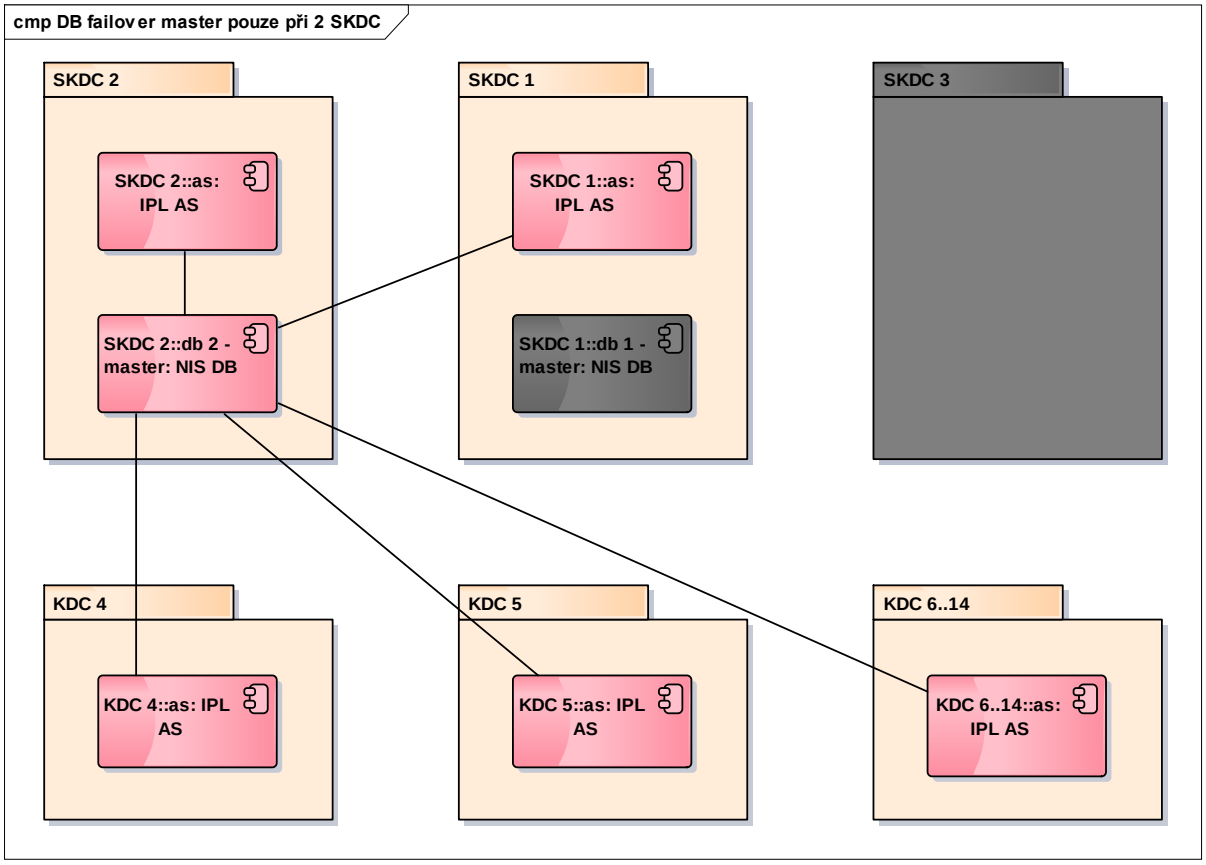
1.3.1.4.3.1 Výpadek slave serveru



Obrázek 17 Výpadek slave databáze při běhu pouze 2 SKDC

- V případě že resource agent vyhodnotí výpadek slave serveru, failover aktivita provede rekonfiguraci master serveru a zastavení replikací. Na funkčním serveru bude nastavena blokace se serverem, který havaroval.
- Aplikační servery zapisují a čtou nadále pouze přes master node.
- Databázová vrstva jako celek je nadále plně funkční bez vlivu na zbytek systému.
- Po obnovení vypadlé databáze je tato zapojena do databázové vrstvy nejprve jako asynchronní node. Teprve po dosynchronizaci jejich dat je možné manuálně přepnout konfiguraci do stavu před výpadkem.

1.3.1.4.3.2 Výpadek master serveru



Obrázek 18 Výpadek master databáze při běhu pouze 2 SKDC

- V případě že resource agent vyhodnotí výpadek master serveru, failover aktivita provede povýšení slave serveru na master. Dále se neprovádí žádný pokus o nastavení replikací. Na funkčním serveru bude nastavena blokace se serverem, který havaroval.
- Aplikační servery zapisují a čtou nadále pouze přes nový master node. K přesměrování zápisů přes nový master nod dojde automaticky pomocí využití nástroje Pgpool.
- Databázová vrstva jako celek je nadále plně funkční.
- Po obnovení vypadlé databáze je tato zapojena do databázové vrstvy nejprve jako asynchronní node. Teprve po dosynchronizaci jejich dat je možné manuálně přepnout konfiguraci do stavu před výpadkem.

1.3.1.4.4 Výpadek všech tří DB serverů – celé databázové vrstvy

V případě výpadku všech serverů databáze událostí nejsou služby SKDC z hlediska IPL dostupné a všechny kraje přechází do ostrovního režimu, včetně aplikačních serverů IPL ve vlastních SKDC.

1.3.1.4.5 Rozdělení sítě

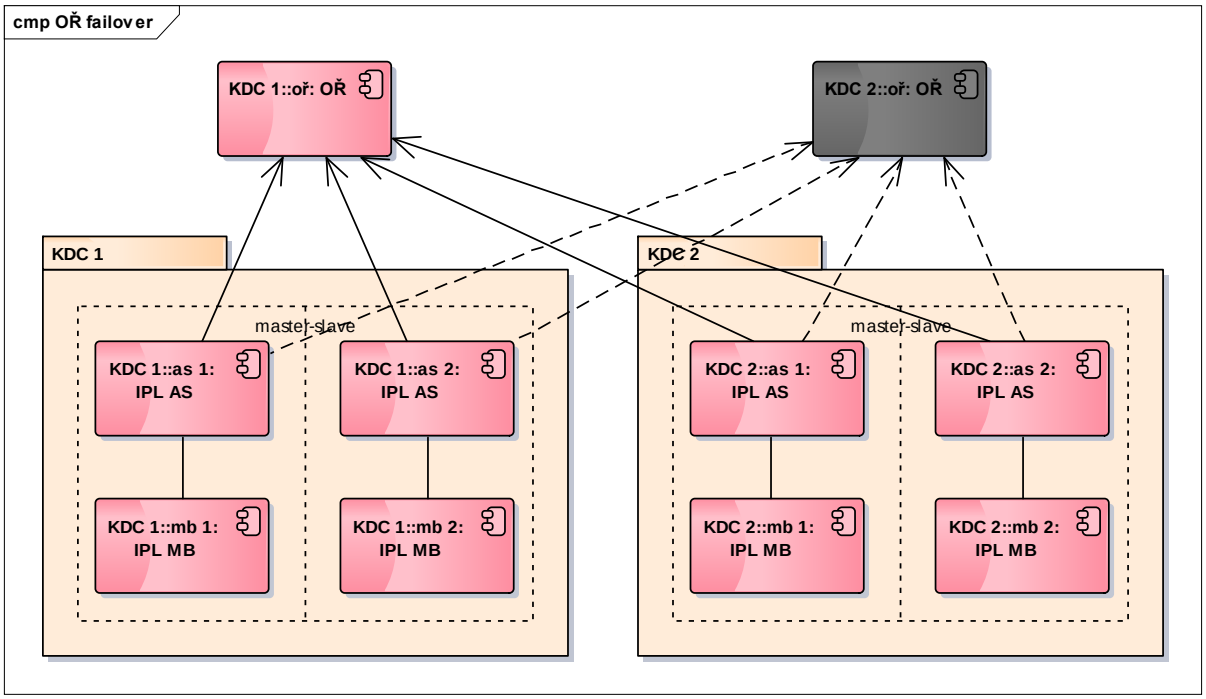
Popisuje chování databázové vrstvy v případě rozdělení sítě ITS na 2, či více nezávislých samostatně funkčních segmentů. Může nastat jedna z těchto situací:

- V nově vzniklém segmentu sítě zůstávají přítomna všechny SKDC. Databázová vrstva zůstává plně funkční ve výchozí konfiguraci. Případná KDC, která se ocitla mimo tento segment, přechází do ostrovního režimu.

- V nově vzniklém segmentu sítě se nalézají dvě SKDC. Databázová vrstva tohoto segmentu se pak chová, jako by došlo k výpadku DB serveru třetího SKDC. Následný stav odpovídá běhu pouze dvou SKDC.
- V nově vzniklém segmentu zůstane pouze jedno SKDC. Databázový server v tomto segmentu se stává automaticky master nodem a chová se tak, jako by došlo k výpadku ostatních dvou serverů.
- V nově vzniklém segmentu se nenalézá ani jedno SKDC. Případná funkční KDC v tomto segmentu pak přecházejí do ostrovního režimu.

Ve všech zmíněných situacích může dojít k situaci, že je více databází povýšeno na master nod. Po opětovném spojení ITS dojde k manuálnímu odstavení jednoho z masterů. Následuje manuálně spuštěná procedura dosynchronizace dat do hlavní databáze včetně možnosti administrátorského zásahu a řešení konfliktů. Po dobu synchronizace jsou události z odstavené databáze nedostupné.

1.3.1.5 Výpadek OŘ, nebo spojení do OŘ



Obrázek 19 Výpadek OŘ

- Každý aplikační server IPL může při normálním režimu zaslat zprávu přímo do libovolného OŘ kteréhokoli kraje.
- V případě výpadku OŘ zůstávají neodeslané zprávy do OŘ uloženy v messagingové vrstvě KDC, kde se nalézá aplikační server, který zprávu pro OŘ vytvořil. Zprávy jsou fyzicky uloženy na vysoce dostupném SAN disku, v úložišti sdíleném mezi oběma MB nody.
- Po obnovení spojení do OŘ jsou mu dodatečně odeslány všechny čekající zprávy pro toto OŘ ze všech KDC.
- Při tomto dodatečném odeslání není systém NIS schopen zachovat časové pořadí vzniku zpráv zasílaných z více KDC.

1.3.1.6 Ostrovní režim

Pojmem „ostrovní režim“ je označována situace kdy běžné KDC ztratí fyzické (páteřní) datové spojení do sítě ITS, ocitá se izolováno, a nemá dostupné žádné SKDC, obvykle ani KDC ostatních krajů. OŘ složek vlastního kraje, který se ocitl v „ostrovním režimu“ zůstávají do postiženého KDC připojeny, neboť jsou sice z logického hlediska součástí sítě ITS, nalézají se však v lokalitě shodné, nebo blízké ku KDC, a jsou připojeny odlišným fyzickým datovým spojením.

Důsledkem „ostrovního režimu“ je nedostupnost databáze událostí pro zpracovatelskou aplikační logiku v postiženém kraji, a nedostupnost OŘ ostatních krajů. IPL postiženého kraje funguje pouze jako bezestavový „přeposílač“ zpráv mezi vlastními OŘ s využitím příslušných transformací a zajišťující privátnost dat během sdílení.

Detailní popis chování KDC/SKDC v ostrovním režimu a jeho komunikace s OŘ je popsán v části B (software) tohoto prováděcího konceptu. Shodného chování, jako v ostrovním režimu dojde i v případě kompletního výpadku všech serverů databázové vrstvy (na všech SKDC).

1.3.1.7 Administrační aplikace

Administrátorská konzole je klasickou webovou aplikací nasazenou na tři aplikační servery IPL, na každém SKDC jeden.

- K přístupu k administrátorské aplikaci lze využívat libovolného load balanceru všech tří SKDC, v rámci jednoho sezení (bez nového přihlášení) však není možné LB střídat.
- Každý server obsahuje stav připojených klientů (stav klientské session), který ale není vzájemně mezi servery replikován.
- Po prvním připojení klienta je udržované připojení na zvolený server pomocí LB a "sticky session", který volí server podle hlavičky "JSESSIONID" v rámci HTTP requestu.
- V případě nedostupnosti aplikačního serveru IPL, je dotaz load balancerem přesměrován na aplikační server dalšího SKDC až do celkového počtu všech tří SKDC.
- V případě přesměrování na nové SKDC je vyžadováno nové ověření identity uživatele – nové přihlášení.

1.3.2 Spolehlivost GIS služeb

Spolehlivost služeb GIS je dostatečně rozpracována v části B-SW, zejména poté v bodech 2.2.3 a 2.2.4.

2 Přílohy

2.1 Seznam obrázků