

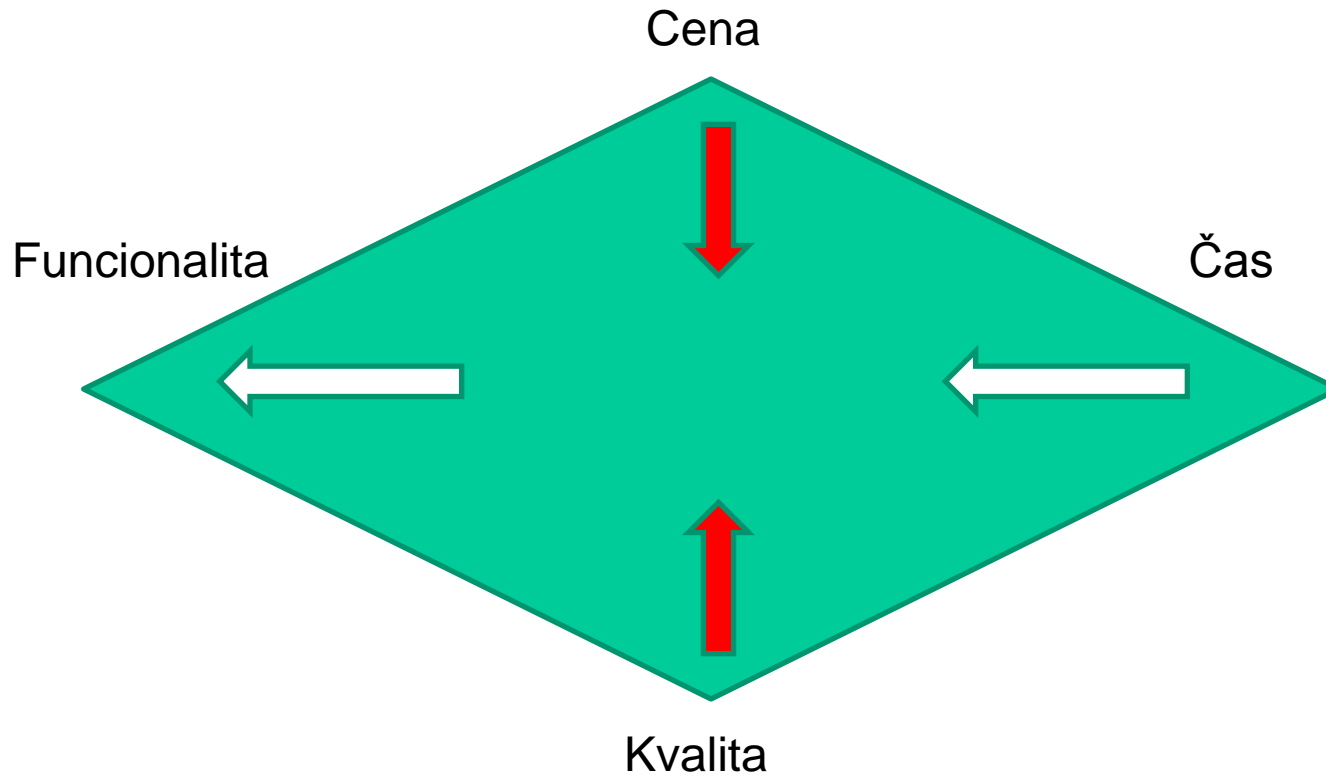
Řešení Oracle pro ukládání mnoha dat

Jak zvládnout uložení
velkého počtu dokumentů

”Co
po
nás
zbude



4 parametry SW řešení



”Co
po
nás
zbude



Co je to kvalita SW?

Metrika	Význam
Spolehlivost (REL)	Míra rizika, že řešení nebude schopno dodat požadovanou funkcionalitu vůbec a nebo jen v omezené míře.
Výkon (PERF)	Schopnost řešení dodat požadovanou funkcionalitu očekávanému počtu (současně přístupujících) uživatelů s očekávanou rychlostí odezvy.
Bezpečnost (SEC)	Míra rizika vyplývající ze zneužití informací, se kterými řešení pracuje.
Udržovatelnost (MAIN)	Schopnost aplikace adaptovat se na nové, budoucí požadavky, či schopnost být přenesen na jiné prostředí (HW, operační systém, atp.)
Uživatelská přívětivost (USB)	
Testovatelnost (TST)	

”Co
po
nás
zbude



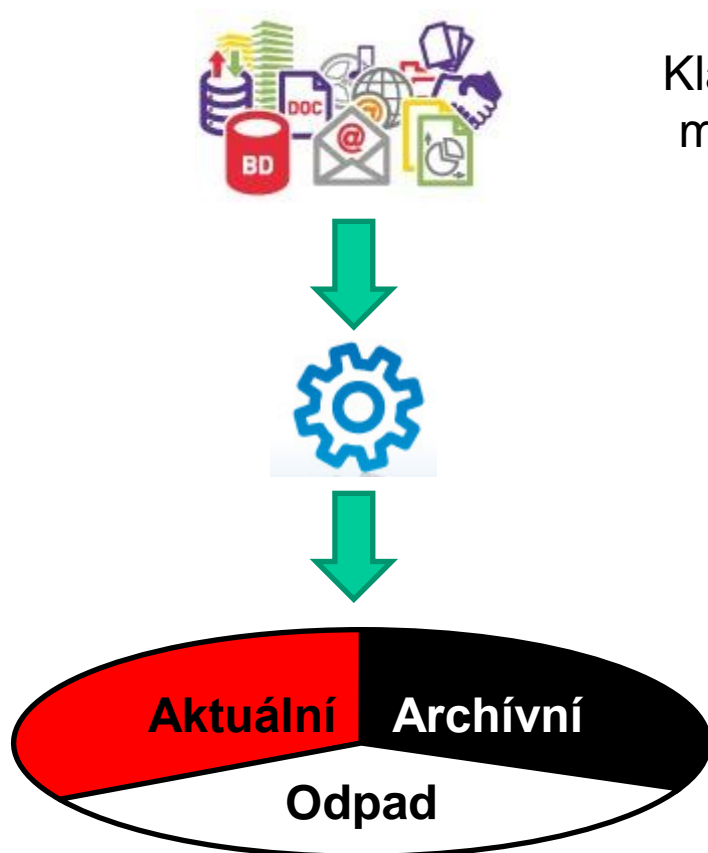
Metriky

Metrika	Stav při spuštění	Dnes	Vývoj do budoucna
Počet dokumentů	Hodně	Denně 150 000 Celkem 54 mil.	Do 20 let až 6 miliard
Průměrná velikost dokumentu	stovky kB	250 kB	stovky kB
Velikost dat	-	31 TB	Jednotky PB
Počet uživatelů	Jednotky na každé z cca 4 000 org.	Desítky na každé z cca 4 000 org.	Potenciálně celá ČR (účastníci sporů), minimálně však všichni pracovníci justice
Zabezpečení	per organizace	per organizace	per organizace + výjimky
Vysoká dostupnost	Ne	Ne	Možná
Disaster Recovery	Backup/restore	Backup/restore	Možná záložní lokalita
HW/Platformní nezávislost	Ne	Ne	Ne
Uživatelská přívětivost	Přístup z prostředí existujících aplikací	Přístup z prostředí existujících aplikací	Přístup z aplikací i napřímo (web browser či MS Office)

”**Co
po
nás
zbude**“



ILM



”Co
po
nás
zbude



Kompresa



”Co
po
nás zbude



Operace vs. objem

- vyhledávání (rychlost odezvy)
 - při užití fulltextu
- zabezpečení
- spolehlivost
- cena za uložení

”Co
po
nás
zbude



Vyhledávání

- přístup k záznamu
- hledání (dotaz)

Tvorba indexu (zejm. fulltextu)

- index zabírá místo
 - Fulltextový index zabírá 1/3 obsahu v textové podobě
- po přidání položky je třeba jej doplnit
 - Cena operace závisí na
 - velikosti souboru
 - počtu položek

”
Co
po
nás
zbude



Vyhledávání

Prezentační vrstva
(přijetí požadavku)
Web Server

Logická vrstva
(business logika)
Aplikační Server

Datová vrstva
(data, indexy)
Databáze, Úložiště



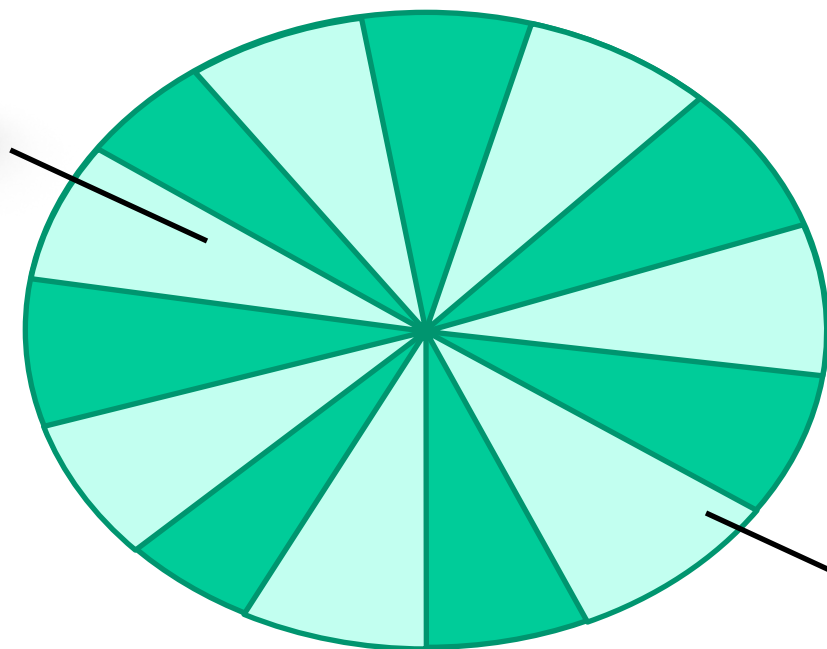
Oracle
Secure
Enterprise
Search

Oracle
Text (DB)

”
Co
po
nás
zbude



Zabezpečení



”Co
po
nás
zbude



Zabezpečení

- v první fázi deklaratorní (metadata)
 - uživatel „vidí“ jen data své organizace
 - side-effect: urychluje vyhledávání
- v druhé fázi je třeba rozšířit
 - spíše než v „organizaci“ se pracuje v „týmu“ (senát)
 - někdy je třeba vidět i „mimo tým“
 - předání spisu na jiný soud
 - nahlížení do spisu (účastníci sporu)
- výjimky mohou být per dokument
 - je třeba myslet na spravovatelnost

”Co
po
nás
zbude



Spolehlivost

- dnes řešeno jako backup/recovery
 - vloni (cca 14TB) výpadek – plné obnovení by trvalo **14 dní**
 - použita metoda *partitioningu* (obnova nejnovějších)
 - do budoucna neudržitelné

”Co
po
nás
zbude



Spolehlivost

- další varianty
 - redundantní ukládání dat (ZFS)
 - záložní lokalita
 - nákladné řešení
 - řeší výpadek na kterékoliv vrstvě, vč. spojení s lokalitou

”Co
po
nás
zbude



Cena za uložení

Základní kapacita systému	100 TB
Možnost rozšíření až na	800 TB
Maximální spotřeba systému diskového úložiště o kapacitě cca 1750 TB ⁽¹⁾	do 3,8kW 2,1kW v Standby režimu
Minimální počet operací	50 000 IOPS

⁽¹⁾ jedná se o dvojnásobnou max. kapacitu z důvodu redundantního uložení

”Co
po
nás
zbude



Co zákazník potřebuje

- ve špičce bude systém užívat asi **5 000 uživatelů**, kteří během **10-15 minut** vytvoří 1 transakci, tj. **5 – 7 dotazů** (READ) na úložiště
- bude vkládáno **10 miliónů dokumentů** (WRITE) **ročně**
- z jiné externí aplikace je generováno **35 miliónu dotazů** (READ) **ročně**
- požadovaná doba odezvy byla **4 sekundy**
- uživatelé většinou pracují jen s daty tohoto či minulého roku (**3 – 10%** celkové kapacity)

Tj. 77 čtecích a 5 operací na zápis za sekundu

”**Co
po
nás
zbude**



Snížení nákladů = pásky

- v daném případě vychází cena řešení na discích **10x vyšší** než při kombinaci disků (5%) a pásek
- pásky se „netočí“ (**žádná spotřeba energie**)
- pásky mají **delší dobu odezvy**
- v kombinaci ze SAM-QFS je možné výrazně **omezit** potřebu **záloh** (a případné obnovy)

”**Co
po
nás
zbude**



Závěr

Best-practices:

- nedovolte, aby kvalita klesla **pod přijatelnou míru**
 - stanovte si **smysluplné metriky** a užívejte je
 - **upravujte metriky** v čase dle nových potřeb a znalostí
 - snažte se anticipovat problémy
-
- **ved'te, na co vynakládáte peníze**

”
Co
po
nás
zbude



Děkuji za pozornost

Jiří Machotka

jiri.machotka@oracle.com

”
Co
po
nás
zbude

