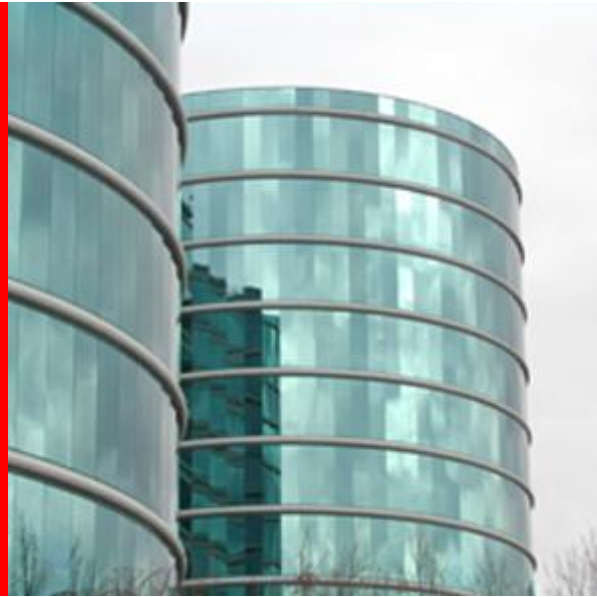


ORACLE®



ORACLE®

Aktuální trendy v oblasti datových skladů a business intelligence

Jakub ILLNER

Solution Architect, Oracle Consulting EE-CIS

Pozice BI & DWH v roce 2009

- Poselství o stavu BI & DWH
- Referenční architektura
- Rychlost, výkon, objem
- Time to Market
- Předpřipravené BI aplikace
- Důraz na datovou kvalitu
- Master Data Management
- BI a Performance Management
- Metadata driven
- Bezpečnost především
- BI & DWH jako služby
- Service Level Agreements

Poselství o stavu BI & DWH

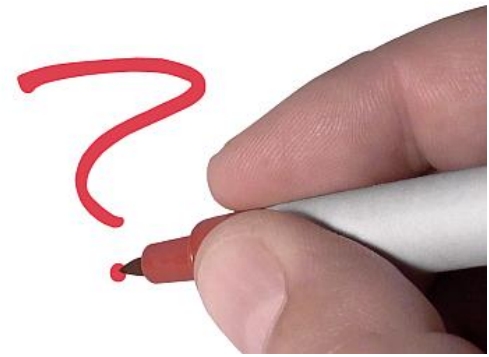
- **V roce 2009 jsou systémy business intelligence a datové sklady považovány za vyspělá a standardní řešení**
- Organizace považují BI a DWH za klíčové systémy pro růst, udržení kompetitivních výhod a řízení firmy
- Systémy BI a DWH se stávají „produkčními systémy“, podporujícími klíčové firemní procesy
- Technologie DWH jsou robustní, stabilní a výkonné
- Technologie BI umožňují širokou škálu využití analytických informací
- Metodiky budování BI a DWH jsou spolehlivé
- Je používána ověřená referenční architektura BI a DWH
- Jsou k dispozici best practices pro návrh datových modelů

Co na to říká Gartner?

Gartner Identifies the Top 10 Strategic Technologies for 2009

- Virtualization
 - Cloud Computing
 - Servers - Beyond Blades
 - Web - Oriented Architectures
 - Enterprise Mashups
 - Specialized Systems
 - Social Software and Social Networking
 - Unified Communications
 - Green IT
 - Business Intelligence (BI)
- BI, the top technology priority in Gartner's 2008 CIO survey, can have a direct positive impact on a company's business performance, dramatically improving its ability to accomplish its mission by making smarter decisions at every level of the business from corporate strategy to operational processes. BI is particularly strategic because it is directed toward business managers and knowledge workers who make up the pool of thinkers and decision makers that are tasked with running, growing and transforming the business. Tools that let these users make faster, better and more-informed decisions are particularly valuable in a difficult business environment.
- <http://www.gartner.com/it/page.jsp?id=777212>

Tak proč ještě nenasazujeme BI & DWH jako Baťa cvičky?



- Je to drahé
- Není to jednoduché
- Neumíme to (nemáme zkušenosti)
- Podceňujeme reálnou složitost
- Nerozumíme businessu
- Máme nekvalitní data
- Máme málo času
- Máme nereálná očekávání
- Nechceme investovat do trvalého rozvoje
- Nemáme podporu organizace

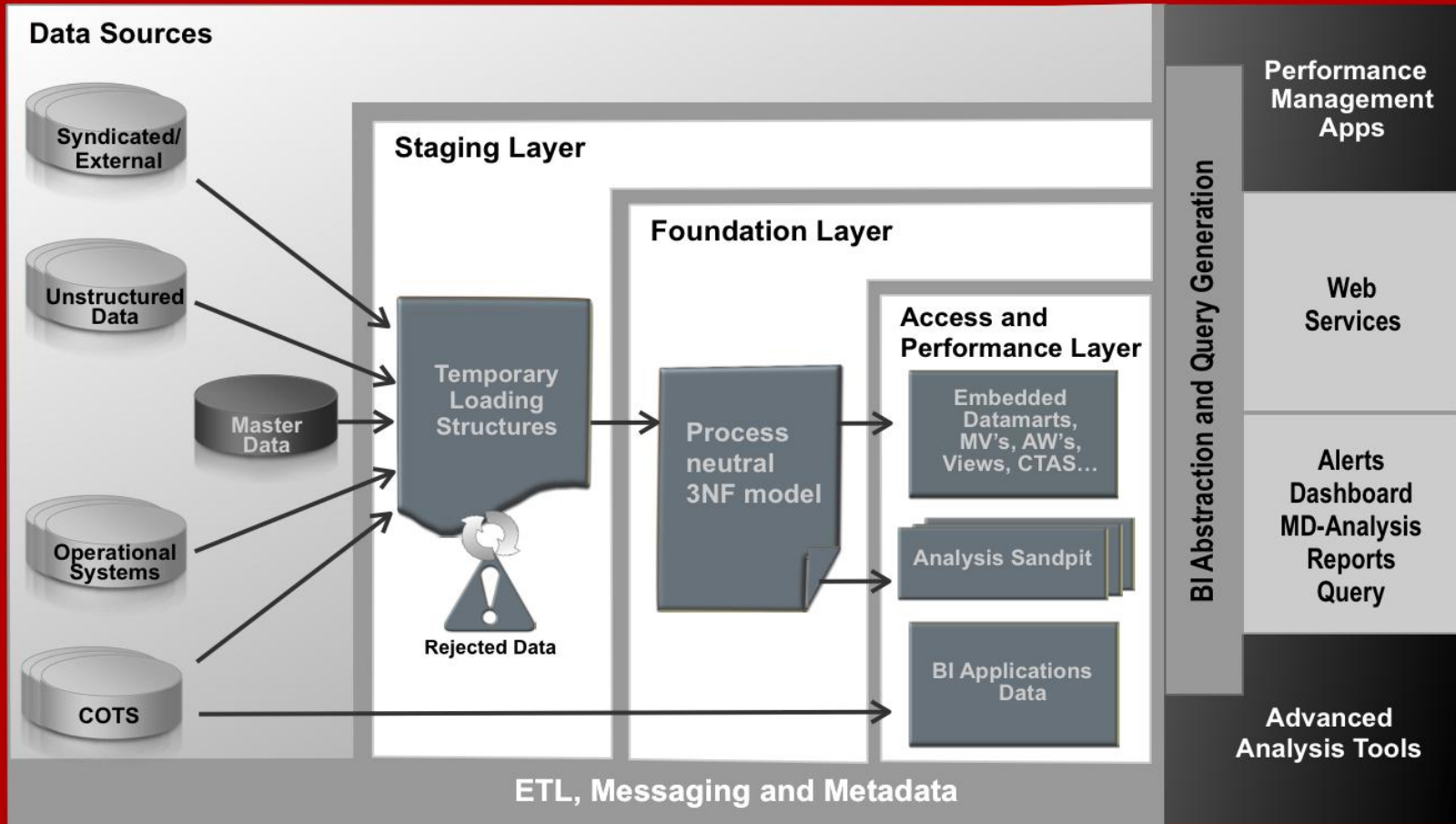
**Jak (relativně) bezbolestně
nasazovat BI & DWH ...**

**a zároveň být v souladu s
současnými trendy?**



Referenční architektura

Security



Principy referenční architektury

- Logické vrstvy datového skladu
 - Staging Layer – pracovní oblast, sloužící pro dočasné uložení a transformaci vstupních dat. Oblast umožňuje oddělit proces extrakce a přenosu dat od procesu zpracování dat a uložení v datovém skladu
 - Foundation Layer – oblast slouží pro uložení historických dat na nejnižší úrovni granularity (detailní data). Pro datový model je zpravidla využita upravená 3NF. Oblast zodpovídá za dlouhodobou správu dat.
 - Access and Performance Layer – oblast poskytující data uživatelům a aplikacím v podobě optimalizované pro jejich potřeby. Obsahuje detailní i agregovaná data, využívá OLAP, materializované view, agregované tabulky, externí data marty apod. Oblast může být kdykoliv naplněna z Foundation Layer.

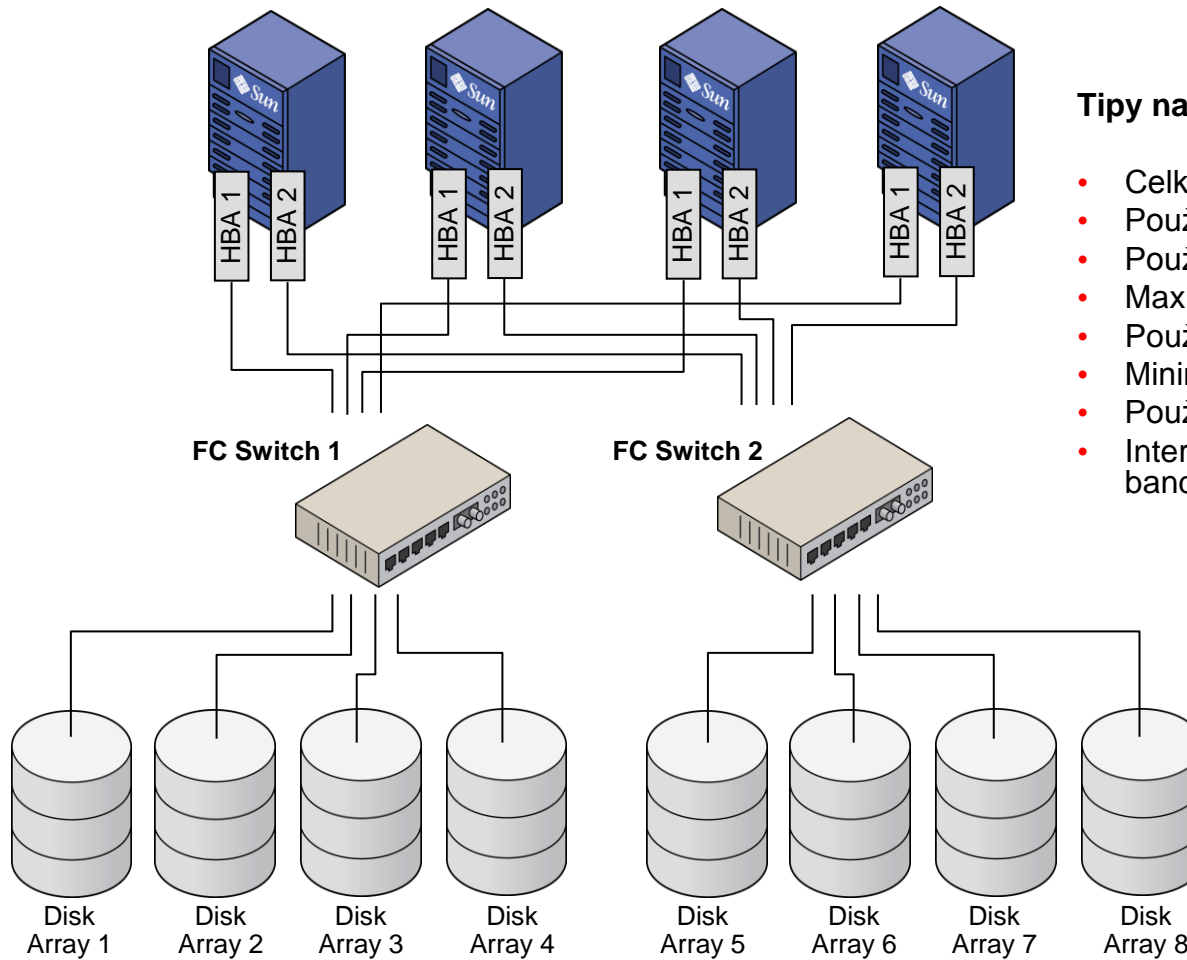
Rychlost, výkon, objem

- Objem dat v datových skladech stále roste – jednotky TB jsou běžné, desítky TB jsou obvyklé, stovky TB jsou již v produkci
- Požadavky na rychlost dotazů jsou klíčové jak pro uživatele standardních reportů, tak pro analytické uživatele pracující iterativním způsobem
- Dostupnost dat v čase D+1 je očekávána, dostupnost s častější, téměř online frekvencí je požadována pro vybrané oblasti

Jak zajistit dostatečný výkon?

- ✓ Dobře navržený logický i fyzický datový model pro oblasti Foundation a Access/Performance Layer
- ✓ Balancovaný systém optimalizovaný na celkový throughput (CPU+Memory+I/O+Disks)
- ✓ Využití „Data Warehouse Appliance“
- ✓ Výkonná databáze Oracle 11g
- ✓ Paralelizace, partitioning, komprese, ASM, RAC, indexace, materializace, resource management
- ✓ Direct load, partition exchange loading
- ✓ Využití multidimenzionální databáze (OLAP)
- ✓ Výkonný BI nástroj pro reporting, analýzy a data mining
- ✓ Výkonný ETL nástroj, zkušení vývojáři

Balancovaný systém

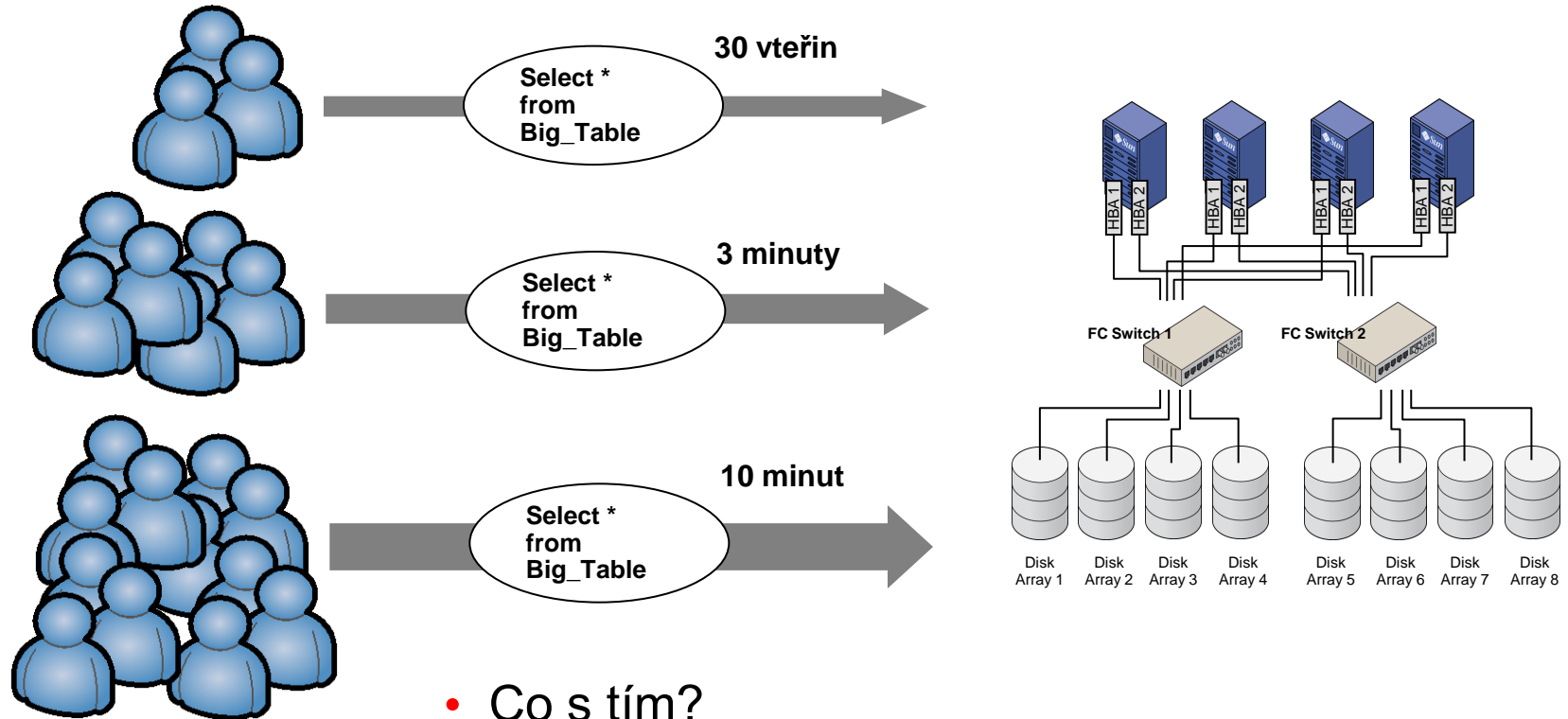


Tipy na balancovaný systém

- Celkový throughput = # core X 200MB
- Použít 1 HBA port per CPU
- Použít 1 disk controller per HBA Port
- Max 10 fyzických disků per controller
- Použít menší disky (146GB nebo 300GB)
- Minimálně 4GB of Memory per core
- Použít RAID 1 s ASM
- Interconnect bandwidth = IO subsystem bandwidth

Jen balancovaný systém nestačí

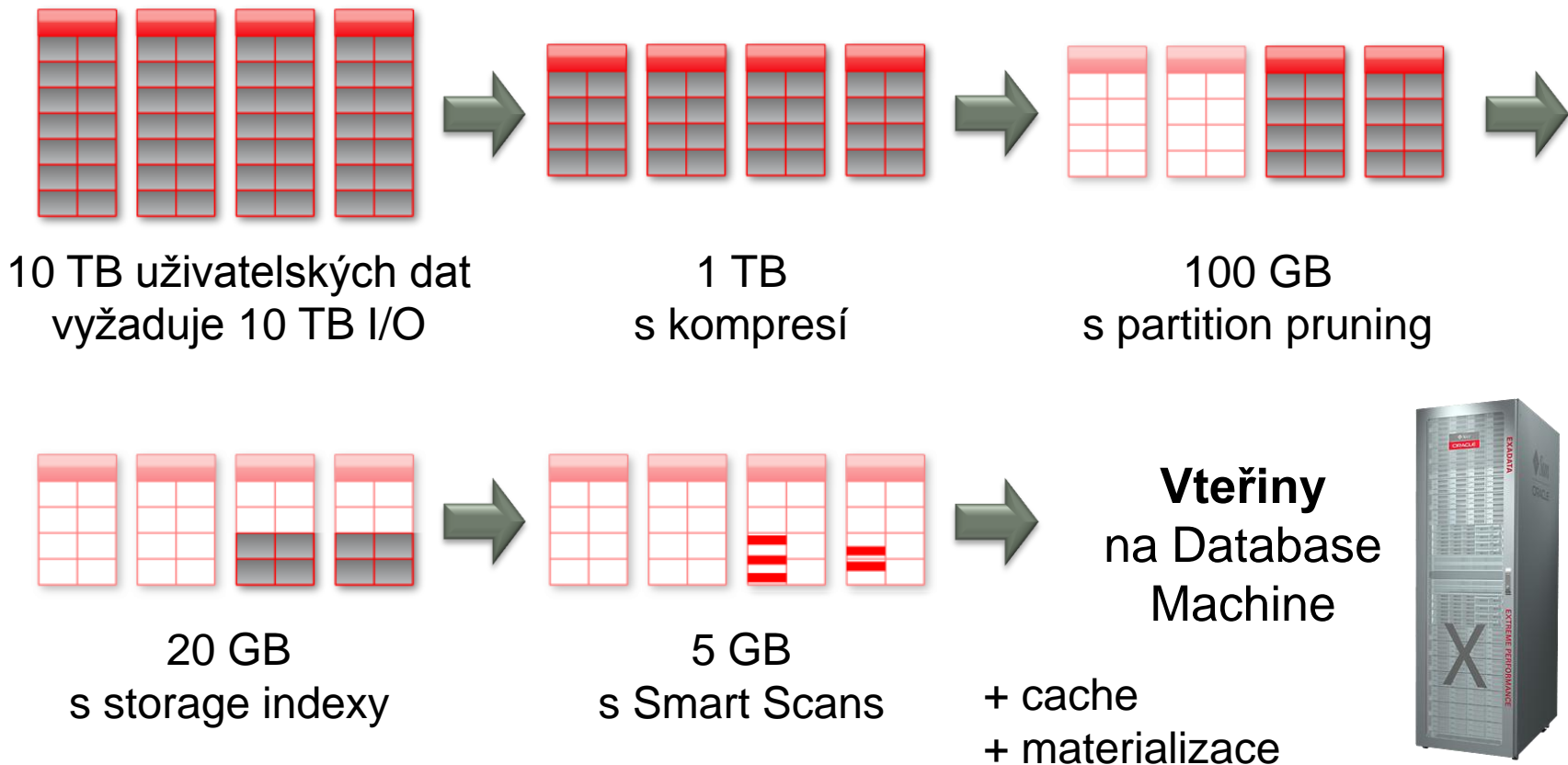
Škálovatelnost pro větší počet uživatelů



- Co s tím?
 - Serializace dotazů, resource management
 - Přidání dalších CPU, disků, switchů (při zachování balancovaného systému)
 - Použití OLAP, sumarizace, cache, komprese

Jen balancovaný systém nestačí

... aneb hrubá síla není všechno ...



Data jsou 10x menší, dotazy 2000x rychlejší ;-)

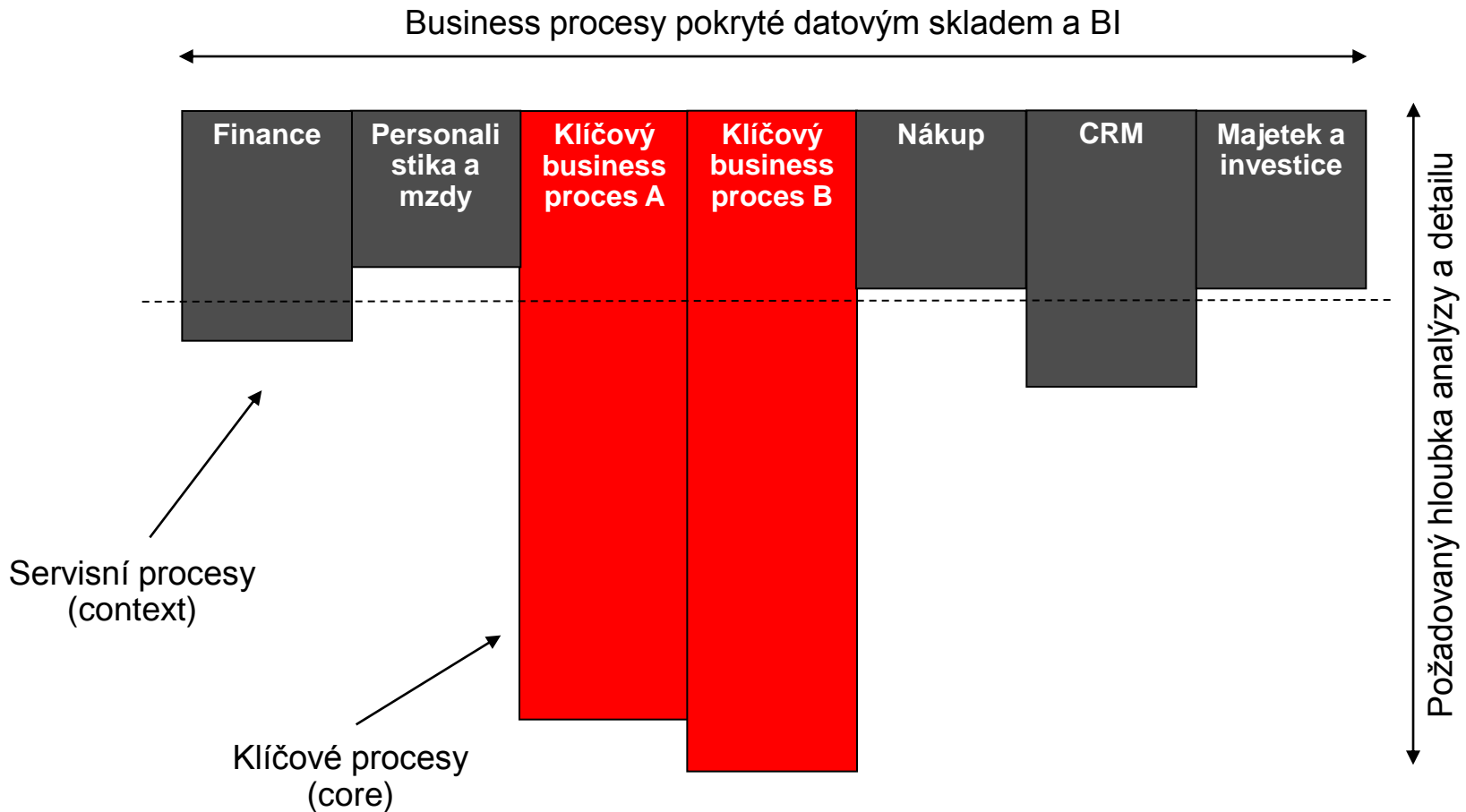
Time to Market

- Požadavky businessu se mohou rychle měnit, v závislosti na nových trendech, produktech, strategii, konkurenci. Uživatelé očekávají, že požadavky budou do datového skladu rychle zapracovány.
- Ale zapracování změn do korporátního datového skladu je složitější proces, který vyžaduje
 - Nový funding pro datový sklad, pokud je stávající budget již vyčerpán
 - Analýzu dopadů změny na další aplikace, uživatele a reporty nad datovým skladem
 - Testování kvality nových změn, ověření výkonnosti pro zapracování změn
 - Koordinaci s release cykly ostatních aplikací a prioritami datového skladu

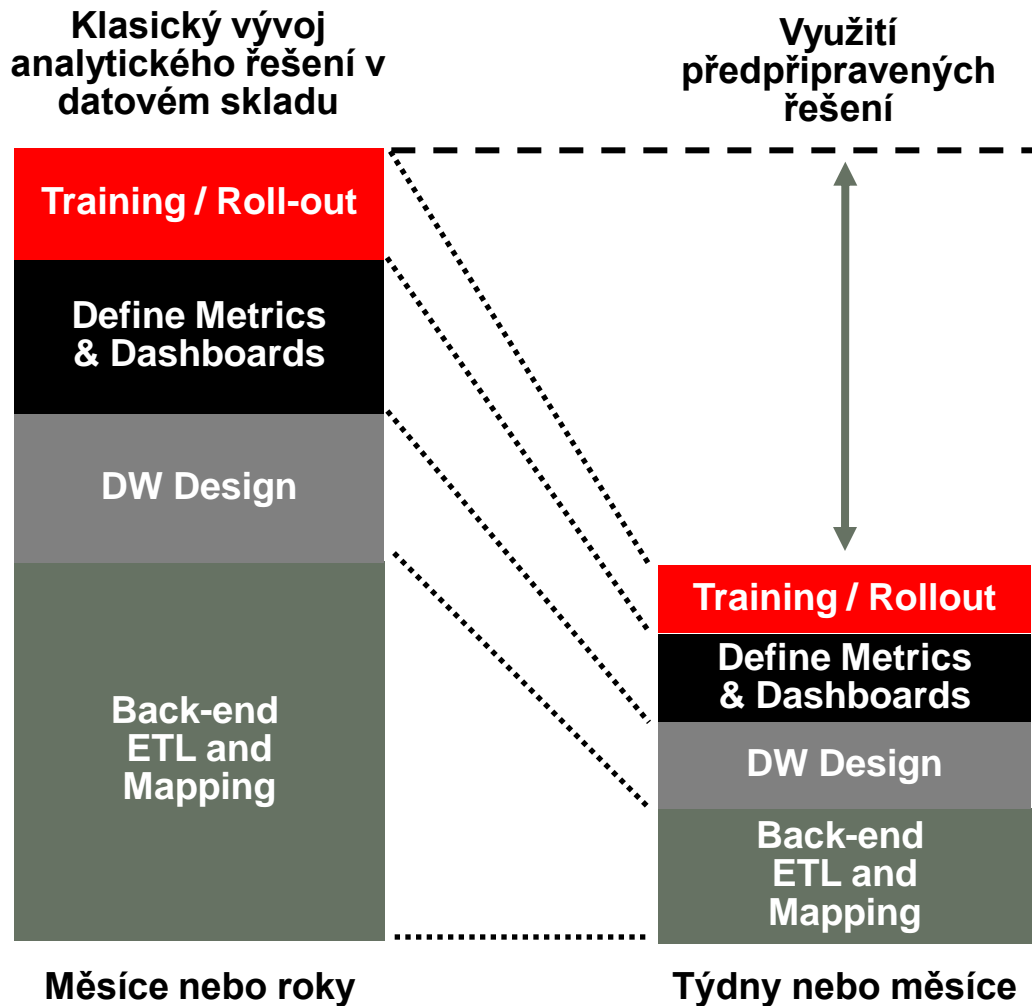
Jak urychlit nasazování změn v BI & DWH?

- ✓ Počítat s potřebou trvalého rozvoje datového skladu, mít k dispozici tým a metodiku pro řízení změn v datovém skladu ve všech fázích - definice, prioritizace, analýza, vývoj, testování
- ✓ Architektura musí podporovat změnové řízení
 - Procesně neutrální datový model
 - Otevřené interface s zdrojovými systémy
 - Analytické hřiště pro adhoc potřeby uživatelů a prototypování nových požadavků
 - Business Intelligence umožňující přístup k datům nejen v datovém skladu
 - Podpora rychlé přípravy testovacích prostředí a scénářů
- ✓ Použít předpřipravené BI aplikace pro rychlé nasazení zcela nových věcných oblastí

Opravdu chcete navrhovat datový sklad a BI pro všechny procesy ve firmě?



Předpřipravené BI aplikace



- Předpřipravené BI aplikace umožňují významně urychlit vývoj a nasazení datových skladů a business intelligence pro **standardní procesy**

- Procurement
- Financials
- HR a Payroll
- Sales
- Marketing
- Service
- Partner Management
- Contact Telephony

- Připravený datový model, KPI, reporty a dashboardy
- Včetně předdefinovaného mapování pro běžné ERP a CRM balíky (Oracle EBS, SAP, Siebel, Peoplesoft)

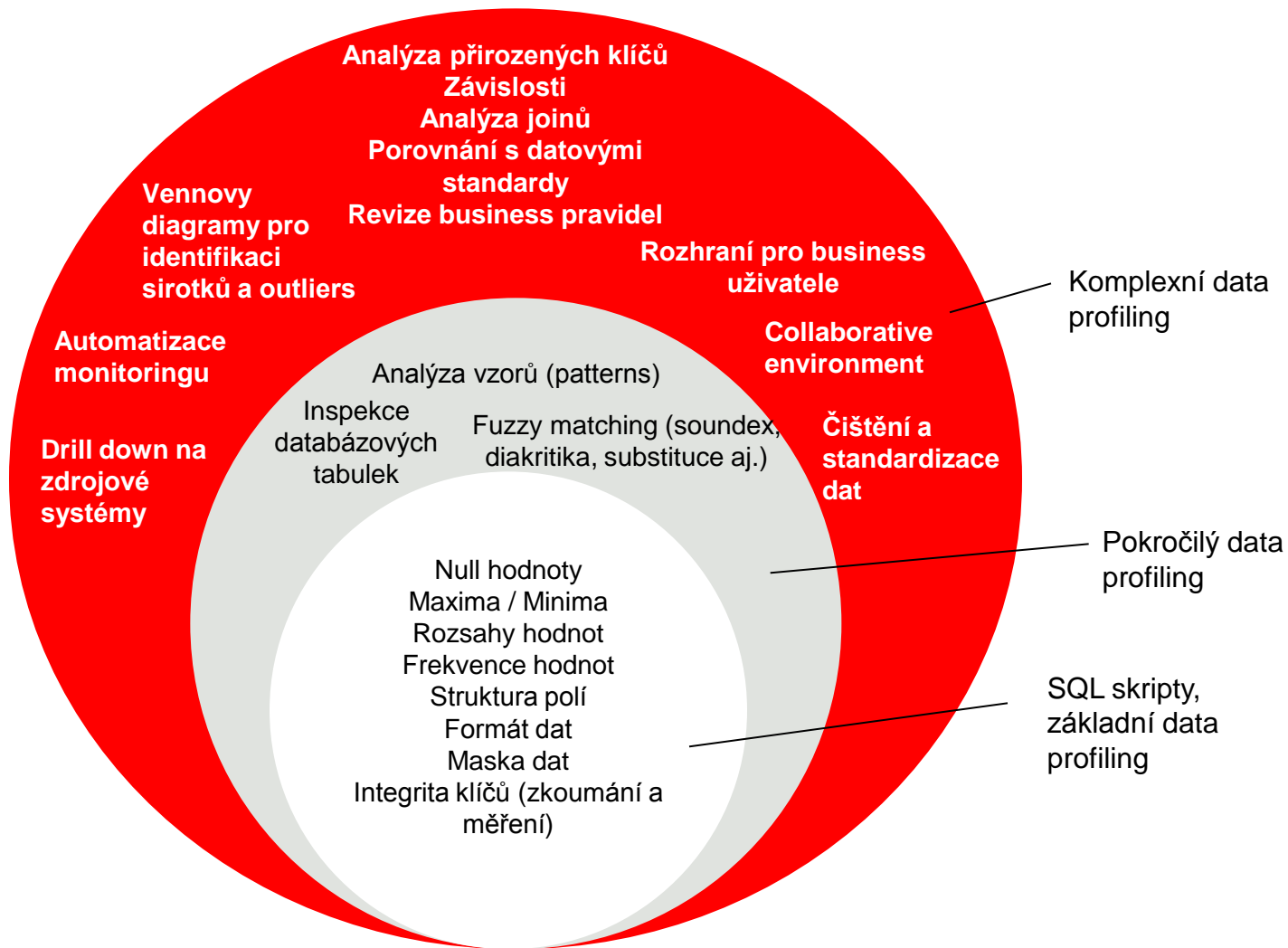
Důraz na datovou kvalitu

- Uživatelé očekávají kvalitní data v datovém skladu; pokud se domnívají, že data nejsou kvalitní, datovému skladu nevěří a nepoužívají jej. A to bez ohledu na to, zda je viníkem datový sklad nebo zdrojový systém.
- Typické problémy s kvalitou v datovém skladu
 - Závislost na kvalitě dat v zdrojovém systému, sebelepší nástroje pro data cleansing a profiling nepomohou, pokud data zcela chybí nebo obsahují systematické chyby
 - Neexistuje proces a zodpovědnosti pro celkové řešení datové kvality ve firmě, datový sklad sám o sobě může řešit jen vybrané problémy s kvalitou
 - Datová kvalita je akceptovatelná při nasazení do produkce ale postupně se zhoršuje, neboť není monitorována a chyby nejsou opravovány
 - Zdrojové systémy mění chování aniž by upozornily datový sklad nebo aniž by změna byla v datovém skladu zapracována
 - Chyby v extrakčních a ETL procesech mohou způsobit chybná nebo chybící data případně může docházet k ztrátě změn (lost updates)
 - Uživatelé mají očekávání na datovou kvalitu (např. reconciliaci s hlavní knihou), která nebyla zohledněna při definici rozsahu a designu projektu

Jak zlepšit kvalitu dat v datovém skladu?

- ✓ Uzavřít SLA s dodavatelem zdrojových dat, zaručující strukturu, kvalitu, čas dodávky a změnové řízení
- ✓ Využívat systémy pro Master Data Management
- ✓ Využívat nástroje pro kontrolu, čištění a obohacení vstupních dat v rámci ETL procesu
- ✓ Provádět pravidelný Data profiling, jak na straně vstupních dat, tak na straně datového skladu
- ✓ Pravidelně rekonziliovat data mezi datovým skladem a zdrojovými systémy
- ✓ Definovat indexy kvality, monitorovat je a vyhodnocovat („Data Quality Data Mart“)
- ✓ Vyjasnit si rozsah aktivit pro datovou kvalitu na začátku projektu, v rámci definice rozsahu projektu

Možnosti analýzy kvality dat

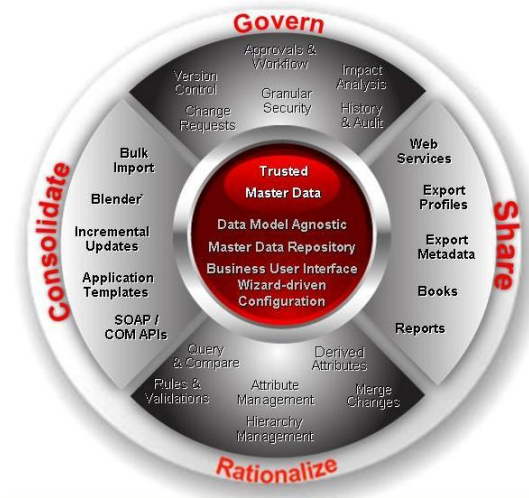


Master Data Management

- Jednou z metod zvyšování kvality je zavádění systémů pro Master Data Management, které centralizují správu klíčových entit ve firmě a umožňují jednotný pohled na data ve všech systémech
 - Zákazníci, dodavatelé, odběratelé (Customer Hub)
 - Produkty (Product Hub)
 - Lokality (Site Hub)
 - Finanční data
 - Analytické číselníky
- Je nezbytné, aby datový sklad byl úzce integrován s Master Data Management systémy a aby dimenze v skladu odpovídaly centrálně spravovaným entitám

Master Data Management

Například konsolidace účetní osnovy



Peoplesoft Balance Sheet (US GAAP)

- 100-000 – Assets
 - 100-100 – Short Term Assets
 - 100-110 – Cash
 - 100-120 – Cash Equivalent
 - 100-200 – Long Term Assets
 - 100-210 – Research
 - 100-220 – Advertising
 - 100-230 - Inventory
- 200-000 – Liabilities
- 300-000 – Equity

SAP Balance Sheet (IFRS)

- 10000 – Assets
 - 11000 – Short Term Assets
 - 11100 – Cash
 - 11200 – Cash Equivalent
 - 15000 – Long Term Assets
 - 15100 – Embedded Derivatives
 - 15300 – Inventory
- 20000 – Liabilities
- 30000 – Equity

konsolidace
mapování



správa
sdílení

Management Accounting & Performance Reporting

- 1 – Assets
 - 1.1 – Short Term Assets
 - 1.1.0 – Other Short Term Assets
 - 1.1.1 – Cash
 - 1.1.2 – Cash Equivalent
 - 1.2 – Long Term Assets
 - 1.2.0 – Other Long Term Assets
 - 1.2.1 – Research
 - 1.2.2 – Advertising
 - 1.2.3 – Inventory
 - 2 – Liabilities
 - 3 – Equity

Master Data Management

Nebylo-by pěkné, kdyby číselníky byly spravované profesionálně?

The screenshot displays the Hyperion System 9 Master Data Management interface. The main window shows a hierarchical tree structure of assets under the heading "Rationalized - Balance Sheet". The tree is organized as follows:

- 1 - Assets
 - 1.1 - Short Term Assets
 - 1.1.0 - Other Short Term Assets
 - 1.1.1 - Cash
 - 1.1.2 - Cash Equivalent
 - 1.2 - Long Term Assets
 - 1.2.0 - Other Long Term Assets
 - 1.2.0.1 - Embedded Derivatives
 - 1.2.1 - Research
 - 1.2.2 - Advertising
 - 1.2.3 - Inventory
 - 2 - Liabilities
 - 3 - Equity

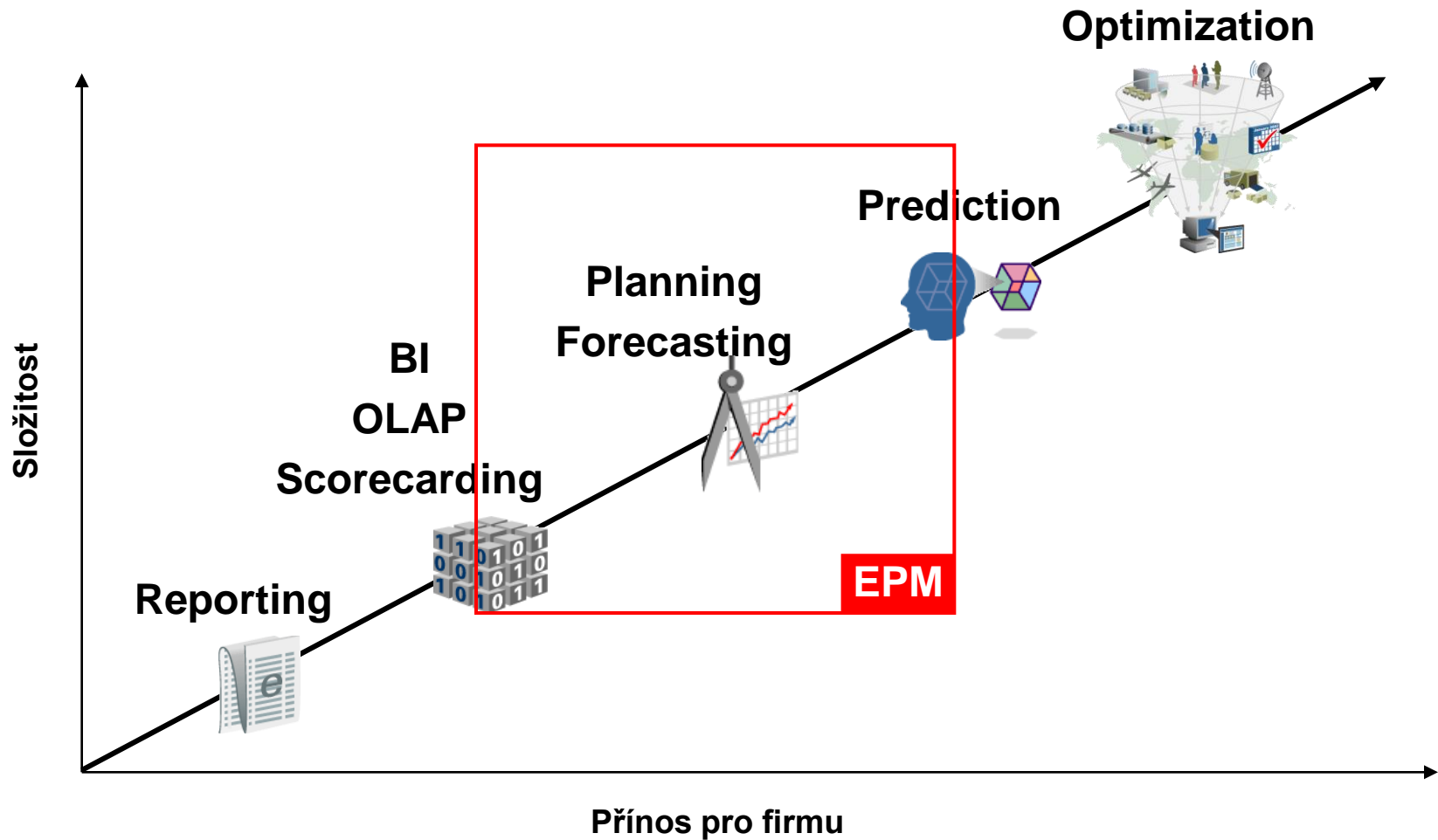
The right-hand pane shows a table of properties for the selected node:

Property	Value
Acct Participation	IFRS
AccountSource	Rational

The interface also includes a menu bar (File, Edit, View, Tools, Admin, Window, Help) and a toolbar with various icons. The status bar at the bottom right indicates the version number "V 9.3.0.39".

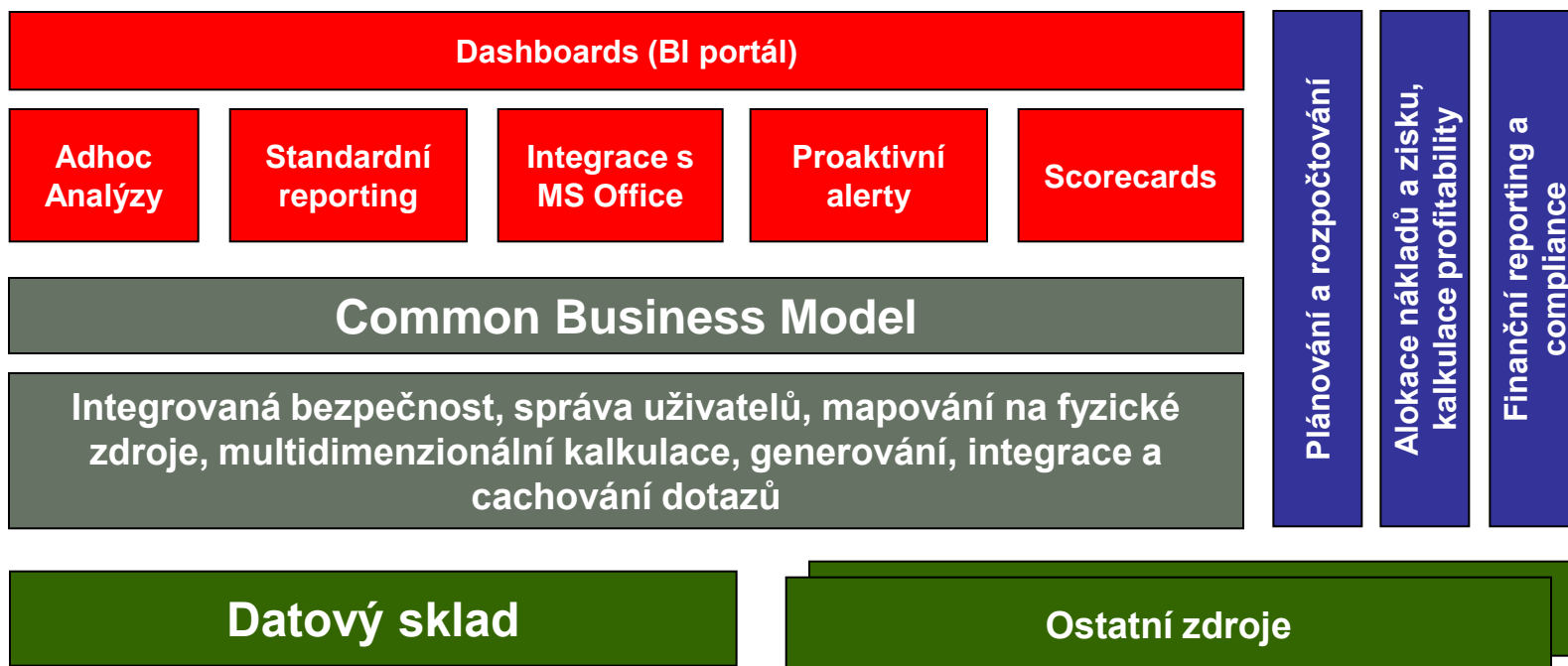
Performance Management

Další vývojový krok



Performance Management

- Datový sklad je sice důležitou komponentou pro správu a analýzu informací ve firmě, sám o sobě ale neumožňuje komplexní využití informací, jejich distribuci běžným uživatelům či manažerům a měření klíčových ukazatelů firmy.
- Proto je vhodné datový sklad doplnit integrovanou platformou pro Business Intelligence a integrovat s nástroji pro správu výkonnosti firmy (EPM – Enterprise Performance Management)



Malá anketa k EPM

- Používá vaše firma systém řízení a odměňování založený na klíčových ukazatelích výkonnosti (KPI)?
 - Kolik z těchto ukazatelů jste schopni sledovat na potřebné úrovni (org. jednotka, produkt, zákazník) v datovém skladu?
- Připravujete rozpočet a forecast na základě modelů, obsahujících technické ukazatele?
 - Kolik z těchto ukazatelů jste schopni poskytovat z datového skladu (historické průměry pro rozpočet, skutečnost pro forecast)?
- Využíváte v kontrolingu ABC modely pro alokaci nákladů na jednotlivé procesy, produkty, oddělení?
 - Kolik z statistik (drivers) jste schopni poskytovat z datového skladu?

Metadata Driven

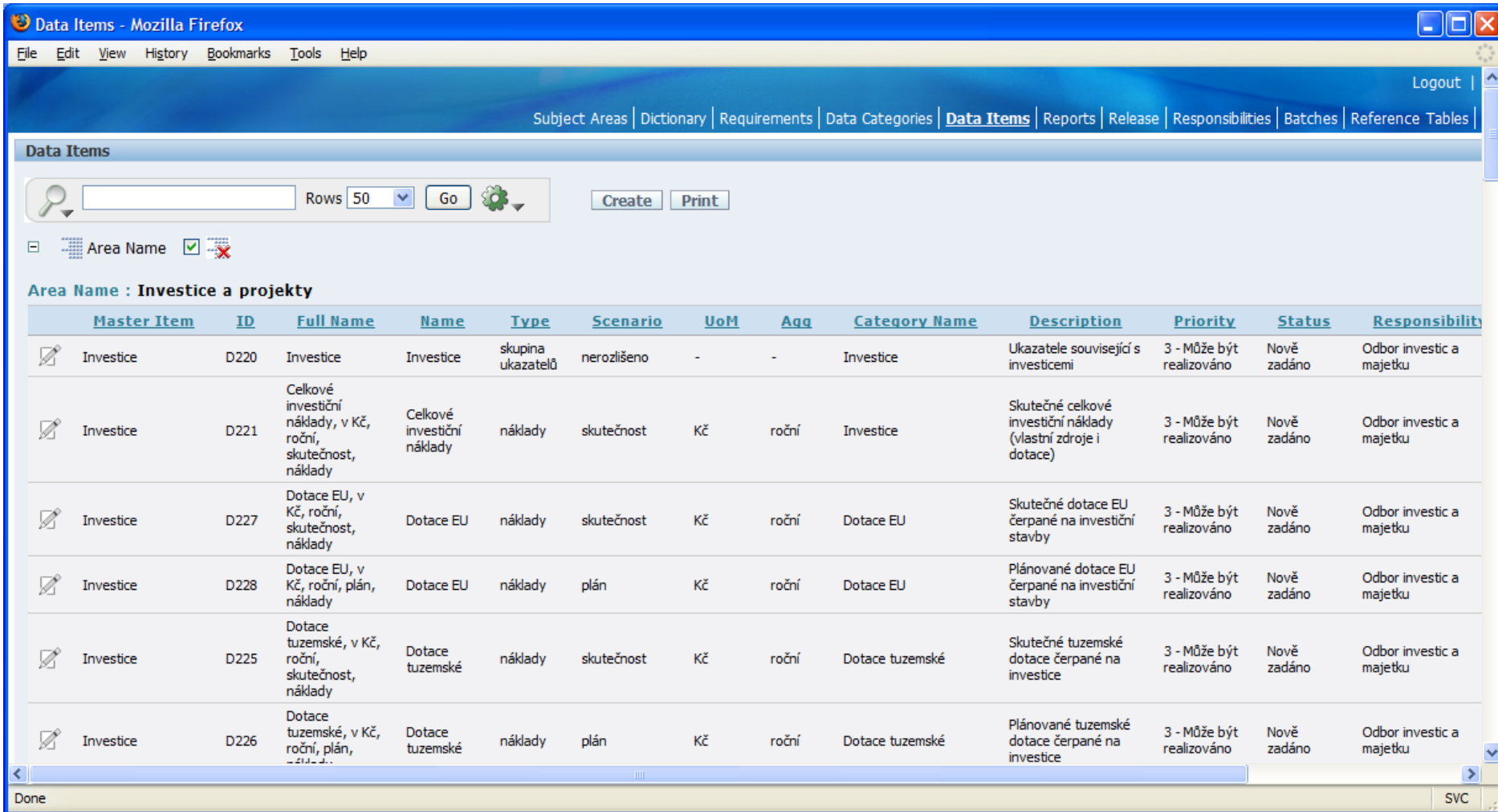
- Klíčová role metadat pro návrh, řízení a dokumentaci BI a datového skladu
 - Aktivní metadata
 - Business model (ukazatele, dimenze, hierarchie, KPI, terminologie)
 - Dependency management (závislosti)
 - Mapování zdrojů na cíle (transformace, pravidla)
 - Stav systému (nahrávání, datová kvalita)
 - Bezpečnost (role, vertikální a horizontální bezpečnost)
 - Change management

Jak efektivně s metadaty pracovat?

- Aktivní business model (fyzický, logický, prezentační)
- ETL procesy řízené a dokumentované metadaty
- Data mart s runtime metadaty (kdy, jak dlouho, kolik)
- Impact Analysis a Data Lineage
- Strukturované logické mapování
- Dokumentace generovaná z metadat
- Portál zpřístupňující metadata uživatelům
- Datový slovník

Master Data Management

Nebylo-by pěkné, kdyby i business metadata byla v repository?



Data Items - Mozilla Firefox

File Edit View History Bookmarks Tools Help

Logout

Subject Areas | Dictionary | Requirements | Data Categories | **Data Items** | Reports | Release | Responsibilities | Batches | Reference Tables

Data Items

Rows 50 Go Create Print

Area Name

Area Name : Investice a projekty

Master Item	ID	Full Name	Name	Type	Scenario	UoM	Aqq	Category Name	Description	Priority	Status	Responsibility
Investice	D220	Investice	Investice	skupina ukazatelů	nerozlišeno	-	-	Investice	Ukazatele související s investicemi	3 - Může být realizováno	Nově zadáno	Odbor investic a majetku
Investice	D221	Celkové investiční náklady, v Kč, roční, skutečnost, náklady	Celkové investiční náklady	náklady	skutečnost	Kč	roční	Investice	Skutečné celkové investiční náklady (vlastní zdroje i dotace)	3 - Může být realizováno	Nově zadáno	Odbor investic a majetku
Investice	D227	Dotace EU, v Kč, roční, skutečnost, náklady	Dotace EU	náklady	skutečnost	Kč	roční	Dotace EU	Skutečné dotace EU čerpané na investiční stavby	3 - Může být realizováno	Nově zadáno	Odbor investic a majetku
Investice	D228	Dotace EU, v Kč, roční, plán, náklady	Dotace EU	náklady	plán	Kč	roční	Dotace EU	Plánované dotace EU čerpané na investiční stavby	3 - Může být realizováno	Nově zadáno	Odbor investic a majetku
Investice	D225	Dotace tuzemské, v Kč, roční, skutečnost, náklady	Dotace tuzemské	náklady	skutečnost	Kč	roční	Dotace tuzemské	Skutečné tuzemské dotace čerpané na investice	3 - Může být realizováno	Nově zadáno	Odbor investic a majetku
Investice	D226	Dotace tuzemské, v Kč, roční, plán, náklady	Dotace tuzemské	náklady	plán	Kč	roční	Dotace tuzemské	Plánované tuzemské dotace čerpané na investice	3 - Může být realizováno	Nově zadáno	Odbor investic a majetku

Done SVC

Bezpečnost především

- S dospíváním datových skladů a zvyšováním počtu uživatelů je nezbytné zajistit odpovídající bezpečnost dat v datovém skladu
- Proč – datový sklad je jeden z mála systémů ve firmě, který obsahuje integrovaná a snadno dostupná data, s osobními informacemi (jména, identifikátory, adresy, kontakty, účty), citlivými finančními informacemi (výnosy, náklady, marže, rozpočet) a obchodními informacemi (počty a kvalita klientů, profily, příležitosti)
- Únik citlivých informací je nejen problematický pro business, ale i legálně postižitelný

Co můžeme udělat pro zajištění bezpečnosti?

- Adresnost a autentikace uživatelů přistupujících k systému
- Jednotné řízení přístupu k datům bez ohledu na použitý nástroj (LDAP, Active Directory)
- Definované role dle funkce, oddělení
- Jasná pravidla pro přidělování rolí (governance)
- Přístupová práva k objektům a privilegiím
- Horizontální bezpečnost (přístupná je jen podmnožina dat)
- Vertikální bezpečnost (citlivé atributy jsou ukryty před neoprávněnými uživateli)
- Auditování přístupu k datům

BI a DWH jako služba

- Datové sklady mohou poskytovat služby ostatním produkčním systémům ve firmě a integrovat tak provozní informace s historií a trendy dostupnými v BI
 - Alerty a trigger generované v datovém skladu
 - Poskytování historických informací
 - Profily a KPI klientů
 - Prediktivní analýza
 - Kalkulace na požádání (skóring, profitabilita)
 - What-If analýza
 - Integrace s procesy v datovém skladu

Předpoklady integrace BI do business procesů

- Garantovaná dostupnost a aktuálnost potřebných informací v datovém skladu
- Architektura a výkon přizpůsobený online integraci (high availability, RAC, oddělení dávkového zpracování a velkých dotazů od online služeb)
- Sjednocený pohled na klíčové entity s ostatními systémy (Master Data Management)
- Architektura
 - Možnost publikovat BI jako Web Service
 - Možnost integrovat BI s ostatními aplikacemi
 - Podpora proaktivních alertů
 - Podpora mobilních zařízení (“Pervasive BI”)

Service Level Agreements

- Formalizace vztahů mezi datovým skladem, uživateli datového skladu a BI a dodavateli zdrojových dat
 - SLA na **dodávku dat ze zdrojových systémů** (frekvence, čas, struktura, dodržení kvality, dopředné informace o plánovaných změnách v struktuře, obsahu; dopředné notifikace o prováděných hromadných změnách)
 - SLA na **dostupnost dat v datovém skladu** (frekvence a čas nahrávání, prodleva mezi změnami dat ve zdroji a jejich dostupností v datovém skladu, samostatně pro denní a závěrková data)
 - SLA na **dostupnost klíčových reportů** (frekvence a plán tvorby reportů, jejich dostupnost pro uživatele, průměrná odezva na klíčové reporty či dotazy)
 - SLA na **dostupnost datového skladu** (časová okna pro uživatelské dotazy a služby, servisní okna, doba recovery po výpadku, frekvence nových verzí)

Předpoklady uzavření a dodržování SLA

- Kvalitní a akceptovaná dokumentace popisující rozhraní mezi datovým skladem a zdrojovými systémy
- Požadavky na dostupnost dat a klíčových reportů v skladu, vycházející z reálných business potřeb a možností systému
- Pravidelné měření a vyhodnocování, jak jsou dodržovány parametry SLA; nejlépe přímo z aktivních metadat v skladu („data mart“ pro servisní parametry skladu)
 - Kdy byla dodána data ze zdrojů, statistika chyb, změn
 - Kdy byla data nahrána do skladu, trvání ETL procesů
 - Kdy byly připraveny standardní reporty
 - Průměrná délka trvání typizovaných dotazů

Trendy v oblasti datových skladů a BI

- ✓ Referenční architektura
- ✓ Rychlost, výkon, objem
- ✓ Time to Market
- ✓ Předpřipravené BI aplikace
- ✓ Důraz na datovou kvalitu
- ✓ Master Data Management
- ✓ BI a Performance Management
- ✓ Metadata driven
- ✓ Bezpečnost především
- ✓ BI & DWH jako služby
- ✓ Service Level Agreements

ORACLE IS THE INFORMATION COMPANY

ORACLE®