

IT a anatomie firmy

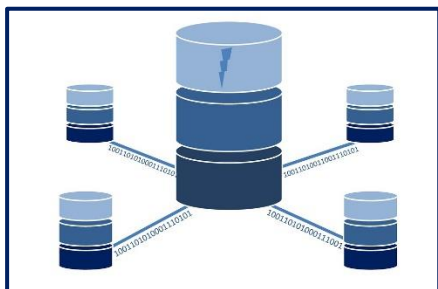
(Datové zdroje)

(pracovní dokument)



Marek, Slezák, Nikola Pátková, Michal Novotný

VŠE Praha, 2023



Mapa podle částí textu

(A) Novotný M: Data Mesh
(B) Pátková, N.: Datové katalogy
(C) Slezák, M.: Metodika datové migrace

Obsah

A. Data Mesh	5
Úvod	6
1. Principy	6
1.1 Federated computational governance	6
1.2 Self-serve datové platforma	6
1.3 Datové domény	6
1.4 Datový produkt	6
2. Přínosy a rizika Data Mesh	9
2.1 Přínosy	9
2.2 Rizika	9
3. Role – vlastník datového produktu	9
4. Závěry a doporučení	9

5.	<i>Zdroje</i>	10
B.	<i>Datový katalog</i>	11
	<i>Úvod</i>	12
1.	<i>Analýza současných přístupů k řešení a užití datových katalogů, jejich možnosti a omezení</i>	12
1.1	<i>Analýza současného trhu</i>	12
1.1.1	SaaS Cloud-based.....	12
1.1.2	Open source.....	13
1.1.3	Commercial Data Catalogs.....	13
1.1.4	On-premises Data Catalogs.....	13
1.2	<i>Současné přístupy k užití datových katalogů</i>	14
1.2.1	Proč implementovat	14
2.	<i>Možnosti datových katalogů</i>	16
3.	<i>Omezení datových katalogů</i>	17
4.	<i>Přehled vybraných dostupných nástrojů datových katalogů</i>	17
4.1	DataHub.....	18
4.2	MS Purview	19
4.3	Informatica	21
4.4	Collibra Data Intelligence Cloud.....	23
5.	<i>Závěr</i>	24
6.	<i>Zdroje</i>	24
C.	<i>Metodika datové migrace</i>	27
1.	<i>Metodika a postup datové migrace</i>	29
1.1	Používejte formální metodiku, která byla vyzkoušená a otestovaná [1].....	29
1.2	Metodika migrace dat by měla postupovat od posouzení, přes plánování až po přesun a validaci [12]	29
1.3	Mějte plán [12] [15].....	29
1.4	Mějte dostatek času na plánování projektu [15][16]	30
1.5	Zkontrolujte, zda jsou k dispozici potřebné zdroje [16] [19].....	30
1.6	Zdokumentujte proces migrace [12][15].....	30
1.7	Vyviňte interní kompetence ohledně migrace dat a nespolehejte zcela na externí zdroje [1] 30	
1.8	Podnik musí být zapojen ve všech fázích projektu, od zahájení až po testování [1] [6].	31
1.9	Získejte písemný závazek vendora [16]	31
1.10	Odstraňte komunikační bariéry [16].....	31
1.11	Automatizujte, pokud je to možné [15]	32
2.	<i>Nástroje pro datovou migraci</i>	32
2.1	Používejte nástroje pro profilování dat před zahájením projektu i v jeho průběhu [1]	32
2.2	Pro čištění dat využijte nástroje pro čištění dat. Nepokoušejte se o čištění ručně [1].	32

2.3	Používejte nástroje pro integraci dat. Nepokoušejte se o ruční kódování [1] [13].....	33
2.4	Ujistěte se, že pro migraci dat používáte správný software [12] [15] [17] [19].....	33
2.5	Je důležité zvážit, zda je možné migraci snadno ukončit a znova spustit [12]	33
3.	<i>Práce s daty</i>	34
3.1	Pochopte data ve zdrojovém systému ještě předtím, než začnete migrovat [6] [15].....	34
3.2	Zaměřte se na správné mapování dat, aby byl zajištěn kompletní přenos dat [16].....	34
3.3	Rámec pro řízení dat má zásadní vliv na celý migrační projekt [1] [6]	34
3.4	Pravidelně zálohujte data pro případ, že by během procesu migrace došlo k problému [17] [19].....	35
3.5	Identifikujte formáty dat, které budete přenášet [17] [19]	35
3.6	Nahrajte data, aby bylo možné zahájit proces migrace [17].....	35
3.7	Investujte do následné údržby, abyste se ujistili, že byla zachována integrita dat [17]..	35
3.8	Omezte počet zprostředkovatelů (vrstev abstrakce) [15].....	35
3.9	Vytvořte si knihovnu opakovaně použitelného kódu [15]	36
3.10	Opravujte kód, ne data [15]	36
4.	<i>Testování</i>	36
4.1	Proveďte v cílovém systému testování, abyste se ujistili, že byla data úspěšně přenesena [17] [19].....	36
4.2	Pravidelně během migračního projektu testujte [16]	37
4.3	Musí být provedeno systémové a uživatelské testování a před spuštěním ostrého provozu je nutné získat podpis [15].....	37
4.4	Plánujte prodloužené období systémových a uživatelských testů [15].....	37
4.5	Vytvořte si vztah s uživateli, kteří testují [15].....	38
4.6	Ujistěte se, že váš kód loguje vše, co dělá [15]	38
4.7	Získejte novou sadu dat pro každé kolo uživatelských testů [15].....	38
5.	<i>Rozpočet</i>	38
5.1	Je-li migrace součástí většího projektu, řešte ji zcela samostatně, a to od sestavování rozpočtu až po testování [1]	38
5.2	Zhodnoťte velikost projektu a zjistěte, jaké zdroje bude vyžadovat a kolik na něj potřebujete v rozpočtu [15] [17] [19].....	39
6.	<i>Použitá literatura</i>	39

A. Data Mesh

(pracovní dokument)



Michal Novotný

VŠE Praha, 2023

Úvod

Data Mesh je podle jeho autorky (Dehghani 2022) nový **přístup k získávání, správě a přístupu k datům pro analytické účely**, který se řeší rostoucí komplexitou správy dat v organizacích a problémy, se kterými se datové architektury v organizacích potýkají. Těmito problémy jsou například centralizované vlastnictví, škálovatelnost či omezená dostupnost. Koncept Data Mesh **je o zásadním posunu v našich předpokladech, architektuře, technických řešeních a sociální struktuře** našich organizací, ve způsobu, jakým spravujeme, používáme a vlastníme analytická data.

1. Principy

Koncept Data Mesh má dle jeho autorky (Dehghani 2022) **čtyři principy**:

1.1 Federated computational governance

Princip federovaného řízení **zajišťuje interoperabilitu všech datových produktů prostřednictvím standardizace** (Jochen et al. 2023). Ve své podstatě jde o vytvoření rámce a zajistit vzájemnou kompatibilitu datových produktů.

1.2 Self-serve datové platforma

Self-serve datová platforma má za **cíl poskytovat funkce a služby, jako je úložiště nebo výpočetní výkon**. Centrální IT poskytuje platformu, která umožňuje to, že se doménové týmy mohou zaměřit pouze na práci, která je pouze doménově orientovaná, tedy že **se mohou zaměřit pouze na vytváření byznysové hodnoty z dat**, místo toho, aby se také staraly o údržbu nástrojů či infrastruktury (Tartow a Mott 2021b).

1.3 Datové domény

Podle (Tartow a Mott 2021a) **datová doména představuje skupinu lidí v organizaci, kteří jsou organizováni kolem společného obchodního účelu či byznysové entity**. Datové domény obvykle odrážejí organizaci z pohledu byznysu. Tedy například pro e-commerce stránku mohou být datové domény uživatelé, obchodníci, produkty nebo marketing. Princip doménového vlastnictví stanovuje, že **doménové týmy mají převzít zodpovědnost za svá data**. Analytická data jsou tedy strukturována kolem domén, do kterých byznysově (doménově) spadají.

1.4 Datový produkt

Posledním principem je chápání dat jako produktu a aplikování produktového myšlení v rámci organizace, tedy tvorba tzv. datových produktů.

Datový produkt lze definovat takto:

„Datový produkt je kombinace datasetu a produktových metadat, která náleží právě jedné datové doméně, má konkrétního vlastníka a vzniká za určitým účelem s cílem naplňovat potřeby svých konzumentů.“

Datový produkt musí splňovat určité charakteristiky, které jsou vypsány v tabulce 1.

Tabulka 1: Charakteristiky datového produktu (zpracováno autorem podle zdrojů (Dehghani 2022; Majchrzak et al. 2023; Jochen et al. 2023))

Podmínka	Popis
Dohledatelnost a snadná dostupnost	Datový produkt musí být jednoznačně pojmenovaný, musí být vy publikován na stanoveném místě a uživatelé musí vědět, kam se v případě potřeby podívat a kde datový produkt hledat

Srozumitelnost a pochopitelnost	Datový produkt musí být pro jeho konzumenty srozumitelný a pochopitelný, respektive dostatečně (nejen byznysově) popsán
Adresovatelnost	Datový produkt musí adresovatelný
Dostatečná zabezpečení	K datovým produktům musí mít přístup jen ti uživatelé, kteří k němu přístup mají
Interoperabilita	Datový produkt musí mít schopnost spolupracovat s dalšími datovými produkty (jeden datový produkt může být zdrojem dalšího datového produktu)
Důvěryhodnost	Datový produkt musí být pro uživatele důvěryhodný, tedy že data jsou správná a že jim může věřit
Hodnota	Datový produkt musí mít hodnotu sám o sobě, nesmí být nutné jeho propojení s dalšími datovými produkty, aby teprve poté dostal hodnotu a smysl
Unikátní doména	Datový produkt patří právě do jedné datové domény, není součástí žádné jiné
Vlastník a vývojový tým	Datový produkt musí mít stanoveného svého vlastníka a vývojový tým, který se stará o vývoj a údržbu datového produktu
Druh	Musí spadat do jednoho z druhů datových produktů
Metadata	O datovém produktu se musí sbírat a uchovávat metadata, která mohou mít různou podobu
Účel	Datový produkt musí vzniknout za jasným účelem a cílem (nemůže se stát, že by datový produkt vznikl jako vedlejší produkt činnosti)
Podmínky konzumace a SLO	Datový produkt musí mít jasně definované podmínky jeho využívání (kdo ho může využívat, kdy, jak)
Vstupní a výstupní porty	Datový produkt musí mít stanovené vstupní porty (přístup ke zdrojům) a výstupní porty (poskytnutí uživatelům)

Obecně platí, že libovolný způsob reprezentace dat, který poskytuje uživatelům nějakou hodnotu, je vhodným kandidátem na to stát se datovým produktem (Majchrzak et al. 2023). Jednoduše řečeno, datový produkt si lze představit **jako dataset splňující výše zmíněné charakteristiky**. Dataset může mít různou podobu, nemá pevně stanovenou formu. Mezi nejlépe představitelnou a zároveň dle autora zdaleka nejčastější podobou datového produktu je libovolný databázový dataset, ať už se jedná o databázovou tabulku, jejich propojení či databázové pohledy.

Datové produkty lze dle jejich orientace rozdělit na **3 základní druhy**:

- datové produkty **zaměřené na zdroj** (source-aligned) – tyto produkty mají za cíl jen přenášet data ze zdrojových systémů (či jiných zdrojů) s minimálními úpravami či transformacemi a poskytovat je ke konzumaci uživatelům (Mott 2022),
- datové produkty **orientované na spotřebitele** či konzumenta (consumer-aligned) - tyto datové produkty přinášejí byznysovou hodnotu díky expertíze a znalostem lidí z oboru (tedy konzumentů),
- **agregované datové produkty** – kombinace (agregace) různých datových produktů z různých domén.

Datový produkt nemůže být datovým produktem, pokud se k němu nesbírají a neuchovávají metadata. Ukázková struktura metadatového minima pro datové produkty je uvedena v tabulce 2.

Tabulka 2: Metadata datového produktu (zpracováno autorem podle zdrojů (Jochen et al. 2023; Majchrzak et al. 2023))

Metadata	Komentář
ID	Unikátní identifikátor datového produktu
Název	Unikátní (byznysový) název datového produktu
Popis	Byznysový popis datového produktu (k čemu slouží a co je jeho obsah)
Vlastník	Vlastník datového produktu (jméno, příjmení, kontakty apod.)
Vývojový tým	Vývojový (údržbový) tým datového produktu (jména, příjmení, kontakty apod.)
Zdroje	Zdrojové systémy a jiné zdroje datového produktu (výpis, případně i popis a odkazy)
Vstupní port(y)	Výpis vstupních portů datového produktu
Výstupní port(y)	Výpis výstupních portů datového produktu
Konzumenti	Výpis současných konzumentů datového produktu
Service-level objective(s)	Výpis SLOs
Podmínky konzumace	Za jakých podmínek lze datový produkt využívat (např. kdo, kdy)
Odkaz	Odkaz na datový produkt
Stav	Zobrazení stavu datového produktu (aktivní, neaktivní apod.)
Datum vytvoření	Datum vytvoření datového produktu (neměnný)
Datum platnosti	Určené datum platnosti datového produktu (mění se v průběhu času)

Verze	Aktuální verze datového produktu
Popis struktury	Popis datasetu datového produktu (popis jednotlivých sloupců, datových typů apod.)

2. Přínosy a rizika Data Mesh

2.1 Přínosy

- díky doménovému vlastnictví dat umožňuje využít naplno agilních způsobů řízení při práci s daty,
- vlastníci a vývojáři mohou rychle a pružně reagovat na změny v oblasti dat,
- vyšší škálovatelnost datového řešení (decentralizace),
- podporuje efektivnější správu dat a lepší spolupráci datových týmů a jednotlivců mezi sebou,
- pomocí publikování datových produktů dělá data (respektive v kontextu konceptu Data Mesh datové produkty) snadno dohledatelná a dostupná pro konzumenty dat k provádění analýz,
- tvorba datových produktů není vzdálena od byznysu a byznysových uživatelů, ba naopak je kladen důraz na propojení byznysové hodnoty,
- jasně definované vlastnictví dat, ale také pevně stanovené odpovědnosti.

2.2 Rizika

- chybí soubor osvědčených postupů, který by implementaci konceptu usnadňoval a usměrňoval,
- konzumenti dat musí dokázat popsat, co přesně potřebují (jaké mají byznysové potřeby),
- nedostatečný důraz na byznysové cíle či problémy,
- nedostatečná komunikace s konzumenty dat,
- velký zásah do firemní kultury (změna architektury, reorganizace a rekvalifikace či nábor zaměstnanců), pravděpodobně doprovázen také složitou implementací.

3. Role – vlastník datového produktu

Vlastník datového produktu je **osoba zodpovědná za nejen byznysové, ale také technické směřování a rozvoj datového produktu**. Tato pracovní role je stěžejní pro úspěšnou implementaci datových produktů. Role vlastníka datového produktu přirozeně vychází z principu, že data jsou vnímána v konceptu Data Mesh jako produkt. Důležitým pravidlem je, že by musí být **pouze jeden vlastník pro konkrétní datový produkt**.

Vlastník je podle (Majchrzak et al. 2023) **odpovědný za**:

- Stanovení důvodu tvorby datového produktu.
- Zjistit očekávání konzumentů a jejich byznysové potřeby.
- Neustálý sběr zpětné vazby.
- Plánování implementace datových produktů a jeho úprav.
- Strategické plánování vývoje datového produktu.

4. Závěry a doporučení

- Koncept Data Mesh je **nový přístup ke správě dat** v organizaci.

- Koncept se snaží řešit **problémy aktuálních datových architektur**, jako je například škálovatelnost či centralizace.
- Mezi **přínosy** patří například efektivnější správa dat, důraz na byznysový kontext, možnost pružně reagovat či vyšší škálovatelnost řešení.
- Mezi **rizika** patří například velký zásah do firemní kultury a architektury, nedostatečná komunikace s konzumenty dat a nedostatečný důraz na byznysové cíle a problémy.
- Koncept má **4 principy** – doménově orientované decentralizované vlastnictví, datové produkty, federated computational governance a self-serve datová platforma.
- Organizace by vždy měla pečlivě **zvážit, zda je pro ni koncept Data Mesh vhodné řešení**, jelikož je časově, finančně a architektonicky náročný na implementaci, a také k němu doposud chybí soubor osvědčených postupů, podle kterého by se mohla organizace řídit.

5. Zdroje

DEHGHANI, Zhamak, 2022. *Data Mesh: delivering data-driven value at Scale*. O'Reilly Media. ISBN 978-1-4920-9239-1.

JOCHEN, Christ, Larysa VISENGERIYEVA a Simon HARRER, 2023. *Data Mesh Architecture*. Online. Dostupné z: <https://datamesh-architecture.com>. [Citováno: 2023-08-05].

MAJCHRZAK, Jacek, Sven BALNOJAN, Marian SIWIAK a Mariusz SIERACZKIEWICZ, 2023. *Data Mesh in action*. Manning. ISBN 978-1-63343-997-9

TARTOW, Colleen a Andy MOTT, 2021a. *Data Mesh: Domain-oriented Ownership & Architecture*. Online. Starburst. Dostupné z: <https://www.starburst.io/blog/data-mesh-and-starburst-domain-oriented-ownership-architecture/>. [Citováno: 2023-11-20].

TARTOW, Colleen a Andy MOTT, 2021b. *Data Mesh: Self-Service Data Infrastructure*. Online. Starburst. Dostupné z: <https://www.starburst.io/blog/data-mesh-starburst-self-service-data-infrastructure/>. [Citováno: 2023-11-20].

B. Datový katalog

(pracovní dokument)



Nikola Pátková

VŠE Praha, 2023

Úvod

Datový management a zpracování firemních dat se poslední roky stává trendem a bohatstvím každé větší společnosti. Společnosti pomocí svých dat a dat dostupných o konkurenčních společnostech mohou zpracovávat analýzy a **data používat jako podklady pro svá strategická plánování**, mohou tedy tak získat konkurenční výhody, pokud dokáží data dostatečně využít.

Ne vždy je snadné získat data, která jsou k reportingu třeba, právě například ve velkých společnostech, které mají k dispozici mnoho dat na různých místech, z mnoha využívaných služeb a aplikací a bez celkového řízení datových toků.

Z tohoto důvodu je vhodné ve společnostech mít **interní tým expertů v oblasti zpracování dat**, kteří nejen že pomohou dostupná data zpracovat, získat z nich dostupné informace a vytvářet tak různé pohledy pro vedení společnosti, plní však také roli trenérů, kteří učí jiné zaměstnance konceptům Business Intelligence, aby si pohodlně z dostupných předpracovaných dat zvládli připravovat pohledy na data sami.

Data je účelné využívat jako **aktivum společnosti a ta musí být aktivně řízena**, včetně zajištění datové kvality. Řízení dat je složitá disciplína zahrnující všechny úrovně společnosti, a to od návrhu strategie až po provádění obchodních procesů. **Základem řízení dat však zůstává zodpovězení základních otázek**, jako jsou: Co jsou zdroje dat? Jakým způsobem jsou používána? Jaké jsou požadavky na data, aby plně podporovala obchodní procesy? Jak se řídí životní cyklus dat? Jak jsou data propojena mezi sebou? Kde data najít? Kdo je jejich vlastníkem a zodpovědnou osobou?

Dalším z **problémů je neefektivní využívání dat**. Mnoho různých týmů a manažerů využívá data ke svým činnostem, často však nevědí, kde a jak mohou data získat, a tak jsou nuceni hledat informace někdy velmi zdlouhavě. Stejně tak se může stát, že více týmů zpracovává a ukládá data stejným způsobem a data jsou tak zbytečně ukládána a zpracovávána vícekrát. Tuto problematiku by měl datový katalog pomoci vyřešit.

Velmi důležitým aspektem je **pravdivost dat**, tedy důvěra v datech pro zákazníky. Data musí být spolehlivým zdrojem, aby se dala využívat jako důležitý podklad ke strategickým rozhodováním celé společnosti. Nelze ji vybudovat ihned, avšak podpůrné procesy především, sdílené znalosti a programy by k tomu měly pomoci.

1. Analýza současných přístupů k řešení a užití datových katalogů, jejich možnosti a omezení

Tato část je zaměřena na **analýzu současného trhu s datovými katalogy**, která se nejprve soustředí na analýzu přístupů k datovým katalogům, následovanou současnými přístupy k užití datových katalogů a v poslední části je uveden přehled několika dostupných datových katalogů, které se liší právě přístupy.

1.1 Analýza současného trhu

K samotné aplikaci datového katalogu lze využít **4 přístupy řešení**, SaaS, Open Source, komerční a on-premises datové katalogy.

1.1.1 SaaS Cloud-based

Model SaaS (Software jako služba) umožňuje uživatelům připojit se ke cloudovým aplikacím a používat je přes internet. SaaS poskytuje **úplné softwarové řešení**, které lze zakoupit na základě

průběžných plateb od poskytovatele cloudové služby. **Pronajímá se jako možnost použití aplikace** a uživatelé se k ní připojují přes internet obvykle pomocí webového prohlížeče. Veškerá podpůrná infrastruktura, middleware, software a data aplikace jsou umístěné v datovém centru poskytovatele služeb. Poskytovatel služeb spravuje hardware a software a v rámci příslušné smlouvy o poskytování služeb zajišťuje dostupnost a zabezpečení aplikace a dat. SaaS dovoluje organizaci začít aplikaci rychle využívat s minimálními pořizovacími náklady. (Microsoft, 2023).

Výhody:

Toto řešení poskytuje jednodušší přístup odkudkoliv pomocí internetového připojení. Jelikož zákaznická organizace nemá pod správou hardware, tak automatické aktualizace a správa hardware je na straně poskytovatele služeb.

Nevýhody:

Nevýhodou tohoto řešení mohou být vysoké poplatky za využívání služby, i když zrovna aktivně nemusí být používána, nebo závislost na kvalitě a rychlosti internetového připojení.

1.1.2 Open source

Open source je software či produkt se zdrojovým kódem, který kdokoliv může kontrolovat, upravovat a vylepšovat. Původně termín vznikl v souvislosti se specifickým přístupem k vytváření počítačových programů, dnes však pojem označuje **širší soubor hodnot, kdy zahrnují principy otevřeného vývoje, výměny, spolupráce, rychlého vytváření prototypů, transparentnosti a komunitně orientovaného rozvoje.**

V tomto přístupu je tedy zákazníkům k dispozici zdrojový kód aplikace a zákazník kromě samotné implementace se tak může podílet na samotném rozvoji produktu. Často je na straně poskytovatele zřízeno **fórum, kde zákazníci, především vývojáři, mohou sdílet svou zpětnou vazbu** či dokonce své rozšířené verze produktu, návrhy opravy chyb a jiné. (Opensource.com, 2023)

Výhody:

Mezi výhody řešení tohoto typu patří rozhodně cena, jelikož nejsou zde žádné pořizovací náklady a jsou vysoce přizpůsobitelné díky přístupu ke zdrojovému kódu.

Nevýhody:

Ke vhodnému nastavení a správě je potřeba dostatečné technické znalosti, nemusí vždy poskytovat dostatečnou zákaznickou podporu, což se může zařadit mezi nevýhody. (Klipfolio, 2023)

1.1.3 Commercial Data Catalogs

Komerční datové katalogy jsou **placená řešení často nabízená s přídatnými funkcemi** a zákaznickou podporou. Jsou nejvhodnější pro větší společnosti, které vyžadují pokročilé funkce a dedikovanou podporu.

Výhody:

Výhodou tohoto typu řešení je, že poskytuje komplexní funkce, nástroje a vyhrazenou zákaznickou podporu.

Nevýhody:

Mezi nevýhody tohoto řešení lze řadit vyšší cenu a počet funkcí, které jsou s datovým katalogem implementovány, avšak nevyužití. (Klipfolio, 2023)

1.1.4 On-premises Data Catalogs

On-Premises datové katalogy jsou **instalovány a provozovány na vlastních serverech** dané společnosti, což je vhodným řešením pro společnosti, jejichž prioritou je kontrola a vlastní infrastruktura.

Výhody:

Výhodou tohoto typu je plná kontrola nad datovým katalogem a jeho zabezpečení, především co se týká hardwarové vrstvy, stejně tak jako že neexistuje přílišná závislost na kvalitě internetového připojení.

Nevýhody:

To, co pro jiné může být výhodou lze však zařadit i mezi nevýhody, jelikož hardwarové vybavení musí být tedy umístěno a spravováno zákaznickou společností a finanční náklady na pořízení a správu tak nemusí být vždy výhodné. Stejně tak jako prvotní nastavení datového katalogu, které může být nejen tedy drahé, ale také velmi náročné. (Klipfolio, 2023)

Datové katalogy tedy jako software mohou být nabízeny několika způsoby a je tedy na zákazníkovi si vybrat, který přístup mu vyhovuje nejvíce. Některé mohou být také **nabízeny jako součást balíčku produktů, či samostatně**. Lze se tedy rozhodovat na základě zabezpečení, ceny, dostupnosti zákaznické podpory, technických dovedností na straně zákazníka či dispozice hardware.

1.2 Současné přístupy k užití datových katalogů

Datový katalog lze přirovnat k existujícímu systému v knihovnách. Je to **v podstatě soubor metadat o dostupných datech**. Avšak aby data a informace získávané z nich byly dostatečně věrohodné, je potřeba nastavit proces, který důvěru uživatelům poskytne. V praxi se ustálil **3 krokový mechanismus**:

1. **Find:** Nejprve analyzovat data, kterými společnost disponuje, uvést jejich lokaci, datové typy a jiné užitečné metainformace.
2. **Explore:** Následuje objasnění vzniku dat a jejich reprezentace a kvalita.
3. **Trust:** Ohodnocení dat na základě vstupů z předcházejících kroků a zároveň vystavení procesu kontinuálního zlepšování, který zajistí zisk zpětné vazby od koncových uživatelů na poskytovaná data, případně odstraní identifikované kvalitativní nedostatky. (Lukšík, 2020)

Datový katalog má za úkol splnit **3 hlavní úkoly**:

1. *Dataset searching.*
2. *Dataset Evaluation.*
3. *Data Access.*

Dataset Searching pomáhá uživatelům **s hledáním požadovaných dat na základě klíčových slov, tagů, businessových výrazů** a vrací výsledky na základě relevance a četnosti. Vyhledávání pomocí přirozeného jazyka umožňuje méně technickým uživatelům najít svá data.

Dataset Evaluation, neboli vyhodnocování datasetu, slouží k **vyhodnocování dat na první pohled** a bez omezeného přístupu. Proto by lidé měli mít možnost zobrazit náhled datasetu, číst uživatelské recenze a anotace kurátorů, zobrazit podrobnosti o kvalitě dat a jejich metadata.

Data Access, tedy přístup k datům, zastává cestu od vyhledávání dat, jejich vyhodnocování až po přístup k nim, která by měla být bezproblémová. Součástí datového katalogu by měly být přístupové protokoly, aby byla zajištěna bezpečnost, soukromí a používání citlivých údajů v souladu s předpisy. (Wells, 2020)

1.2.1 Proč implementovat

Implementace datového katalogu v první řadě pomáhá zmapovat dostupná data ve společnosti za účelem jejich zpřehlednění. Dalším přínosem je **lepší dostupnost informací o datech** ve společnosti pro všechny zainteresované zákazníky, kterými jsou zaměstnanci společnosti, především businessoví a datoví analytici. Ti tato data analyzují za účelem jejich pochopení a nalezení nejlepšího využití. (Wells, 2020). Dalším z důvodů, proč implementovat datový katalog je **vize zlepšení kvality dat**.

Pojem datová kvalita může být definován jako „the degree to which a set of characteristics of data fulfills requirements“. (Transition Support, 2023)

Datová kvalita může být měřena dle metrik, definovaných v práci „Metriky, monitoring a řídicí proces Data Governance“. Jedná se o 6 identifikovaných metrik, kterými jsou Přesnost, Úplnost, Aktuálnost, Konzistence, Včasnost a Nestálost. Obecné příklady metrik jsou inspirovány zdrojem a vlastními zkušenostmi autorky.

1. Přesnost (Precision)

- a. Přesnost odráží **míru, jakou data správně odráží skutečnost**. Problémy s přesností si můžeme představit například syntaktické, jako jsou překlepy, nebo semantické, jakožto vyplnění hodnoty do nesprávného sloupce.
- b. Obecný příklad metriky:
 - i. Počet hodnot v neodpovídajícím sloupci.
- c. Konkrétní příklad:
 - i. V tabulce zaměstnanců se ve sloupci job_title nachází 17 % datumů, které neodpovídají žádné hodnotě označení pracovní pozice a svým formátem vyčnívají.

2. Úplnost (Completeness)

- a. Úplnost zobrazuje, **jak kompletně jsou vyplněny jednotlivé atributy**. V současné praxi není obvyklé rozlišovat mezi povinnými a nepovinnými atributy, a tak často může prázdná hodnota (NULL) dávat smysl, případně alternativní hodnota, jako „1900-01-01“ pro datové hodnoty či „Neuvedeno“ pro textové. Existují i tzv. „nice to have“ atributy, kde je prázdná hodnota z technického hlediska povolena, avšak z hlediska podniku může být její vyplnění považováno za důležité. Při analýze úplnosti lze do negativního hodnocení úplnosti zahrnout i alternativní hodnoty, jejichž příklady byly uvedeny výše.
- b. Obecný příklad metriky:
 - i. Výskyt „1900-01-01“ u nenulových atributů (požadováno vyplnění).
- c. Konkrétní příklad metriky:
 - i. 60 % tréninkových dat nemá vyplněný začátek tréninku.

3. Aktuálnost (Currentness)

- a. Aktuálnost dat je definována **časem mezi změnou informací o sledovaném objektu na zdrojovém systému a její změnou v datech**.
- b. Obecný příklad metriky:
 - i. Časový interval mezi aktualizací dat ze zdrojového systému a replikou v datech.
- c. Konkrétní příklad:
 - i. Doba trvání přiřazení nově vytvořených účtů k definovaným rolím v interním systému je v 8 % případech až 48 hodin.

4. Konzistence (Consistency)

- a. Konzistence udává, **jak jednotný je obsah dat o jednom unikátním objektu napříč různými částmi organizace**. Optimálního stavu je dosaženo při existenci jediné platné verze dat označované jako „verze pravdy“ (*single point of truth*).
- b. Obecný příklad metriky:
 - i. Počet různých verzí informací o zaměstnanci společnosti napříč systémy udržující data o zaměstnancích.
- c. Konkrétní příklad:
 - i. Z celkových 5 systémů udržující data o zaměstnancích má paní Pátková ve 2 systémech lišící se informace o týmu v rámci organizace, jehož je v současné době součástí.

5. Včasnost (Timeliness)

- a. Včasnost je dimenze, která je orientovaná na čas, sleduje, **zda uživatelé mají svá data dostupná včas k tomu, aby mohli správně a včas provádět rozhodnutí**.

- b. Obecný příklad metriky:
 - i. Procentuální podíl situací, kdy manažeři oblastí podnikání nedostanou včas informace o aktuálním počtu zaměstnanců v jejich příslušné oblasti v porovnání s celkovým počtem sledovaných časových okamžiků.
- c. Konkrétní příklad:
 - i. V 90 % případů dostane manažer oblasti podnikání informaci o aktuálním počtu svých zaměstnanců včas, kdy v rámci měsíce bylo provedeno sledování aktuálního stavu na denní bázi a z 30, kdy ve 27 dnech byla obdržena data včas.

6. Nestálost (Volatility)

- a. Nestálost **sleduje frekvenci změn dat v čase**. Ačkoliv ji lze považovat spíše jako dimenzi, která slouží jako základ pro podpůrné metriky, může indikovat klíčové problémy a být důležitým aspektem například v případě fraud monitoringu, což je sledování podvodů, kde může sloužit k identifikaci potenciálních nepravostí.
- b. Obecný příklad metriky:
 - i. Počet změn hodnoty atributu za vybrané časové období.
- c. Konkrétní příklad:
 - i. Hodnota grade se u paní Pátkové za poslední týden změnila 8x.

(Vacek, 2014)

Podobně jako u prediktivní analytiky v podnikové informatice, datový katalog může napomoci v těchto oblastech, jakými jsou Analýza dostupných zdrojů dat, Organizace, integrace a čištění dat a Identifikace relevantních vztahů. (Pour, 2017)

V případě **analýzy dostupných zdrojů je kromě přehledu vyhodnocována i jejich kvalita, dostupnost, konzistence a konsolidace**. Pro organizaci, integraci a čištění dat datový katalog může podpořit právě přehledností a přístupností, kdy lze identifikovat právě nedostatky v kvalitě dat. A jelikož některé datové katalogy zahrnují funkce data mining, například pro data lineage či jiný pohled na identifikaci relevantních vztahů, lze je považovat také za nástroj, který svým způsobem podporuje i aktivity spojené s řešením prediktivní analytiky. (Pour, 2017).

2. Možnosti datových katalogů

Možnosti datových katalogů jsou různé dle daného softwaru. Nicméně obecně lze uvést několik příkladů, jaké možnosti a funkce mohou datové katalogy nabízet. Zdrojem pro tuto část práce byly především samotné datové katalogy a popis služeb, které nabízejí.

Jednou z možností datových katalogů je **katalogizace dat**, tzn. že data katalogizují metadata o datech a usnadňují tak identifikaci a porozumění datům. Fungují tedy dost podobně, jako systém knihoven. Tato možnost také úzce souvisí s další možností, **centralizace dat**, kdy datový katalog poskytuje centralizovaný přístup ke všem datovým zdrojům, což vede k usnadnění jejich správy a využití. S centralizovanými daty s metadaty je také možností snadněji řídit **bezpečnost a řízení přístupu k datům**, což může vést k vyšší transparentnosti využívání dat.

Další z možností je **metadata management**, tedy uchování a správa metadat o datech, včetně popisů, schémat, klíčových slov a dalších, což podporuje další z možností, kterou je snadnější hledání a vyhledávání dat na základě různých kritérií, která právě prohledávají uvedená metadata. Tuto možnost nabízejí pomocí interaktivních uživatelských rozhraní. Další související možností ulehčující vyhledávání a samotnou organizaci dat je možnost přidávat tagy a anotace k datům.

Datové katalogy nabízejí často **možnost integrovat s existujícími nástroji a technologiemi**, jako například databázové systémy, nástroje Business Intelligence a další.

Obecně k řízení metadat většina datových katalogů nabízí podporu pro různé **metadatové standardy** umožňující konzistentní a interoperabilní správu metadat, ať už se jedná o firemní standardy nebo ISO standardy.

Některé datové katalogy nabízejí **oznamování a sledování změn v datech a metadatech**, což usnadňuje jejich správu a udržování dat aktuálních a relevantních. Tato oznamování a sledování lze u některých katalogů také získávat jako upozornění, lze je omezit pouze na vybrané změny či události v datech. Tato funkce může být podporována možností **auditování a sledování změn**, pro zajištění integrity dat. Mimo těchto oznámení nabízejí některé datové katalogy také **tvorbu reportů a statistik o používání metadat**, což lze využít nejen ke zlepšení samotné správy metadat, ale také k analýze a plánování.

Kromě výše zmíněných funkcí lze vybrané datové katalogy využít také jako **formu dokumentace** a nástroj pro vytváření a udržování dokumentace datových zdrojů.

Některé z datových katalogů nejsou omezeny pouze na malé datové sady, ale umožňují správu a zpracování větších objemů dat, tzv. Big Data. (Ham a Crockett, 2021)

3. Omezení datových katalogů

Jako mnoho jiných nástrojů, tak i datové katalogy mají svá omezení, které mohou narušovat efektivitu a zpomalovat či přímo bránit přínosům jejich implementace. Mezi taková omezení lze zmínit:

1. **Neaktuálnost metadat** – metadata o datech v katalogu je nutné udržovat aktuální, metadata mohou zastarat a nepřesně zobrazovat aktuální stav dat, především v dynamicky se měnících datových satech. Proto je nutné metadata pravidelně kontrolovat a aktualizovat.
2. **Složkost a náklady na implementaci** – vybrané typy datových katalogů může být náročné implementovat a následně spravovat, jiné zase mohou být příliš nákladné. Zejména pokud společnost nedisponuje dostatečnou kapacitou, financemi a technologickými zkušenostmi. Proto je důležité dbát na kvalitě a přesnosti prvotní analýzy při výběru řešení.
3. **Uživatelské rozhraní** – některé datové katalogy nemají vhodné uživatelské rozhraní a může být obtížné pro pochopení a užívání koncovými uživateli. V takovém případě je risk, že uživatelé budou hledat jiné cesty k přístupu k datům či více vytěžovat interní podporu datového katalogu.
4. **Kvalita metadat** – aby byla zajištěna efektivita datového katalogu, musí mít metadata o datech dostatečnou kvalitu. Pokud nejsou správná, úplná, aktuální, mohou vést ke ztrátě důvěryhodnosti dat.
5. **Bezpečnost a řízení přístupu** – pokud datové sady obsahují citlivá data, je klíčové data dostatečně zabezpečit, aby nebyla data ohrožena.
6. **Škálovatelnost** – pro společnosti s větším objemem dat a větším počtem uživatelů je důležité, aby byl datový katalog škálovatelný a umožnil pokrytí zdrojů hardwarových i softwarových.
7. **Nedostatek standardizace** – některé datové katalogy mohou aplikovat různé standardy a formáty metadat, která mohou mít problém s integrací s jinými systémy.
8. **Kontext** – datové katalogy mohou mít omezenou schopnost zachovat kontext kolem dat, což může vést k potížím uživatelů pochopit souvislosti s jinými daty.
9. **Náročnost udržování a správy** – nedostatek prostředků k udržování a správě datového katalogu může vést ke snížení jeho efektivitu.

Zdrojem této kapitoly byla inspirace v předchozích kapitolách vlastním zpracováním a internetové zdroje. (Klipfolio, 2023)

4. Přehled vybraných dostupných nástrojů datových katalogů

Výběr datových katalogů pro tento přehled byl kromě dostupných datových katalogů v rámci společnosti inspirován také trendem uvedeným společností Gartner. Byly vybrány 4 nástroje od různých poskytovatelů, a to **DataHub, Microsoft Purview, Informatica a Collibra**.

Pro nabízené funkce byly zpracovány tabulky shodné pro vybrané datové katalogy. Porovnání bylo vypracováno na základě informací dostupných na stránkách poskytovatele.

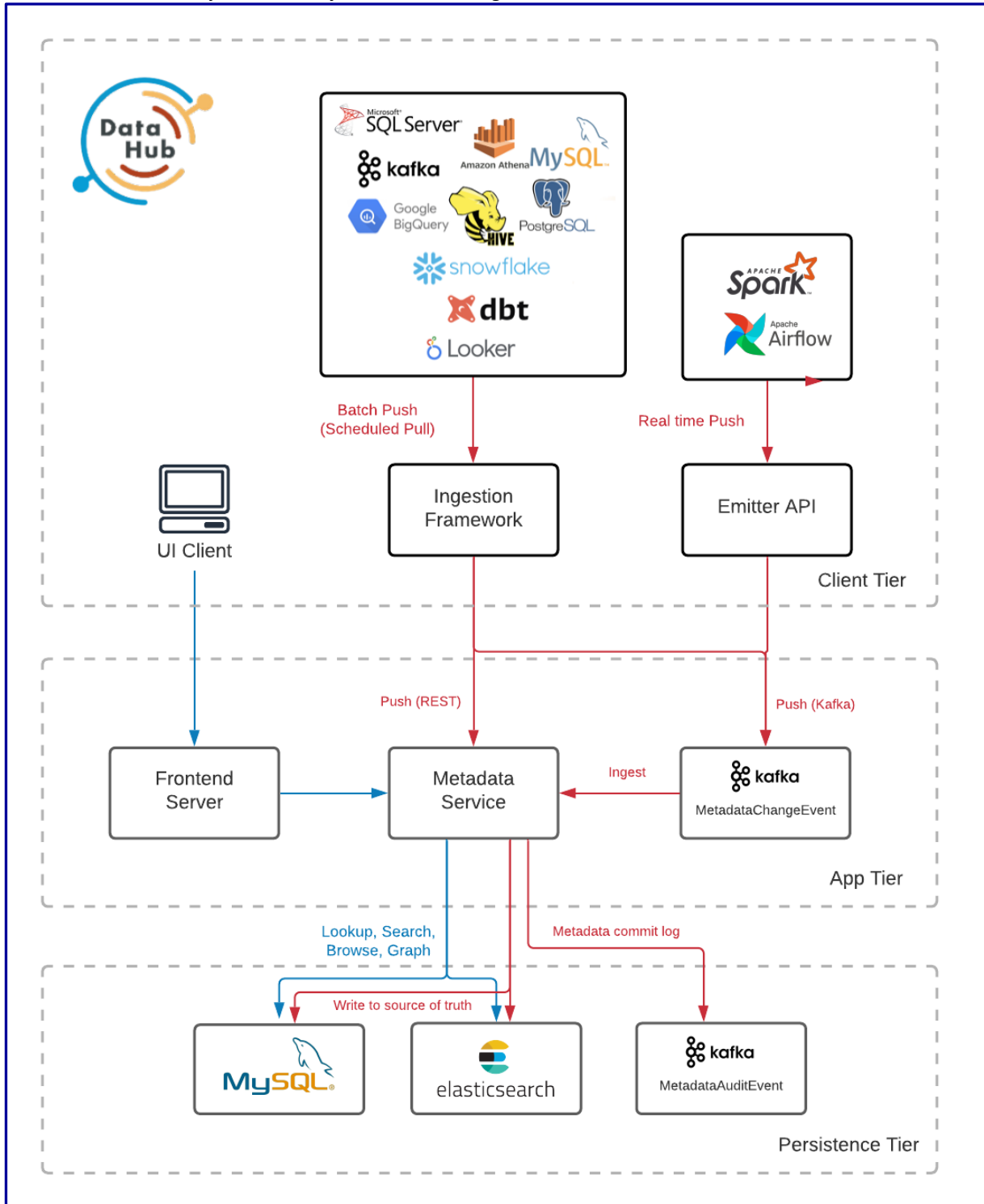
4.1 DataHub

DataHub je **open source datový katalog** vyvinutý ve spolupráci společností Acryl Data a LinkedIn. Poskytuje základní funkce datového katalogu a jako open source řešení navíc umožňuje vlastní rozšíření a správu. Nabízí například **předpřipravené integrace** se systémy Kafka, Airflow, MySQL, MS SQL Server, Postgres, LDAP, Snowflake, Hive, BigQuery a další. (DataHub, 2023)

Tabulka 1: DataHub provedení vybraných kritérií (vlastní zpracování)

Funkce	Provedení
Metadata Management	Správa a uchování metadat o datech, popisy, tagy, a jiné, dataset stats. Metadata lze spravovat také skrze UI.
UI/Vyhledávání	Vyhledávání umožněno skrze uživatelské rozhraní.
Bezpečnost a řízení přístupu	Umožněno skrz administraci - vytáření uživatelů, skupin a pravidel přístupů.
Podpora pro metadatové standardy	Součástí repozitáře je metadata služba, podporuje však i federované služby komunikující s centrálním vyhledávacím indexem.
Reporty a statistiky	Nabízena možnost tvorby automatických reportů na základě vybraných kritérií.
Oznamování a zasílání upozornění	Ano, lze nakonfigurovat jako rozšíření, například do MS Teams, Slack.
Centralizovaný přístup	End-to-End Lineage
Auditování a sledování změn	Ano, v rámci Govern in Real Time
Integrace s dalšími nástroji	DataHub lze integrovat s nástroji třetích stran, například synchronizace Tagů do Snowflake systému, nebo vytváření JIRA ticketů a jiné.

Obrázek 1 dokumentuje **architekturu datového katalogu** DataHub.



Obrázek 1 DataHub architektura (DataHub, 2023)

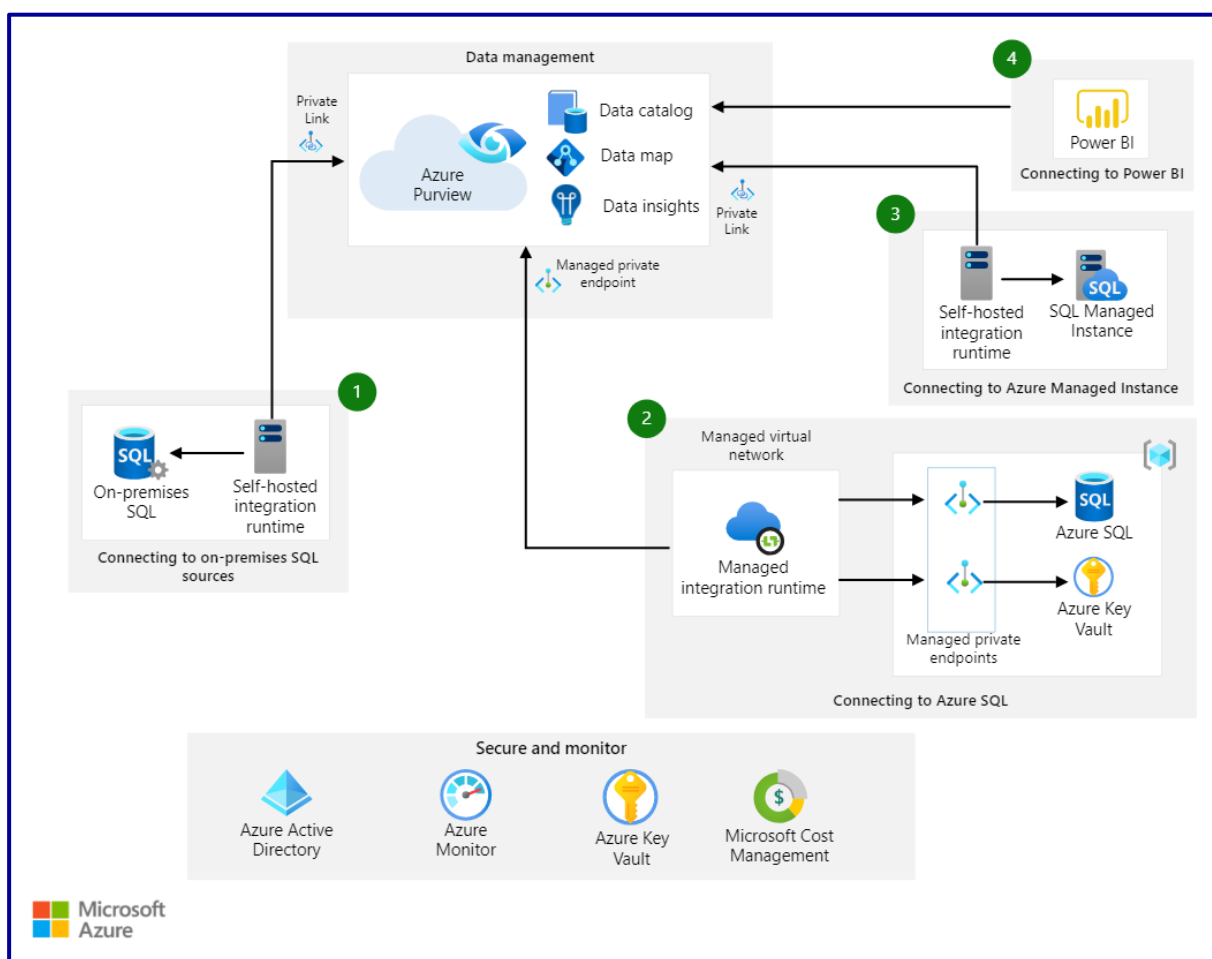
4.2 MS Purview

MS Purview je datový katalog společnosti Microsoft. Slouží **k centralizovanému řízení dat a metadat** v organizaci a lze ho plně integrovat s ekosystémem Microsoft, jako například MS Azure, Azure Data Factory, Power BI a jiné. (Microsoft, 2023).

Tabulka 2: MS Purview provedení vybraných funkcí (vlastní zpracování)

Funkce	Provedení
Metadata Management	MS Purview umožňuje správu a uchování metadat o datech, včetně popisů, schémat a dalších klíčových informací.
UI / Vyhledávání	Vyhledávání je umožněno skrze uživatelské rozhraní.
Bezpečnost a řízení přístupu	Umožněno pomocí vytváření uživatelů, správu rolí. Další funkce pro zabezpečení dat a metadat, včetně řízení přístupu, jsou například služba Information Protection, Insider Risk Management, Data Loss Prevention, Customer Lockbox, Information Protection SDK, Message Encryption.
Podpora pro metadatové standardy	Purview podporuje různé metadatové standardy, což usnadňuje konzistentní správu metadat, například pomocí MS Data policies, Data Estate Insights.
Reporty a statistiky	Uživatelé mohou generovat reporty a statistiky o používání dat a metadat.
Oznamování a zasílání upozornění	Microsoft Purview umožňuje nastavit upozornění na změny v datech a metadatech nebo například na policies která nemají vyřešené problémy.
Centralizovaný přístup	Data Map
Auditování a sledování změn	MS Purview nabízí například moduly MS Purview Audit pro auditování, a MS Purview eDiscovery a Data Lifecycle Management pro sledování změn, nebo například životní cyklus dat.
Integrace s dalšími nástroji	Platforma Purview má silnou integraci s Azure Data Services, což zahrnuje Azure Synapse Analytics a Azure Data Lake Storage.

Obrázek 2 dokumentuje architekturu MS Purview:



Obrázek 2 MS Purview architektura (Microsoft, 2023)

4.3 Informatica

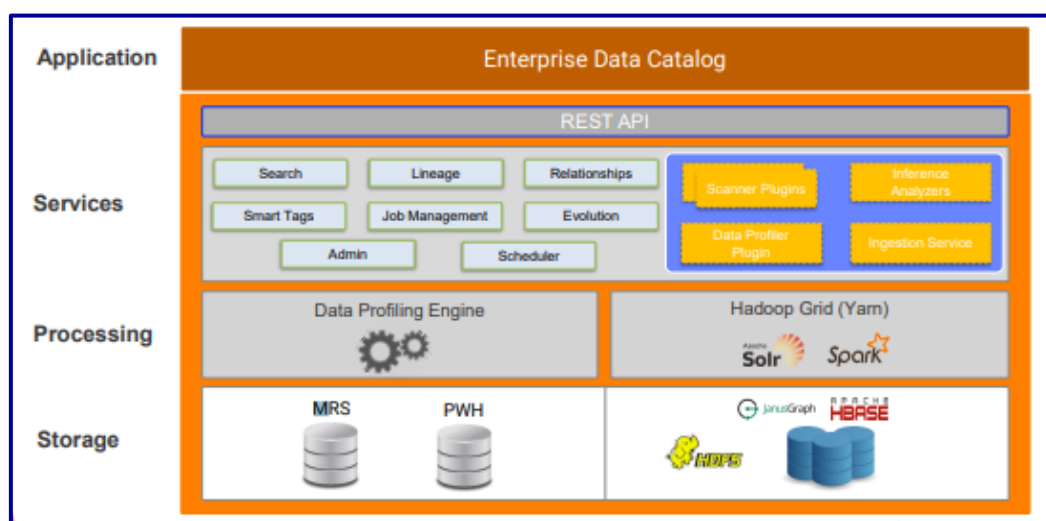
Americká akciová společnost Informatica, založena v roce 1993 zakladateli Gaurav Dhillon a Diaz Nesamoney v Kalifornii, USA. Dle informací dostupných na oficiálních stránkách byla společnost v roce 2023 lídrem na trhu v oblastech Data Quality Solutions, Data Integration Solutions a Master Data Management Solutions hodnocené společností Gartner, Magic Quadrant. Nabízí **celkové cloudové řešení** Intelligent Data Management Cloud na bázi engine CLAIRE umělé inteligence ke správě dat jakéhokoliv typu, schématu, složitosti nebo pracovní zátěže v jakékoliv lokaci na jediné platformě. (Informatica, 2023)

Tabulka 3 Informatica provedení vybraných funkcí (vlastní zpracování)

Funkce	Provedení
Metadata Management	Definice a správa metadat
UI/Vyhledávání	Umožněno skrze uživatelské rozhraní.

Bezpečnost a řízení přístupu	Obsahuje funkce pro zabezpečení dat a metadat a umožňuje řízení přístupu k nim.
Podpora pro metadatové standardy	Informatica podporuje různé metadatové standardy, a nabízí možnost integrace standardizace dat, ověřování a jiné, podporováno službou Informatica Cloud Data Governance and Catalog.
Reporty a statistiky	Uživatelé mohou generovat reporty a statistiky o používání dat a metadat.
Oznamování a zasílání upozornění	Informatica umožňuje nastavit upozornění na určité změny nebo události a zasílat oznámení srkze UI, email či jiné.
Centralizovaný přístup	End-to-End data lineage, holistické pohledy na vztahy mezi daty.
Auditování a sledování změn	Platforma umožňuje provádět audit změn v datech a metadatech, nebo například prediktivní analýzu.
Integrace s dalšími nástroji	Integrace s databázovými systémy, BI nástroji a další.

Obrázek 3 prezentuje architekturu produktů společnosti Informatica.



Obrázek 3 Informatica architektura (Informatica, 2023)

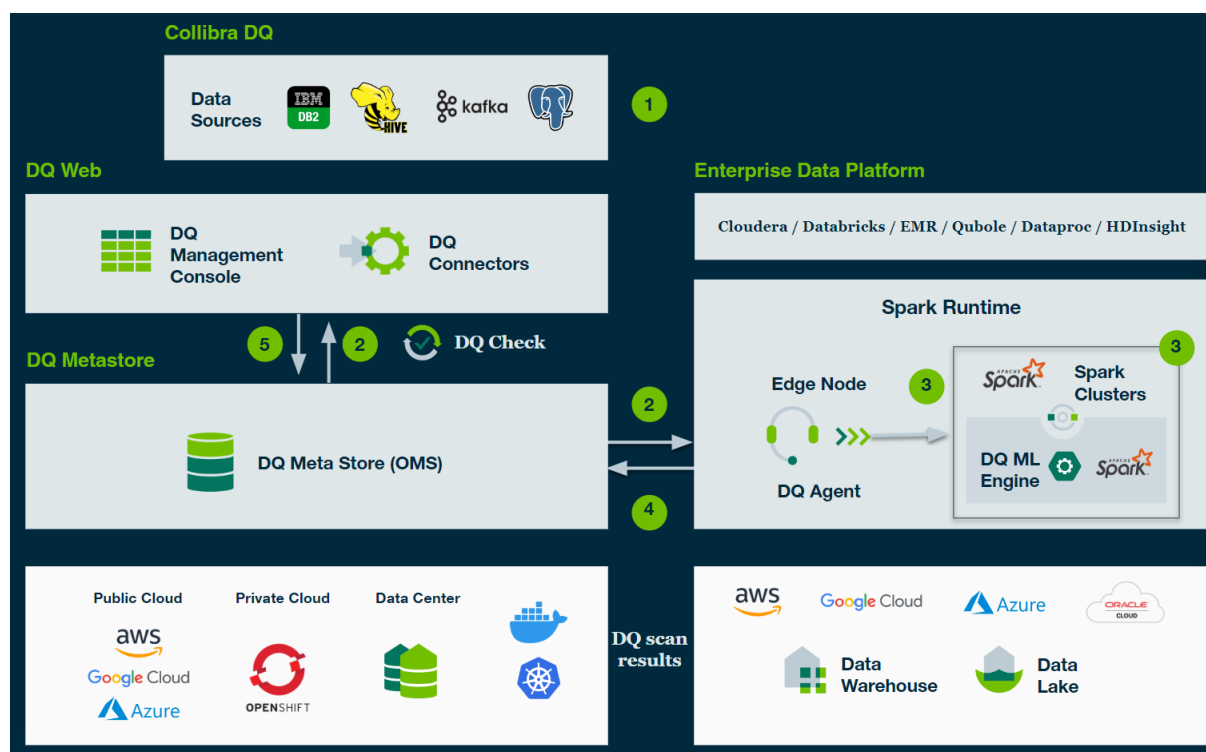
4.4 Collibra Data Intelligence Cloud

Společnost Collibra nabízí podnikové řešení Collibra Data Governance Center, jejíž součástí je právě Collibra datový katalog. Používá **sémantické vyhledávání**, **nabízí end-to-end vizualizaci datových toků**, **automatizaci klasifikace a kategorizace datasetů**, business procesů a intuitivní workflow engine. (Collibra, 2023)

Tabulka 4 Collibra provedení vybraných funkcí (vlastní zpracování)

Funkce	Provedení
Metadata Management	Business, technické i bezpečnostní metadata.
UI/Vyhledávání	Umožněno skrze uživatelské rozhraní.
Bezpečnost a řízení přístupu	Umožňuje řízení přístupu, asset-level security, a současně dodržuje měnící se regulační požadavky v celosvětovém měřítku.
Podpora pro metadatové standardy	V rámci automatizace data governance.
Reporty a statistiky	Data quality reports, usage reports
Oznamování a zasílání upozornění	Skrze UI
Centralizovaný přístup	Nativní business a technické data lineage, one-stop-shop.
Auditování a sledování změn	Umožňuje sledovat změny v datasetech, zabezpečení i uživatelské změny.
Integrace s dalšími nástroji	Platforma Collibra umožňuje integraci s existujícími nástroji a systémy, včetně databázových systémů, business intelligence nástrojů a dalších.

Obrázek 4 dokumentuje architekturu Collibra Data Intelligence Cloud:



Obrázek 4 Collibra architektura (Collibra, 2023)

5. Závěr

Všechny datové katalogy poskytují **velmi podobné možnosti**, které byly zkoumány. Disponují vlastními rozhraními a často je lze integrovat s dalšími produkty poskytovatele.

Microsoft si zakládá na integraci s ostatními systémy a aplikacemi společností vyvinutými, je tedy vynikající volbou pro společnosti, která využívá i jiné produkty v rámci Azure. Je nabízen v rámci cloudových služeb Microsoft Azure.

Informatica datový katalog se specializuje na řešení integračních a datových správcovských potřeb. Nabízí silné integrační a ETL možnosti, je možnost nástroj nasadit jak v cloudu, tak on-premises. Je tak vhodný pro organizace hledající komplexní řešení pro integrační, datovou správu a data governance.

Collibra datový katalog funguje jako samostatný produkt, není přímo spojen s cloudovým ekosystémem a lze ho nasadit i v rámci on-premise. Cílí na širší škálu uživatelů s různými datovými infrastrukturami a specializuje se na data governance.

DataHub jako jediný z výše uvedených nástrojů poskytuje open source řešení. Klade tedy vyšší nároky na technické možnosti a znalosti svých zákazníků, nabízí však návody na svých stránkách a velkou komunitu uživatelů a vývojářů. Nabízí samostatné řešení s předpřipravenými rozhraními na nejvíce používané případy, například skenery na typy databází, které jsou nejčastěji používány.

6. Zdroje

APACHE SOFTWARE FOUNDATION, 2023. *Apache Hive* [online] [vid. 2023-12-02]. Dostupné z: <https://hive.apache.org/>

AREVALO, Isabel, 2023. *Správa dat napříč aktivy Azure SQL pomocí Microsoft Purview - Azure Architecture Center* [online] [vid. 2023-12-01]. Dostupné z: <https://learn.microsoft.com/cs-cz/azure/architecture/solution-ideas/articles/azure-purview-sql-estate-architecture>

BAARS, H., KEMPER, H., 2008. Management Support with Structured and Unstructured Data—An Integrated Business Intelligence Framework [online] [vid. 2018-09-23]. Dostupné z: <https://www.semanticscholar.org/paper/Management-Support-with-Structured-and-Unstructured-Baars-Kemper/2809521eecd9fb9a11332a96bab273bbae62e14>

COLLIBRA, 2023a. *Collibra Data Catalog | Collibra* [online] [vid. 2023-12-01]. Dostupné z: <https://www.collibra.com/us/en/products/data-catalog>

COLLIBRA, 2023b. *Data Citizens drive change with access to the right data* [online] [vid. 2023-12-01]. Dostupné z: <https://www.collibra.com/us/en/company/who-we-are>

DATABRICKS INC., 2021. What is Hadoop Distributed File System (HDFS). *Databricks* [online] [vid. 2023-12-02]. Dostupné z: <https://www.databricks.com/glossary/hadoop-distributed-file-system-hdfs>

DATAHUB, 2023a. *A Metadata Platform for the Modern Data Stack | DataHub* [online] [vid. 2023-12-01]. Dostupné z: <https://datahubproject.io/>

DATAHUB, 2023b. *Deployment Guides | DataHub* [online] [vid. 2023-12-01]. Dostupné z: <https://datahubproject.io/docs/category/deployment/>

DRESCH, Aline, Daniel PACHECO LACERDA a José Antônio Valle ANTUNES, 2015. *Design science research: a method for science and technology advancement*. Switzerland: Springer. ISBN 978-3-319-07374-3.

HAM, Kenneth a Dustin CROCKETT, 2021a. *Evaluation of Data Catalog Software for Hanford Site Environmental Datasets* [online]. PNNL--31960, DVZ-RPT--066, 1832173. [vid. 2023-12-01]. Dostupné z: doi:[10.2172/1832173](https://doi.org/10.2172/1832173)

HAM, Kenneth D. a Dustin L. CROCKETT, 2021b. *Evaluation of Data Catalog Software for Hanford Site Environmental Datasets* [online]. PNNL-31960; DVZ-RPT-066. B.m.: Pacific Northwest National Lab. (PNNL), Richland, WA (United States) [vid. 2023-12-01]. Dostupné z: doi:[10.2172/1832173](https://doi.org/10.2172/1832173)

INFORMATICA, 2023. *Informatica* [online] [vid. 2023-12-01]. Dostupné z: <https://www.informatica.com>

JSON, 2023. *JSON* [online] [vid. 2023-12-01]. Dostupné z: <https://www.json.org/json-en.html>

KLIPFOLIO, 2023. *What is a Data Catalog? Definition, Challenges, and More | Klipfolio* [online] [vid. 2023-12-01]. Dostupné z: <https://www.klipfolio.com/resources/data-stack/data-catalog>

KUBERNETES, 2023. *Production-Grade Container Orchestration. Kubernetes* [online] [vid. 2023-12-01]. Dostupné z: <https://kubernetes.io/>

LUKŠÍK, Marek, 2020. *Datová kvalita a datové katalogy* [online]. Dostupné z: <https://katalog.vse.cz/Record/000658742>

MBI, 2023. *MBI - Management Byznys Informatiky - home page* [online] [vid. 2023-12-01]. Dostupné z: <https://mbi.vse.cz/>

MBI-AF, 2023. *MBI-AF | IT A ANATOMIE FIRMY* [online]. [vid. 2023-12-01]. Dostupné z: <https://www.mbi-af.cz/>

MICROSOFT, 2023a. *Co je SaaS? Software jako služba | Microsoft Azure* [online] [vid. 2023-12-01]. Dostupné z: <https://azure.microsoft.com/cs-cz/resources/cloud-computing-dictionary/what-is-saas>

MICROSOFT, 2023b. *Learn about Microsoft Purview* [online] [vid. 2023-12-01]. Dostupné z: <https://learn.microsoft.com/en-us/purview/purview>

MICROSOFT, 2023c. *Prostředí Integration Runtime - Azure Data Factory & Azure Synapse* [online] [vid. 2023-12-01]. Dostupné z: <https://learn.microsoft.com/cs-cz/azure/data-factory/concepts-integration-runtime>

NEDZELSKÝ, Toman, 2019. *Metoda dynamické alokace lidských zdrojů v projektovém řízení - Vysokoškolské kvalifikační práce - Vysoká škola ekonomická v Praze* [online] [vid. 2023-12-01]. Dostupné z: <https://vskp.vse.cz/77486>

OPENSOURCE.COM, 2023. *What is open source?* | *Opensource.com* [online] [vid. 2023-12-01]. Dostupné z: <https://opensource.com/resources/what-open-source>

POUR, Jan, 2017. *Prediktivní analytika v podnikové informatice* | *Česká společnost pro systémovou integraci* [online] [vid. 2023-12-01]. Dostupné z: <http://old.cssi.cz/prediktivni-analytika-v%C2%A0podnikove-informatice>

REDHAT, 2020. *Red Hat OpenShift enterprise Kubernetes container platform* [online] [vid. 2023-12-01]. Dostupné z: <https://www.redhat.com/en/technologies/cloud-computing/openshift>

SHEVEK, 2021. *What Data Catalogs Don't Reveal*. *DevOps.com* [online] [vid. 2023-12-01]. Dostupné z: <https://devops.com/what-data-catalogs-dont-reveal/>

TRANSITION SUPPORT, 2023. *Quality - what it means* [online] [vid. 2023-12-01]. Dostupné z: https://transition-support.com/what_is_quality.html

VACEK, Martin, 2014. *Metriky, monitoring a řídicí proces Data Governance - PDF Stažení zdarma* [online] [vid. 2023-12-01]. Dostupné z: <https://docplayer.cz/10899103-Metriky-monitoring-a-ridici-proces-data-governance.html>

VONDRA, Marek, 2023. *Lekce 1 - Docker - Teorie a instalace* [online] [vid. 2023-12-01]. Dostupné z: <https://www.itnetwork.cz/docker-teorie-a-instalace>

WELLS, Dave, 2020. *What Is a Data Catalog? - Importance, Benefits & Features* | *Alation* [online] [vid. 2023-12-01]. Dostupné z: <https://www.alation.com/blog/what-is-a-data-catalog/>

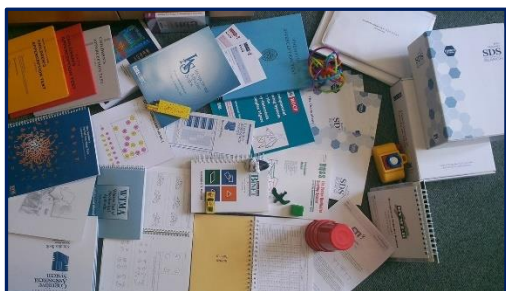
C. Metodika datové migrace

(pracovní dokument)



Marek Slezák

VŠE Praha, 2023



[1] Metodika a postup datové migrace	
[2] Nástroje pro datovou migraci	[3] Práce s daty
[4] Testování	[5] Rozpočet
[6] Přehled zdrojů	



Účelem tohoto dokumentu je poskytnout **přehled všech podstatných doporučení a návodů** ve vztahu k řízení a řešení **datové migrace** v rámci zejména aplikačních projektů. Dokument představuje speciální dílčí **dokument k základnímu dokumentu Řízení IT** na portále MBI-AF. Dokument vychází **z celé řady zdrojů obsažených v závěrečném přehledu** a na bázi příslušných citací vybírá z nich ta nejpodstatnější doporučení.

1. Metodika a postup datové migrace

1.1 Používejte formální metodiku, která byla vyzkoušená a otestovaná [1]

Tento osvědčený postup se podílí na úspěšném dokončení projektů. Rozlišovat také musíme skutečnost, zdali se jedná o metodiku interní, která byla vytvořena vlastními silami, nebo metodiku, kterou nám poskytl systémový integrátor, či dodavatel. Metodiky poskytované systémovými integrátory a dodavateli mají totiž větší úspěšnost než metodiky interní, což může být dáno skutečností, že jsou interní metodiky méně otestované [1].

Využívejte proto metodiky, které již byly ověřeny. Metodiku můžete také stanovit na základě předchozích zkušeností z migračních projektů, pokud jimi disponujete. **Výhodou migračních projektů** je skutečnost, že i přes to, že se liší v různých oblastech, **jádro zůstává pořád stejné**, a tudíž i postup se diametrálně neliší, a proto můžete stejný, či obdobný postup využít také příště. Počítejte ale také s tím, že v určitých oblastech budete kvůli různým požadavkům zainteresovaných stran a dalším skutečnostem nuceni metodiku modifikovat. **Na závěr každého projektu si definujte tzv. lessons learned** – co fungovalo, a co naopak nefungovalo a je potřeba do budoucna zlepšit. Také to vám může pomoci vylepšovat metodiku pro hladší průběh budoucích projektů.

1.2 Metodika migrace dat by měla postupovat od posouzení, přes plánování až po přesun a validaci [12]

Jak již bylo zmíněno, snažte se využívat metodiku, která byla otestována. Váš **postup by měl zahrnovat plánování**. Během této fáze společně s dalšími zainteresovanými stranami **naplánujte, jak dojde k přesunu dat ze zdrojového do cílového prostředí a vytvořte projektový plán**, který budete v průběhu projektu dodržovat. Ve fázi plánování také stanovte, jaký hardware nebo software budete k migraci potřebovat, jaká bude architektura migrace, specifikujte zdroje, postupy migrace a plány nasazení a testování. Do plánování zahrňte nejen pracovníky vašeho týmu, ale také další strany, které se na projektu podílí. Také ostatní strany budou mít v rámci projektu zodpovědnost za vykonávání některých činností, proto je důležité, aby měly o projektovém plánu přehled.

Ve druhé fázi se zaměřte na **přesun dat ze zdrojového do cílového prostředí**. Získejte, nainstalujte a nakonfigurujte potřebný software, který budete k migraci potřebovat, získejte data a analyzujte zdrojové a cílové struktury, a začněte tvořit transformační kód, který data transformuje do cílových struktur. Počítejte také s tím, že vám fáze transformací dat pravděpodobně zabere nejvíce času společně s následným testováním dat. Nepodceňte testování, které může výrazně ovlivnit průběh projektu. V první řadě proveďte systémové testování, dále poté uživatelské testování.

Validace by pak měla být poslední fází, během které se ověří, **zda byla data úspěšně přesunuta ze zdrojového do cílového prostředí** podle plánu. Pokud se na projektu podílíte jako externí systémový integrátor, nezapomeňte připravit pro vedení společnosti, pro kterou jste data migrovali, prezentaci se shrnutím projektu. V závěru projektu také vypracujte dokumentaci a předejte ji klientovi. Získané zkušenosti zpracujte během tzv. lessons learned, kdy si stanovíte, co během projektu fungovalo, co nefungovalo a můžete zlepšit do dalšího projektu. Na základě těchto zjištění můžete modifikovat metodiku pro další projekty.

1.3 Mějte plán [12] [15]

Plánování migrace může pomoci **určit, kde by se mohly vyskytnout problémy a způsoby**, jak se jim vyhnout. Mějte plán a plňte ho. Plánujte migraci dat (nebo integraci) jako projekt a sledujte svůj postup podle plánu. **Plánujte neúspěch**; plánujte špatné požadavky; plánujte klienty, kteří nereagují nebo jsou nerozhodní. Plánujte zkrátka vše. [15]

Zároveň se snažte mít **projektový plán co nejvíce konkretizovaný** tak, aby obsahoval kompletní seznam všech činností, které je potřeba vykonat, spolu s časem potřebným pro jejich provedení a identifikací odpovědných osob. Právě projektový plán vám umožní všechny tyto skutečnosti vidět.

Počítat je také nutné se situacemi, ve kterých může dojít ke změně požadavků a projektový plán tak bude muset být přizpůsoben novým podmínkám. Počítejte proto s tím, že změny mohou nastat také v průběhu projektu, a to nejen v důsledku předem špatně stanovených požadavků, ale také například v souvislosti s opomenutím některých činností nebo v důsledku změn požadavků a rozhodnutí zainteresovaných stran.

1.4 Mějte dostatek času na plánování projektu [15][16]

Je důležité, abyste si **na plánování projektu vyhradil dostatek času**. Plánování totiž může ovlivnit průběh celého projektu, a proto by neměly být opomenuty žádné kritické kroky. Fáze plánování by měla zahrnovat jednotlivé kroky, které je během průběhu migračního procesu provést, a zohlednění zdrojů odpovědných za jednotlivé úkoly [16].

Pokud je v projektu datové migrace **zainteresováno více stran, nezapomeňte je do procesu tvorby projektového plánu zapojit**. Každá strana bývá zodpovědná za provedení různých kroků v rámci projektu. Proto je klíčové, aby se všechny strany aktivně zapojily do tvorby projektového plánu a byly informovány o svých odpovědnostech. Nezapomeňte také všem stranám zajistit přístup k projektovému plánu, díky čemuž budou moci pravidelně monitorovat jeho postup.

1.5 Zkontrolujte, zda jsou k dispozici potřebné zdroje [16] [19]

Zajištění dostupnosti zdrojů, které jsou zodpovědné za jednotlivé úlohy, **hraje klíčovou roli** během projektu. Klíčové zdroje by měly být zapojeny od začátku projektu a neměly by se jeho součástí stávat až v jeho průběhu. Pokud je do migrace zapojeno více pracovišť, je potřeba, aby se zapojily zdroje z každého pracoviště v případě, že mezi pracovišti existují rozdíly [16].

V případě, že vás v průběhu migračního projektu opustí některý člen týmu, je nezbytné nalézt kvalifikovanou náhradu. Současně zajistěte, aby byl **nový člen týmu plně seznámen se všemi aspekty** projektu a jeho rolí v něm. Pro tyto účely mu poskytněte odpovídající školení nebo konzultace. Je důležité počítat s možností náhrady člena týmu a také s tím, že následné zaškolení nového člena bude vyžadovat čas a úsilí. Nepodceňte tuto činnost, jelikož správné zaškolení nového člena týmu je klíčové pro zachování kontinuity a efektivitu projektu.

Zároveň zajistěte, aby byly **od počátku projektu k dispozici zdroje, od jejichž činností se odvíjejí další kroky**. Například, pokud je některý člen týmu odpovědný za implementaci prostředí, musí tuto činnost dokončit v předem stanoveném termínu, aby nezdržoval postup dalších kroků projektu, které jsou na implementaci prostředí závislé.

1.6 Zdokumentujte proces migrace [12][15]

Zdokumentujte celý proces migrace. To vám pomůže vyškolit zaměstnance, předat know-how a zjednodušit nebo zefektivnit další projekt datové migrace, což pomůže snížit náklady i rizika. Tvorbou dokumentace rovněž umožníte přístup k informacím o provedené migraci a poskytnete auditorům komplexní přehled o průběhu projektu.

Pokud se na projektu podílíte jako externí systémový integrátor, **předejte vytvořenou dokumentaci se v závěru projektu klientovi**. Tato dokumentace klientovi poskytne odborné znalosti, které může využít a aplikovat v případě budoucích projektů datové migrace. Snažte se také dokumentaci předat ve stručné a srozumitelné formě, aby byla čitelná a použitelná. Díky tomu může dokumentace sloužit také pro další zlepšování procesů datové migrace.

1.7 Vyviňte interní kompetence ohledně migrace dat a nespolehejte zcela na externí zdroje [1]

Pokud jako společnost nedisponujete dostatečným know-how nebo zdroji pro provedení migrace, můžete **oslovit externího systémového integrátora**, který má s podobnými nebo stejnými projekty zkušenosti, a který migraci dat provede za vás. I přesto je však důležité, abyste byly součástí migračního

projektu od začátku do konce, jelikož i vy budete mít v rámci projektu zodpovědnost za vykonávání některých činností. Díky tomu budete mít zároveň přehled o průběhu projektu. Vaše zapojení do projektu je klíčové a je nutné, abyste aktivně komunikovali a spolupracovali s ostatními zainteresovanými stranami v projektu.

1.8 Podnik musí být zapojen ve všech fázích projektu, od zahájení až po testování [1] [6]

Jedná se podle [1] o **nejkritičtější faktor úspěchu**. Zákazníkům, produktům nebo službám totiž rozumí mnohem více byznys než IT. Pokud se pracovníci těmito vztahy a jejich zachováním nezabývají, může projekt selhat. Podobně hovoří také [6], podle které je třeba **určit, kdo rozumí obchodním datům**, se kterými se pracuje. Je proto důležité zapojit také jiné podnikové útvary a lidi, kteří data používají [6].

Je třeba si uvědomit, že pokud v projektu vystupuje **více zainteresovaných stran, není podnik samotný jedinou stranou, která by do projektu migrace měla být zapojena**. Z hlediska jeho úspěšného dokončení by do něj měly být aktivně zapojeny všechny strany, které se na projektu podílí, jelikož každá z těchto stran je v projektu zodpovědná za vykonávání některých činností. Zajistěte také, aby byly všechny strany zapojeny od začátku projektu, díky čemuž budou mít dostatečný přehled a rozumět kontextu.

1.9 Získejte písemný závazek vendora [16]

Pro zajištění úspěchu v projektu je důležité, aby byl **zapojen také dodavatel systému**, pokud jeho služeb využíváte. Dodavatel systému, do kterého data migrujeme, by měl být přítomen po celou dobu projektu [16]. Zároveň zajistěte, aby byl dodavatel zapojen od samého počátku projektu a mohl se tak věnovat svým činnostem. Zajistěte také **pravidelný kontakt s dodavatelem**, nechte se informovat o postupu na jeho straně a informujte ho o postupech na vaší straně.

Pokud se na projektu podílíte **jako externí systémový integrátor**, podpis o spolupráci s dodavatelem systému pravděpodobně bude mít váš klient, který jeho služeb využívá. Proto v případě, že vendor neplní závazky, které ze smlouvy vyplývají, reportujte tuto skutečnost klientovi, který s vendorem smlouvu uzavřel, a který následně stanoví rozsah škod a jejich náhradu. V případě, že klienta o problémech nebudete informovat, sami se do nich můžete dostat, což zkomplikuje průběh migrace.

1.10 Odstraňte komunikační bariéry [16]

Mnoho problémů, které během projektu vznikají, jsou způsobeny komunikačními bariérami – **špatné pochopení problému, nejasné vysvětlení, jazykové rozdíly, nejasnosti v odpovědnostech, nedostatečné informace** atd. [16]. Odstraňte proto komunikační bariéry mezi všemi členy týmu a zároveň mezi všemi stranami projektu. To vám pomůže usnadnit průběh migračního projektu.

V případě, že nerozumíte významu dat nebo způsobu, jak data mapovat, **komunikujte s relevantními stranami a diskutujte řešení problému**, aniž byste se pokoušeli o vlastní řešení. Následné chyby mohou mít negativní vliv na průběh celého projektu.

Pokud pracujete **na mezinárodním projektu**, popřípadě na projektu, kde se používá jiný jazyk než váš mateřský jazyk, zajistěte, že **členové týmu ovládají potřebný jazyk** a jsou schopni aktivně komunikovat s ostatními stranami a řešit případné problémy.

Na závěr každého meetingu shrňte, co bylo na dané schůzce probráno, jaké jsou následující kroky, a kdo je za ně zodpovědný. Dělejte si také během každé schůzky poznámky, které pak mohou sloužit jako vodítko nebo důkaz, pokud některá ze zainteresovaných stran bude tvrdit něco jiného, než tvrdila minule.

1.11 Automatizujte, pokud je to možné [15]

Vše, co lze automatizovat, by mělo být automatizováno. **Automatizace zahrnuje všechny přesuny a transformace dat, včetně přesunu původních zdrojových dat do stagingového prostředí.** Automatizujte úlohy, které provádíte opakovaně se stejným postupem, včetně přípravných prací po dodání nových dat, jako je kontrola datové kvality nebo aplikace transformací, abyste minimalizovali riziko chyb a zvýšili efektivitu procesu.

Pro dosažení automatizace můžete **využít vytváření skriptů nebo procedur, které umožní automatické nahrávání, či formátovat** potřebných dat podle stanovených pravidel. Tím ušetříte čas a minimalizujete riziko chyb, které by mohly vzniknout při ručním zpracování dat.

2. Nástroje pro datovou migraci

2.1 Používejte nástroje pro profilování dat před zahájením projektu i v jeho průběhu [1]

Když společnost **Bloor Research provedla v roce 2011 průzkum** mezi společnostmi, které se projekty datové migrace zabývají, zjistila, že zhruba **polovina společností, které profilování využily, tak učinily před stanovením rozpočtu projektu.** Z této poloviny poté 72% společností dokončilo projekt včas a v rámci rozpočtu. Naopak společnosti, které profilování využily později, měly úspěšnost dokončení pouze 52% z nich. Jednou z hlavních příčin byl uveden nedostatečný přehled o problémech s kvalitou dat, který profilování dat poskytuje [1].

Profilovací nástroje vám pomohou porozumět datům a ušetřit velké množství času. Doporučuji používat profilovací nástroje také v případě, že migrujete menší množství dat. Díky nim lépe porozumíte kvalitě dat, se kterými pracujete, čímž se budete schopni vyvarovat chybám v datech, jako jsou duplicity nebo chybějící hodnoty, a nutnosti jejich oprav během testování. Využíváním nástroje pro profilování dat získáte také dostatečný přehled o datech, se kterými pracujete, díky čemuž můžete odstranit chyby a zvýšit tak jejich kvalitu ještě před samotným transformováním.

2.2 Pro čištění dat používejte nástroje pro čištění dat. Nepokoušejte se o čištění ručně [1]

Studie společnosti Bloor Research [1] uvádí, že společnosti, které v dané studii **během řešení migračního projektu nevyužívali nástroje pro čištění dat, měli pouze 55% šanci na úspěšné dokončení** projektu. Jak již bylo zmíněno v předchozí kapitole, špatná kvalita dat je jednou z příčin neúspěšně dokončeného dokončení projektu. Tyto nástroje vám mohou pomoci vyčistit data od duplicit nebo chybných formátů dat, a zvyšovat tak jejich kvalitu.

Abyste mohli s fází čištění začít, je nejprve nutné **získaná zdrojová data analyzovat.** Důkladnou analýzou můžete odstranit mnoho chyb v datech již v jejich počátcích. S čištěním dat tak začnete co nejdříve. Chyby, které v datech v počátcích neodstraníte se následně mohou objevovat během fáze testování, což může prodloužit tuto fázi a být tak v rozporu s projektovým plánem.

Důležité také je, abyste data **nečistili dle vlastního uvážení, pokud jim dostatečně nerozumíte.** Pokud v projektu vystupujete například jako externí systémový integrátor, většinou nemáte dostatečný přehled o byznysovém významu dat, se kterými pracujete. Je proto vhodné dohodnout se se zaměstnanci klienta, kteří datům, se kterými pracují, rozumí, jakým způsobem data čistit. Nikdy nečistěte data dle svého vlastního uvážení.

V neposlední řadě **využívejte nástroje pro čištění dat** a nepokoušejte se o manuální čištění. Zejména při zpracování rozsáhlých datových souborů vyžaduje manuální čištění nadměrnou časovou náročnost

a je náchylnější ke vzniku chyb. V případě, že by došlo k migraci více sad dat, bylo by nutné každou tuto sadu ručně čistit. Tento proces pomohou nástroje pro čištění dat usnadnit.

2.3 **Používejte nástroje pro integraci dat. Nepokoušejte se o ruční kódování [1] [13]**

Podobně jako v případě nástrojů pro čištění dat, by měly být **používány také nástroje pro integraci dat**. Tyto nástroje umožní automatizovat některé úlohy a vyhnout se tak ručnímu kódování, které může být zdouhavější, náročnější na provedení, a náchylné na chyby. Společnosti, které integrační nástroje ve studii [1] nevyužívaly, měly pouze 50% šanci na úspěšně dokončení projektu.

Nástroje pro datovou integraci mají v migračním projektu svoji nezpochybnitelnou roli. Ruční kódování s sebou nese riziko v podobě nesouladu a obtížné údržbě kódu. Každý vývojář v tomto případě vytváří kód podle svého uvážení, což může vést k nekonzistenci výsledného produktu.

Využívejte proto **centrální nástroje pro datovou integraci**. Ty vám navíc mohou pomoci generovat metadata a šablony stejném stylu pro každý mapping, což vám zajistí konzistenci a snadnou orientaci, i když bude každý mapping vytvářet někdo jiný. Díky tomu budete moci kontrolovat a upravovat práci kolegů, když to bude potřeba. Využíváním integračních nástrojů dojde nejen k úspoře času, ale také k minimalizaci rizika vzniku chyb a větší konzistenci mezi jednotlivými mappingy, což oceníte například při jejich kontrole.

2.4 **Ujistěte se, že pro migraci dat používáte správný software [12] [15] [17] [19]**

Produktů, které je možné k migraci využít, je celá řada. Každý produkt obsahuje různé funkcionality a nabízí se za různé ceny. Důkladně proto zvažte, který softwarový produkt nejlépe odpovídá vašim požadavkům.

Existují případy, kdy mají **společnosti vyvinuté vlastní nástroje pro tyto účely**. Zaměstnanci jsou s těmito nástroji seznámeni, což následně usnadňuje jejich použití během migračního projektu. Ne vždy je ale vhodné interní nástroj pro migraci dat využít, jelikož ne vždy tento nástroj splňuje všechny vaše požadavky. Pokud tomu tak je, vyberte jiný nástroj, který je na trhu k dispozici.

2.5 **Je důležité zvážit, zda je možné migraci snadno ukončit a znova spustit [12]**

Požadavkem, který je třeba zvážit, je **možnost vrátit migraci zpět**. V podstatě jde o to, že pokud se během migrace něco pokazí, je možné migraci ukončit a spustit znovu. To může být u některých technologií problematické, zatímco u jiných je to snadné [12].

Během migračního projektu je **zásadní, abyste měli možnost provést rollback migrace**. Existuje několik způsobů, jak toho dosáhnout. Pravidelně si vytvářejte **zálohu migrační databáze**, abyste v případě jejího smazání nebo poškození mohli data obnovit ze zálohy. Zároveň ukládejte různé verze záloh a nepřepisujte stále jen jednu zálohu. Tím získáme více možností pro provedení rollbacku v případě potřeby.

Dalším způsobem, který vám umožní udělat částečný rollback, je **ukládání transformačních skriptů pomocí verzování**. Při každé změně transformačního kódu ukládejte tyto změny do nové verze mapování a ponechte si verze původní. V případě, že vaše poslední verze mapování způsobí v databázi nežádoucí změny v datech, můžete využít verze původní a obnovit data do původního stavu. Pokud váš migrační nástroj umožňuje tuto funkcionalitu, je to nesporná výhoda.

3. Práce s daty

3.1 Pochopte data ve zdrojovém systému ještě předtím, než začnete migrovat [6] [15]

Společnosti často podceňují důležitost posouzení stavu dat před samotným procesem migrace. **Datová kvalita má významný vliv na úspěch migrace, a proto plně pochopte zdrojová data a jejich struktury, než začnete s jejich transformací.** Ptejte se, komu budou data sloužit, co s nimi hodlají uživatelé dělat, či zda v nich něco nechybí. To může výrazně usnadnit následný proces migrace. Správnost mapování dat do cílového systému závisí kromě mnoha dalších faktorů právě na pochopení, a proto tento krok nepodceňte.

Jedním z prvních kroků po získání zdrojových dat je jejich analýza. Zjistěte množství a účel entit, jejich atributy, vazby mezi entitami, formáty dat, primární a cizí klíče, datové typy nebo datové objemy. Pokud si něčím nejste jisti nebo nechápete význam poskytnutých dat, vždy tyto skutečnosti konzultujte a nepokoušejte se o vlastní řešení. To platí také v případě, pokud jste externím systémovým integrátorem, který nemá hlubší byznysové porozumění migrovaných dat a plně nechápe jejich využití. **Vždy proto identifikujte klíčové osoby, které migrovaným datům rozumí** a snažte se správně pochopit význam dat, aby došlo k jejich správné migraci. Nesprávné pochopení dat povede k chybné transformaci a následným problémům během testování.

Pouhé pochopení zdrojových dat tedy nestačí. **Pochopit musíte také struktury cílového systému,** do kterých data migrujete. Pokud budete chápat pouze zdrojové struktury, nikoliv však cílové, je málo pravděpodobné, že se vám transformace dat povede tak, jak očekáváte. V případě nejasností vše konzultujte s klientem nebo dodavatelem systému, kteří mají hlubší povědomí o datech.

3.2 Zaměřte se na správné mapování dat, aby byl zajištěn kompletní přenos dat [16]

Mapování je proces, během kterého **se přiřazují atributy původních datových struktur k atributům cílových datových struktur.** Kromě správného mapování zjistěte, zda se shodují datové typy, délky znaků jednotlivých atributů nebo povinnosti atributů. Kvalita mapování má přímý vliv na kvalitu výsledných dat a platí, že čím kvalitněji mappingy provedete, tím menší je pravděpodobnost výskytu chyb v datech.

Správné mapování dat je úzce souvisí s jejich správným pochopením. Při mapování je důležité mít detailní znalosti o struktuře a významu původních dat, stejně jako o očekávané struktuře a požadavcích cílového systému. Proveďte proto před samotným mapováním důkladnou analýzu datových struktur.

Kvalitu mapování nelze jednoznačně posoudit pouhým pohledem na mapované atributy, a proto je nezbytné **provést důkladné testování.** Přesto je možné snížit riziko nesprávného mapování. Pokud se objeví situace, kdy není jasné, jak daný atribut mapovat do cílového systému, konzultujte tuto skutečnost s někým, kdo má znalosti týkající se migrovaných dat a cílových struktur. Tyto kroky mohou přispět k minimalizaci rizik spojených s nesprávným mapováním atributů.

3.3 Rámec pro řízení dat má zásadní vliv na celý migrační projekt [1] [6]

Je vhodné, aby společnosti měly **zavedený rámec pro řízení dat bez ohledu na to, zda souvisí s migračním projektem** [1]. Pokud nemáte takový rámec, může pro vás projekt migrace představovat dobrý odrazový můstek. Kvalitní rámec pro řízení dat pro vás bude nespornou výhodou, pokud se k datové migraci odhodláte. **Rámec vám umožňuje vyšší kvalitu dat,** což má velký vliv na průběh migračního projektu. Bez rámce pro řízení dat je pravděpodobné, že se bude kvalita dat časem zhoršovat. To bude v případě migračních nebo jiných projektu vyžadovat větší úsilí během jejich čištění. Mějte proto zavedený rámec pro řízení dat a pokud ho ještě nemáte, zvažte jeho implementaci. Implementace rámce pro řízení dat už je ale na samostatný projekt, který vyžaduje specifické požadavky.

3.4 Pravidelně zálohujte data pro případ, že by během procesu migrace došlo k problému [17] [19]

Záloha dat zajišťuje, že pokud během datové migrace **dojde k problémům, či chybám, které způsobí ztrátu dat, zůstanou i přesto tato data bezpečně zálohována** a dostupná pro další práci. Zálohu dat provádějte pravidelně během migračního projektu. Nemělo by se vám stát, že poslední záloha je několik dní nebo týdnů stará. Vytvářejte si také různé verze záloh a nepřepisujte jednu zálohu stále dokola. V případě potřeby tak můžete vybrat, jakou zálohu obnovíte.

Pokud to váš migrační nástroj umožňuje, **zálohujte také transformační kód pomocí verzování**. Tím získáte různé verze transformačního kódu, které je možné v případě potřeby obnovit, stejně jako v případě zálohování databáze. Zálohování transformačního kódu je klíčové pro minimalizaci dopadů chyb či ztráty dat.

Provádějte také kontroly zálohy dat a testujte její obnovení, abyste ověřili, že jsou zálohy správně vytvořené a fungují, jak mají. To pomáhá minimalizovat riziko překvapení během obnovy.

3.5 Identifikujte formáty dat, které budete přenášet [17] [19]

Během migračních projektů se může stát, že datové typy ve zdrojovém a cílovém systému nejsou shodné. Může také nastat situace, kdy zdrojový atribut povoluje více znaků než atribut v cílovém systému, což může způsobit problémy při přenosu některých dat. Proveďte proto **analýzu datových struktur v počáteční fázi migračního procesu, abyste identifikovaly formáty, umístění, citlivost dat, entity**, atributy, vazby mezi entitami, primární a cizí klíče, datové typy, objemy, se kterými budeme pracovat.

Pokud se objeví situace, kdy například zdrojový atribut neodpovídá formátu nebo pravidlům atributu v cílovém systému, **řešte co nejrychleji tyto problémy ve spolupráci s ostatními stranami**, které se na projektu podílejí, a diskutujte vhodné řešení. Nepokoušejte se o řešení sami bez předchozí konzultace.

3.6 Nahrajte data, aby bylo možné zahájit proces migrace [17]

Aby bylo možné spustit proces migrace, je nezbytné **zajistit dostupnost dat, která budou přenášena do nového systému**. Data musí být získána s ohledem na jejich citlivost a zabezpečení. Domluvte si vhodný způsob předávání dat, který minimalizuje riziko změny, poškození nebo ztráty. Nezapomeňte také připravit prostředí, do kterého budou data nahrána.

Pochopte entity ve zdrojovém a cílovém systému, jejich atributy, vazby mezi entitami, formáty dat, primární a cizí klíče, datové typy a objemy dat. To je nezbytné pro správnou analýzu dat. V případě nejasností komunikujte problémy se všemi relevantními stranami projektu a vyjasněte si je.

3.7 Investujte do následné údržby, abyste se ujistili, že byla zachována integrita dat [17]

Tento postup neplatí, pokud jste například externí systémový integrátor, který daný systém nespravuje. Pokud však jako klient nebo dodavatel **vlastníte informační systém, do kterého byla data migrována, je vhodné po dokončení migrace zajistit, aby migrovaná data zůstala v bezproblémovém stavu** a nedošlo k jejich ztrátě. Pro minimalizaci tohoto rizika investujte do následné údržby systému. To umožní zlepšení a optimalizaci výkonu, zajištění bezpečnosti, odstranění chyb a prodloužení životnosti systému. Je také vhodné vytvořit rámec pro řízení dat v případě, že ho ještě nemáte. To pomůže stanovit pravidla pro ukládání dat a zvýšit jejich kvalitu.

3.8 Omezte počet zprostředkovatelů (vrstev abstrakce) [15]

Při přesunu dat z jednoho systému do druhého nebo z jednoho formátu do jiného hrozí, že se něco ztratí nebo změní. Pokud jsou například data v souborech .xlsx, vy je otevřete v aplikaci Excel a

rozhodnete se je uložit ve formátu .csv, aplikace Excel může vypustit počáteční nuly nebo převést text na čísla a přidat desetinná místa. Výsledkem tak mohou být soubory, které nejsou správně naformátované. **Vyhnete se proto otevírání dat v kancelářských balíčcích. Zároveň požadujte, aby byla data dodávána v původním formátu a snažte se omezit počet zprostředkovatelů.** To je jedním z klíčových faktorů pro udržení dat v původní podobě.

Požádejte také společnost, která zdrojová data posílá, o **vytvoření a zdokumentování postupu exportu dat a posílání**. To umožní standardizaci a minimalizaci rizika změn v situacích, kdy data exportuje někdo jiný. Pokud by se však stalo, že data v souboru obsahují jiné formáty než posledně, reportujte tuto skutečnost společnosti, která vám data poskytl.

Také vy si **vytvořte dokumentaci pro nahrávání datových souborů do databáze**. Tímto krokem zajistíte standardizaci postupu při nahrávání dat. S pečlivě zdokumentovaným postupem budete vždy postupovat stejně a zajistíte, že data budou nahrána tak, jak potřebujete.

3.9 Vytvořte si knihovnu opakovaně použitelného kódu [15]

Některé transformace se musí provádět opakovaně, a to nejen v rámci jednoho projektu. Pište proto **kód tak, aby se dal používat opakovaně**. Příkladem takové transformace je formátování telefonních čísel [15]. Vytvoření knihovny opakovaného kódu vám ušetří mnoho času a pomůže minimalizovat chyby. **Knihovna může obsahovat kódy pro různé části projektu, včetně skriptů pro nahrávání a transformaci dat**, funkcí pro změny formátování záznamů v atributech, nebo kódů pro spouštění procedur. Díky opakovaně použitelnému kódu snadněji udržíte konzistenci v transformacích dat a budete schopni aplikovat změny na více místech najednou.

3.10 Opravujte kód, ne data [15]

Během procesu migrace je možné narazit na data, která je nutná opravit. **Opravujte je pomocí transformačního kódu, nepravujte data samotná**. To zajistí, že v případě nové migrace budou nová data správně transformována a nemusí tak docházet k jejím dalším opravám [15]. Tento přístup nejen šetří čas a snižuje riziko chyb, ale také zajišťuje snadnější sledování provedených úprav. Pokud byste upravovali přímo zdrojová data, mohlo by se časem stát obtížným identifikovat, která konkrétní data byla upravena.

V souvislosti s tím je také důležité uvést, **abyste neupravovali kód tak, aby jeho podmínky platily pro konkrétní výjimky ve zdrojových datech**. Naprogramujte obecnější podmínky, které lze aplikovat na každý záznam atributu. Pokud se v migrovaných datech objeví například některé emailové adresy obsahující dva zavináče, není vhodné vytvářet v transformačním kódu výjimku pro tyto konkrétní emailové adresy, ale pro všechny adresy, které mohou dva zavináče obsahovat. Tento přístup vám umožní snadnou aplikaci transformačního kódu na různé sady dat.

4. Testování

4.1 Proved'te v cílovém systému testování, abyste se ujistili, že byla data úspěšně přenesena [17] [19]

Testování tvoří podstatnou část migračního procesu a nemělo by být opomíjeno. Právě **testování ověřit, že byla data namigrována do cílového systému v požadovaném stavu a neobsahují chyby, duplicitu**, neodpovídající datové typy, které by mohly mít negativní vliv na následné rozhodování podniku. Zajistí se také to, že budou migrovaná data bezpečná a na správném místě [19].

Testování je fází, kterou se nevyplatí podceňovat a vždy si pro ni **vyhrad'te dostatek času**. Množství chyb, které se v datech mohou objevit, ovlivňuje množství nutných oprav, s čímž je nutné počítat.

Testování samotných dat ne vždy ověří, zda byla data migrována správně do cílového systému a zda jsou v požadovaném stavu. Samotná data se mohou zdát na první pohled v pořádku, nemusí však odpovídat kontextu ostatních dat. **Testujte proto data v kontextu zbylých dat.** Dále testujte byznys procesy nad těmito daty. Bez těchto činností nepoznáte, zda jsou migrovaná data v požadovaném stavu.

Důležité je, abyste vy, nebo jakákoliv strana zodpovědná za testování, měly **připravené vhodné testovací prostředí, které co nejvíce odpovídá reálným podmínkám cílového systému.** Tato skutečnost vám pomůže odhalit chyby v datech a zjistit, zda jsou data migrována správně.

4.2 Pravidelně během migračního projektu testujte [16]

Testování je dalším krokem, který hraje v migračním projektu významnou roli, jelikož **pomáhá identifikovat a odstranit chyby v datech, které ovlivňují datovou kvalitu.** Může se stát, že nad daty v původním systému nebyl aplikován rámec řízení dat, který se ovšem v cílovém systému nachází. Vznikají tak nové požadavky na migrovaná data a je potřeba otestovat, zda data tyto požadavky splňují.

Začněte také **s testováním co možná nejdříve a vyhrad'te si na něj dostatek času.** Během testovací fáze je pravděpodobné, že se v datech objeví velké množství chyb, které bude následně nutné opravit. To může mít za následek prodloužení fáze testování a následný časový skluz.

Vytvořte si také **jednotné místo, kam se budou ukládat veškeré nalezené chyby (tzv. bugy).** Neposílejte si informace o chybách v datech a jejich řešení prostřednictvím emailů. Brzy se v těchto vláknech začnete ztrácet a přestanete tak mít přehled o řešených chybách. Použití specializovaných nástrojů pro správu chyb umožní sledování, přiřazování odpovědností, prioritizaci a zaznamenávání stavu řešení chyb. Tím zvýšíte přehlednost a efektivitu během průběhu testování.

Vytvořte **po každé opravené chybě kód, který ověří, zda byla chyba v datech opravena** a následně vytvořte tabulku, která bude sloužit jako souhrn těchto kódů, ověřujících, zda jsou data opravena. Tímto postupem získáte přehled o tom, zda se opravené chyby v datech nevyskytují znova.

4.3 Musí být provedeno systémové a uživatelské testování a před spuštěním ostrého provozu je nutné získat podpis [15]

Fáze testování by měla být **rozdělena na systémovou a uživatelskou část. Jako první proved'te testování systémové,** kdy ověříte, zda jsou data správně integrována do cílového systému, a respektují jeho strukturu a pravidla. Jakmile je systémové testování dokončené, přejděte na testování uživatelské. To mohou provádět pouze skuteční uživatelé systému. **Uživatelské testování je důležité, protože uživatelé mají přímý kontakt s daty a mohou odhalit chyby a nedostatky,** které by jinak mohly být přehlédnuty.

Před spuštěním samotné migrace je nezbytné, aby proběhlo **její schválení a došlo k podpisu zúčastněných stran projektu,** kterými mohou být například zástupci klienta a dodavatelé. Podpisem dojde k potvrzení, že zúčastněné strany souhlasí s migrací dat a uvedením systému do ostrého provozu. Podpis zároveň zajišťuje, že jsou všechny strany o datové migraci informovány a souhlasí s ní.

Důležité je uvědomit si, že pokud data nebudou odpovídat požadavkům klienta a dodavatele systému (v případě, že je v projektu přítomen), může se stát, že souhlas k migraci nedají, například z důvodu chyb v datech, a migrace neproběhne. Zajistěte proto, **aby byla data v požadované kvalitě, a to ideálně v termínu, který je pro migraci stanoven.**

4.4 Plánujte prodloužené období systémových a uživatelských testů [15]

V rámci testování může být nalezeno mnoho chyb. Zároveň je těžké odhadnout, kolik času budou mít uživatelé v rámci uživatelského testování. **Naplánujte proto dostatečný čas na celou fázi testování, včetně systémového a uživatelského testování.** Pokud migrujete data do systému dodavatele, je důležité, aby také on byl na systémové testování připraven a vyhradil si dostatek času. **Informujte také uživatele o začátku uživatelského testování a zahajte testování co nejdříve,** aby bylo možné včas

odhalit všechny chyby. V neposlední řadě si naplánujte si prodloužené období pro systémové a uživatelské testování.

4.5 Vytvořte si vztah s uživateli, kteří testují [15]

Pokud se data migrují do vašeho systému a nacházíte se ve fázi uživatelského testování, je důležité **vysvětlovat uživatelům, co děláte a proč se ptáte na určité otázky**. Projevujte zájem o problémy uživatelů a ochotu jim pomoci [15]. Pokud se na projektu podílíte jako externí systémový integrátor, máte o starost méně. V kontaktu s uživateli je především strana, do jejichž systémů jsou data migrována. Vytvořte si **komunikační linku s klientem, který s vámi bude sdílet informace od uživatelů**. V případě potřeby je důležité, abyste klienta v komunikaci podporovali, například vysvětlováním nebo řešením vzniklých problémů.

4.6 Ujistěte se, že váš kód loguje vše, co dělá [15]

Definujte, **co se bude během projektu datové migrace logovat**. Zároveň se ujistěte, že kód zaznamenává vše, co dělá: kdy se proces spustí, kdy se dokončí, počty zpracovaných záznamů, ale také informace o tom, jaké kroky byly v průběhu migrace provedeny, kdo je provedl a jaké byly výsledky každého kroku. Když se následně zjistí **problém s migrací, většina klientů vás požádá, abyste prošetřili, co se pokazilo a proč**. Mít k dispozici logy vám může ušetřit dny práce [15].

Logování však není důležité jen pro dohledávání chyb, ale slouží také auditorům během auditu, kdy může pomoci rekonstruovat průběh migrace dat. **Logy proto zahrňte také do dokumentace projektu**, Také ostatní zainteresované strany budou mít přehled o tom, jaké kroky byly během migračního projektu prováděny.

4.7 Získejte novou sadu dat pro každé kolo uživatelských testů [15]

Během projektu datové migrace **získejte novou sadu dat pro každé kolo uživatelských testů**. Nezapomeňte, že chcete proces otestovat kompletně od začátku do konce. To znamená, že musíte otestovat nejen své postupy nasazení a kód, ale také schopnost exportovat data konzistentním a přesným způsobem [15].

Podobně jako bylo popsáno v postupu, který se týkal minimalizace počtu prostředníků, i zde je vhodné si **zdokumentovat, jakým způsobem budete data exportovat**. Vytvořte si dokument, který krok po kroku popíše tento postup. To umožní standardizaci a minimalizaci rizika změn v situacích, kdy data exportuje někdo jiný a exportovaná sada dat je stále ve stejném formátu.

5. Rozpočet

5.1 Je-li migrace součástí většího projektu, řešte ji zcela samostatně, a to od sestavování rozpočtu až po testování [1]

Existence samostatného rozpočtu je důležitá, pokud je migrace součástí většího projektu. Podle [1] snižuje neoddělení rozpočtu migrace od celkového projektu šanci na úspěšné dokončení projektu až 4x. Kromě odděleného rozpočtu od celkového projektu by měla být migrace také odděleně testována. Je-li testování datové migrace začleněno do celkového testování, snižuje se pravděpodobnost úspěšného dokončení projektů včas nebo v rámci rozpočtu.

5.2 Zhodnoťte velikost projektu a zjistěte, jaké zdroje bude vyžadovat a kolik na něj potřebujete v rozpočtu [15] [17] [19]

Aby bylo možné projekt realizovat, stanovte **rozsah aktivit, které je nutné v rámci projektu provést**. Rozsah projektu může být ovlivněn mnoha faktory. **Zvažte počet systémů, které je potřeba migrovat a také na počet systémů, do kterých máte migraci provést**. Dále zvažte rozdílnost datových struktur. V případě, že jsou datové struktury 1:1, zabere jejich transformace méně času, než když jsou struktury rozdílné. To má samozřejmě také vliv na pracnost a časovou náročnost projektu, a tedy i na rozpočet. Dále **berte ohled také na komplexnost projektu**.

V případě, že se rozhodnete pro koupi zahraničního portfolia (popřípadě pokud migrujete zahraniční portfolio jako externí systémový integrátor), může dojít k situaci, kdy se jednotlivé právní normy obou zemí liší. Rozdílné jsou tak procesy, číselníkové hodnoty a další skutečnosti, které mohou mít následně vliv na pracnost. Všechny tyto faktory proto zvažte, plánujte a stanovte na základě nich rozpočet. Zároveň **určete vhodné zdroje pro realizaci projektu, včetně lidských zdrojů** (vývojáři, analytici, testéři atd.) a technologií (hardware a software). Jsou-li známy rozsah a zdroje, stanovte náklady projektu a na základě toho poté stanovte rozpočet projektu.

6. Použitá literatura

- [1] HOWARD, Phillip. Data Migration – 2011. [online]. 2011. [cit. 2023-01-11]. Dostupné z: <https://www.bloorresearch.com/restapi/download/YrXfE8OGLjFOZi1SheSH9fterlzCB7ZqN91HIYk7Op7tgpIC>
- [2] Balakrishnan Tara, Gnanasambandam Chandra, Santos Leandro, Srivathsan Bhargs. Cloud-migration opportunity: Business value grows, but missteps abound. [online]. 2021. [cit. 2023-01-11]. Dostupné z: <https://www.mckinsey.com/industries/technology-media-and-telecommunications/our-insights/cloud-migration-opportunity-business-value-grows-but-missteps-abound>
- [3] MAHANTI, Rupa. Data Governance and Data Management: Contextualizing Data Governance Drivers, Technologies, and Tools. 2021. [cit. 2023-01-20]. Dostupné z: <https://www.proquest.com/legacydocview/EBC/6723145?accountid=17203>.
- [4] MAHANTI, Rupa. Data Quality: Dimensions, Measurement, Strategy, Management, and Governance. 2019. [cit. 2023-01-23]. Dostupné z: <https://www.proquest.com/legacydocview/EBC/6262212?accountid=17203>.
- [5] AZEROUAL, Otmane a Meena JHA. Without Data Quality, There Is No Data Migration. Big Data and Cognitive Computing [online]. 2021, vol. 5, no. 2, s. 24. [cit. 2023-01-24]. Dostupné z: <https://www.proquest.com/scholarly-journals/without-data-quality-there-is-no-migration/docview/2544472698/se-2>
- [6] BERTOLUCCI, Jeff. 10 Big Data Migration Mistakes. Informationweek – Online [online]. 2012. ISSN 19383371. [cit. 2023-01-24]. Dostupné z: <https://www.proquest.com/trade-journals/10-big-data-migration-mistakes/docview/1032866116/se-2>.
- [7] WORSCH, Filip. Fúze společností z hlediska projektu migrace dat [online]. Praha, 2013 [cit. 2023-05-25]. Dostupné z: <https://theses.cz/id/isa7tp/>. Diplomová práce. Vysoká škola ekonomická v Praze.
- [8] NĚMCOVÁ, Alžběta. Vytvoření a aplikace metodiky pro hodnocení kvality provedení datové migrace [online]. Praha, 2015 [cit. 2023-01-25]. Dostupné z: <https://theses.cz/id/xrsh1r/>. Diplomová práce. Vysoká škola ekonomická v Praze.
- [9] CLÉMENT, Delphine & Hassine-Guetari, Soumaya & Laboisie, B. Data Quality as a Key Success Factor for Migration Projects. 2012. [cit. 2023-01-25]. Dostupné z: https://www.researchgate.net/publication/280751278_Data_Quality_as_a_Key_Success_Factor_for_Migration_Projects
- [10] MORRIS, J. Practical Data Migration, 2020. [cit. 2023-01-25]. Dostupné z: <https://www.proquest.com/legacydocview/EBC/6370586?accountid=17203>.

[11] CHEMUTURI, M. Mastering IT Project Management: Best Practices, Tools and Techniques. [online]. 2013. [cit. 2023-01-25]. Dostupné z: <https://www.proquest.com/legacydocview/EBC/3319551?accountid=17203>

49

[12] IBM. Best practices for data migration. [online]. 2009. [cit. 2023-01-25]. Dostupné z: https://www.academia.edu/9778539/IBM_Global_Technology_Services_Best_practices_for_data_migration_Methodologies_for_assessing_planning_moving_and_validating_data_migration_Best_practices_for_data_migration_Contents

[13] RUSSOM, P. Best Practices in Data Migration. 2006. [cit. 2023-01-25]. Dostupné z: http://download.101com.com/pub/TDWI/Files/TDWI_Monograph_BPInDataMigration_April2006.Pdf

[14] TSAKUNOVA, Margarita. Migrace datového skladu z MS SQL Server on-premise prostředí do cloudu [online]. Praha, 2022 [cit. 2023-01-25]. Dostupné z: <https://theses.cz/id/ayyil/>. Diplomová práce. Vysoká škola ekonomická v Praze.

[15] MASRI, D. Developing Data Migrations and Integrations with Salesforce: Patterns and Best Practices, 2018. [cit. 2023-01-26]. Dostupné z: <https://www.proquest.com/legacydocview/EBC/5622564?accountid=17203>.

[16] NOREEN, Neysa, R.H.I.A. Data Migration Lessons Learned. Journal of AHIMA [online]. 2013, vol. 84, no. 9, s. 40-42. ISSN 10605487. [cit. 2023-01-26]. Dostupné z: <https://www.proquest.com/trade-journals/data-migration-lessons-learned/docview/1437254005/se-2>.

[17] IntelligentHQ. Why is data migration important? [online]. London: Newstex, 2020. Copyright – Copyright Newstex Nov 4, 2020; Poslední aktualizace - 2022-10-19. [cit. 2023-01-26]. Dostupné z: <https://www.proquest.com/blogs-podcasts-websites/why-is-data-migration-important/docview/2457158229/se-2>.

[18] SARMAH Simanta Shekhar. Data migration. [online]. [cit. 2023-01-26]. Dostupné z: <http://article.sapub.org/10.5923.j.scit.20180801.01.html>

[19] Abou_el_ela Abdou Hussein. Data Migration Need, Strategy, Challenges, Methodology, Categories, Risks, Uses with Cloud Computing, and Improvements in Its Using with Cloud Using Suggested Proposed Model (DMig 1). [online]. [cit. 2023-01-26]. Dostupné z: <https://www.scirp.org/journal/paperinformation.aspx?paperid=106592>

[20] ZKOUMALOVÁ, Barbora. Migrace systémové databáze elektronického obchodu [online]. Brno, 2016 [cit. 2023-01-27]. Dostupné z: <https://theses.cz/id/fkgh2e/>. Diplomová práce. Vysoké učení technické v Brně.

[21] Sebastian-Coleman, L. Measuring Data Quality for Ongoing Improvement: A Data Quality Assessment Framework. [online]. 2013. [cit. 2023-01-28]. Available from: <https://www.proquest.com/legacydocview/EBC/1106491?accountid=17203>.

[22] Microsoft Azure. What is Data Migration?. [online]. [cit. 2023-01-28]. Dostupné z: <https://azure.microsoft.com/en-us/resources/cloud-computing-dictionary/what-is-data-migration/#defining-data-migration>

[23] IBM. What is data governance?. [online]. [cit. 2023-01-28]. Dostupné z: <https://www.ibm.com/topics/data-governance>

50

[24] MAHANTI, R. Data Governance Success: Growing and Sustaining Data Governance. 2021. [online]. [cit. 2023-01-30]. Dostupné z: <https://www.proquest.com/legacydocview/EBC/6826318?accountid=17203>.

[25] WRIGHT, G. Best Practise. [2022]. [cit. 2023-03-20]. Dostupné z: <https://www.techtarget.com/searchsoftwarequality/definition/best-practice>