

Kvalitní data jsou základem pro všechny podnikové procesy. Naše řešení umožňuje monitorování kvality dat, řízení jejich čistění a deduplikace, nebo podporuje jejich validní zadávání do Vašich systémů.

Řešení datové kvality (DQ) společnosti AMBICA s.r.o. má za cíl **maximalizaci kvality dat**, kterými společnost disponuje. K tomuto cíli je využívána kombinace reaktivních a proaktivních opatření, která dokáží datovou kvalitu zvýšit. Dále jsou aplikována opatření umožňující sledovat stav a vývoj kvality shromažďovaných dat. Díky všem těmto opatřením je možné **kontinuálně zlepšovat kvalitu dat**, společně se zvyšováním jejich budoucí použitelnosti. Dále lze také zamezit nečekanému zhoršení datové kvality či se vrátit do stavu před touto událostí.



Schéma:

Řešení datové kvality se skládá z **více komponent**, které mají společný cíl zvýšení kvality dat, se kterými podnik disponuje. Každá komponenta má v řešení své místo, a pokud již není společností ošetřena jinak, pak má potenciál pro další zvýšení datové kvality. Vše však záleží na stávajícím řešení, jeho sofistikovanosti a úrovni práce s daty ve společnosti.

Jednotlivé komponenty:

Každá komponenta řešení datové kvality má svůj specifický účel, kterým pomáhá k jejímu zvýšení. Jednotlivé prvky se **mohou aplikovat i samostatně**, největší efekt ale přinese komplexní řešení.

Měření kvality dat

Mechanismus ohodnocení kvality předpokládá **měřitelnost** jednotlivých atributů. Mají-li být atributy měřitelné, musí existovat jasná pravidla, kdy je atribut v pořádku. Ať už se v praxi jedná o délku záznamu, přítomnost či absenci určitého znaku, formátování textu včetně přesné velikosti písmen či existenci hodnoty v číselníku přípustných hodnot. V praxi tedy může být metrikou kvality například u e-mailového kontaktu shoda záznamu se vzorem předepisujícím povolené znaky a jejich množství. Další možností je kontrola toho, jestli existuje zadaná hodnota v číselníku, například jestli je zadávána adresa v registru a tím pádem se jedná o existující adresu.

Omezení vstupu nevalidních dat

Primárním úkolem je **zabránění nebo minimalizování příslušnosti nových nevalidních dat do systému**. Řešením je aplikování validačních pravidel do formulářů pro vstup dat. Naše **validační komponenta** je schopná přidat validační pravidla do již hotových, zejména webových formulářů. Uživatel formuláře je tak nejen okamžitě informován o správnosti zapsané hodnoty, ale navíc dostává komfort našeptávačů, které zrychlí zadávání a učiní jej pohodlnějším.

Při zachycení chyby hned na počátku procesu sbírání dat je lepší možnost opravit chybu v souladu se skutečností. Každá pozdější případná oprava obsahuje určitou míru nejistoty o správnosti navrženého opatření. Aplikace validačních pravidel do desktopového formuláře umožňuje v praxi řešení nastávajících situací: Například pokud se na prodejně zaměstnanec pokusí zanést do systému hodnotu adresy, která neexistuje. Zaměstnanec bude na místě upozorněn, že se nejedná o vyhovující formát adresy. Tím pádem nejenže bude mít možnost chybu opravit na místě v přítomnosti zákazníka, ale ještě mu může být našeptána správná adresa, kterou si systém dohledá podle registru **RÚIAN** (Registr územní identifikace, adres a nemovitostí). Validací komponenta umístěná do formuláře na webu bude mít stejnou schopnost, ale může fungovat bez přítomnosti personálu. Stejný příklad může být v případě validační komponenty **ověřování platnosti adresy** v e-shopu, kdy může být přímo zákazník upozorněn, že se nejedná o platnou adresu. E-shopu je pak předána adresa s jasným určením jednotlivých částí adresy. Je zřejmé, zda text představuje ulici, místní část nebo obec a zda číslo je číslem popisným, orientačním nebo evidenčním.

Validační komponenta

Validační komponenta představuje soubor entit, v kterých lze určit a ověřovat formální správnost zadaného obsahu. Jednotlivé entity lze do vlastního formuláře použít a rozšířit podle reálných potřeb. Následující formulář je příkladem obvyklého kontaktního formuláře, který je daty využíván v elektronických systémech.

Kontaktní údaje	
Jméno	<input type="text"/>
Příjmení	<input type="text"/>
E-mail	<input type="text"/>
Mobil	<input type="text"/>
Adresa	<input type="text"/>

→ Doručovací údaje

Uložit záznam

→ Další údaje, které ještě umíme validovat

→ Způsob indikace správnosti uvedené hodnoty

Pravidla validace

Validace je prováděna porovnáváním proti vzorům nebo na shodu s existující číselníkovou hodnotou.

Následně lze ověřit, zda je číslo kompletní zadání informací o adresním bodu včetně GPS souřadnic, jejich možná použití doložit přílohou s viděnou mapou společnosti Geonim.cz. GPS souřadnice lze doplnit i pro starší adresy pomocí další komponenty.

Prezentace adresního bodu

Pro zadání validní adresy lze získat kompletní zadání informací o adresním bodu včetně GPS souřadnic, jejich možná použití doložit přílohou s viděnou mapou společnosti Geonim.cz. GPS souřadnice lze doplnit i pro starší adresy pomocí další komponenty.

Ověřování a zjištění entit

Stejný typ entity lze ve formuláři využít i opakovaně, jak to dokládá příklad adresy v kontaktních a doručovacích údajích. Ově entity pracují na sobě zcela nezávisle.

Uložení údajů: Stejný typ entity lze ve formuláři využít i opakovaně, jak to dokládá příklad adresy v kontaktních a doručovacích údajích. Ově entity pracují na sobě zcela nezávisle.

Ukládání údajů

Uložení údajů: Stejný typ entity lze ve formuláři využít i opakovaně, jak to dokládá příklad adresy v kontaktních a doručovacích údajích. Ově entity pracují na sobě zcela nezávisle.

Validační kontroly

Validační kontroly: Stejný typ entity lze ve formuláři využít i opakovaně, jak to dokládá příklad adresy v kontaktních a doručovacích údajích. Ově entity pracují na sobě zcela nezávisle.

Přínosy

- Snížení nákladů**
Snižování nákladů se děje ve více rovinách, ale hlavně dochází ke snížení nákladů na kontaktování jednotlivých osob, kdy je eliminováno množství nedoručitelné pošty, či pošty vícekrát odeslané na jednu adresu.
- Zvýšení efektivity**
Díky zvýšení kvality dat budou mít data větší vypovídací hodnotu, protože bude možné si například v databázi nákupů a zákazníků spojit duplikovaná data, udělat si obrázek o nákupním chování jednotlivých zákazníků, jejich možných budoucích nákupech, frekvenci s jakou nakupují. Zákazníci s vyšším nákupním potenciálem pak budou díky opraveným kontaktním informacím lépe dosažitelní a bude se tím i zvyšovat efektivita marketingové kampaně zacílená na zákazníky.

Výsledek

Výsledkem aplikace datové kvality je konsolidované úložiště s validními a strukturovanými daty bez opakujících se záznamů o stejných entitách. Taková data jsou snadno použitelná pro další zpracování. To samo o sobě snižuje náklady v podobě nižší pracovní a časové náročnosti spojené s manipulací s daty. Hlavním přínosem je ovšem vyšší spolehlivost dat, která se projeví zpětnou vazbou při práci s daty.

Typické scénáře použití

Scénář 1 – změření kvality zákaznických informací

Formou služby provedeme ve vašich systémech změření kvality kontaktních informací vašich zákazníků (vzorku dat nebo celé databáze) a doporučíme možná opatření pro zlepšení kvality dat.

Scénář 2 – jednorázové vyčištění či zkvalitnění zákaznických informací

Formou služby provedeme jednorázové (i opakovaně) vyčištění, případně deduplikaci vašich kontaktních zákaznických informací (vzorku dat nebo celé databáze) včetně monitorování kvality v průběhu čištění (lze dokladovat pozitivní posun kvality).

Scénář 3 – zamezení vstupu nevalidních dat

Do vašeho e-shop systému můžeme (ve spolupráci s jeho dodavatelem) integrovat univerzální validační komponentu, která zajistí ověření kontaktních zákaznických informací už při jejich vstupu do systému. Validací komponenta zároveň poskytne vašim zákazníkům vyšší uživatelský komfort při zadávání informací.

Scénář 4 – implementace systému pro trvalé zajištění kvality zákaznických informací

Můžeme (typicky i postupně) implementovat jednotlivé komponenty systému datové kvality, které vám umožní mít kvalitu dat pod neustálou kontrolou. Od průběžného měření, monitoringu a reportingu datové kvality (dokumentace trendů, identifikace problémů a příležitostí), přes čištění a deduplikaci historických dat nebo automatizované čištění aktuálních denních přírůstků, až po podporu manuálního čištění a proaktivních opatření při vstupu dat (univerzální validační komponenta).

Prvotní čištění dat

Na začátku projektu datové kvality většinou již zákazník má svá data, která je potřeba opravit podle stejných validačních pravidel, jako tomu je u dat nově vstupujících do systému. Prvotní čištění dat tedy slouží k **vyčištění dříve nashromážděných dat** a k jejich úpravě do požadovaného formátu pomocí čistících mechanismů. Počet záznamů, které čištění neobsáhne, je úměrný míře progresu zvolené strategie automatizovaného čištění. Při volbě progresivnější strategie bude počet konečných nevalidních dat nižší. Se stoupající progresí strategie čištění však roste riziko, že oprava přinese nežádoucí významovou změnu hodnoty, proto tato míra závisí pouze na zadavateli a před provedením automatizovaného čištění je tato fáze otestována na vzorku dat pro správnou funkčnost.

Centrální uložení a struktura, historizace událostí

Data integrovaná z různých zdrojů jsou uložena na centrálním uložení. Důležitou vlastností dobrého datového uložení je zavedení **historizace událostí**, což umožňuje **nahlížení do kompletní historie změn**. Samozřejmostí je možnost kdykoliv se vrátit do jakéhokoliv předchozího stavu. Tyto vlastnosti slouží k maximalizaci bezpečí při práci s daty, která bývají často klíčovou devizou společnosti. V praxi pak lze dohledat změny například v databázi kontaktních informací zákazníků, kde kromě samotného dohledání změn bude v případě špatného zadání nebo jakékoli chyby možné vrátit databázi do stavu před touto událostí. Velkou devizou našeho řešení datové kvality je fakt, že právě k centrálnímu uložení jsou připojeny jednotlivé systémy a **všichni uživatelé tak pracují s aktuálními daty.**

Reporting datové kvality

Reporting pomáhá sledovat stav datové kvality v čase a **dokumentovat její historický vývoj**. Díky nastaveným metrikám lze sledovat, zda se datová kvalita vyvíjí správným či špatným směrem a popřípadě lze zjistit, proč se tomu tak děje. Pokud například začne náhle klesat kvalita dat po zavedení nového vstupu dat do systému, pak lze předpokládat, že nám do systému vnikají nevalidní data a je třeba tento problém nějak vyřešit. Další výhodou reportingu datové kvality je možnost **porovnávání datové kvality od jednotlivých poboček či prodejců**. Díky těmto funkcionalitám bude možné například ty proškolit ty pobočky, kde nejsou data vkládána v požadovaném formátu, a zvýšit tak kvalitu dat této pobočky a potažmo i celé databáze. Další funkcionalitou je operativní reporting, který může vyselektovat data pro manuální čištění.



Čištění přírůstků a manuální dočištění dat

Čištění přírůstků má za úkol minimalizovat možnou chybu lidského faktoru tím, že **dočišťuje data nově vstupující do systému**. Dále lze také manuálně dočistit data, která nebyla vyčištěna automaticky. Zde již dočištění závisí na obsluze, která musí vyhodnotit správnou variantu hodnoty, či případně doplnit chybějící hodnotu pomocí doplňovacího dotazu. Po manuálním dočištění by ideálně mělo dojít k úplné opravě záznamů či jejich vyřazení z evidence, ale reálně však bývají ponechány i částečně opravené záznamy, protože i ty se mají využít a mají svou hodnotu. V praxi pokud byl opraven telefon a e-mail, ale nebyla dohledána adresa, pak je záznam ponechán v evidenci, protože lze kontaktovat danou osobu prostřednictvím opravených kontaktů (e-mail, telefon). **Samotné čištění přírůstku vypadá většinou tak, že jsou denně v noci čištěna data, která do systému během dne přibyla.**

Deduplikace dat

Často se stává, že se díky více vstupním bodům dat do systému vyskytují záznamy o stejném subjektu v databázi vícekrát. Tato duplikace je však nežádoucím efektem, protože zvyšuje náklady (zvyšuje se databáze, zpomaluje se práce s ní, dochází k odesílání více letáků na jednu adresu atd...). Proto je dalším krokem deduplikace dat, která spočívá ve **sloučení a nahrazení těchto násobných záznamů pouze jedním záznamem**. Deduplikace kromě snížení nákladů zvýší kvalitu dat, kdy například v analýze zákaznického chování jsme schopni **spojit celou zákaznickou historii**, místo duplikovaných záznamů, které vystupují jako jednotliví zákazníci a mají pro další analýzy sníženou vypovídací hodnotu.

Reference

Mountfield, a.s. vlastní řešení datové kvality od společnosti AMBICA již od roku 2010. Samotné řešení obnášelo kompletní zastřešení datové kvality ve společnosti mající miliony zákazníků. Díky řešení od společnosti AMBICA bylo možné dále využít data, zasáhnout daleko více zákazníků marketingovými kampaněmi a snížit náklady na nedoručitelnou korespondenci.

Mountfield

Společnost Mountfield, a.s. také ušetřila na výdajích za duplicitní rozesílání letáků a dále na takto kvalitní datové základně postavila systém reportingu, který slouží k další spoustě rozhodovacích procesů.

Jak na to

Pokud Vás informace o řešení datové kvality zaujaly, neváhejte si tuto platformu prakticky vyzkoušet. Přístup k validačním komponentě, dalším informacím a praktickým ukázkám získáte na webových stránkách **www.ambica.cz** v sekci Produkty a řešení/BI a DW/Datová kvalita. Pokud by Vám však tyto praktické ukázky nestačily, můžete nás v případě zájmu o individuální prezentaci nebo o konzultaci k možnostem využití datové kvality ve Vaší konkrétní společnosti kontaktovat na **ambica@ambica.cz**.



AMBICA s.r.o.

AMBICA s.r.o. je česká softwarová a konzultační společnost s více než desetiletou dobou působení na trhu. Naše společnost dodává inteligentní softwarová řešení a služby se specializací na **Business Intelligence a Data Warehousing**.

Portfolio řešení společnosti AMBICA pokrývá významný rozsah potřeb v oblasti podpory obchodních procesů, a to od problematiky datové integrace a datové kvality, přes sofistikovanou analytickou a reportingová řešení až po systémy pro podporu plánování a řízení výkonu organizací a další specializované aplikace. Při vývoji produktů a realizaci zákaznických řešení se AMBICA opírá jednak o platformu Microsoft a jednak o vybrané open source technologie.

Díky kombinaci komerčních a open source prvků v architektuře našich řešení si u nás vybere opravdu každý řešení přesně na míru potřebám společnosti. Neváhejte a kontaktujte nás na e-mailové adrese **ambica@ambica.cz** či na telefonním čísle **+420 384 390 700** pro další informace.