

Univerzita Pardubice
Dopravní fakulta Jana Pernera

Přístup k BIG DATA v oddělení logistiky ve Škoda Auto a.s.

Lukáš Motl

Bakalářská práce
2025

Univerzita Pardubice
Dopravní fakulta Jana Pernera
Akademický rok: 2024/2025

ZADÁNÍ BAKALÁŘSKÉ PRÁCE

(projektu, uměleckého díla, uměleckého výkonu)

Jméno a příjmení: **Lukáš Motl**
Osobní číslo: **D21212**
Studijní program: **B1041A040002 Technologie a management v dopravě**
Specializace: **Dopravní management a marketing**
Téma práce: **Přístup k BIG DATA v oddělení logistiky ve Škoda Auto a.s.**
Zadávací katedra: **Katedra dopravního managementu, marketingu a logistiky**

Zásady pro vypracování

Bakalářská práce bude obsahovat:

- vymezení základních pojmů a literární rešerši v oblasti přístupu k Big Data,
- analýzu současného stavu přístupu k Big Data v oddělení logistiky ve společnosti Škoda Auto a.s.,
- návrhy na zlepšení v přístupu k Big Data v oddělení logistiky ve společnosti Škoda Auto a.s.

Rozsah pracovní zprávy: **35-45 stran**
Rozsah grafických prací: **dle doporučení vedoucí/ho**
Forma zpracování bakalářské práce: **tištěná/elektronická**

Seznam doporučené literatury:
dle pokynů vedoucí/ho práce

Vedoucí bakalářské práce: **Ing. Dalibor Gottwald, Ph.D.**
Katedra dopravního managementu, marketingu
a logistiky

Datum zadání bakalářské práce: **31. října 2024**
Termín odevzdání bakalářské práce: **12. května 2025**

L.S.

doc. Ing. Ladislav Řoutil, Ph.D.
děkan

Ing. Pavla Lejsková, Ph.D.
vedoucí katedry

V Pardubicích dne 24. dubna 2025

Práci s názvem Přístup k BIG DATA v oddělení logistiky ve Škoda Auto a.s. jsem vypracoval samostatně. Veškeré literární prameny a informace, které jsem v práci využil, jsou uvedeny v seznamu použité literatury.

Byl jsem seznámen s tím, že se na moji práci vztahují práva a povinnosti vyplývající ze zákona č. 121/2000 Sb., o právu autorském, o právech souvisejících s právem autorským a o změně některých zákonů (autorský zákon), ve znění pozdějších předpisů, zejména se skutečností, že Univerzita Pardubice má právo na uzavření licenční smlouvy o užití této práce jako školního díla podle § 60 odst. 1 autorského zákona, a s tím, že pokud dojde k užití této práce mnou nebo bude poskytnuta licence o užití jinému subjektu, je Univerzita Pardubice oprávněna ode mne požadovat přiměřený příspěvek na úhradu nákladů, které na vytvoření díla vynaložila, a to podle okolností až do jejich skutečné výše.

Beru na vědomí, že v souladu s § 47b zákona č. 111/1998 Sb., o vysokých školách a o změně a doplnění dalších zákonů (zákon o vysokých školách), ve znění pozdějších předpisů, a směrnicí Univerzity Pardubice č. 7/2019 Pravidla pro odevzdávání, zveřejňování a formální úpravu závěrečných prací, ve znění pozdějších dodatků, bude práce zveřejněna prostřednictvím Digitální knihovny Univerzity Pardubice.

V Pardubicích dne 5. 5. 2025

Lukáš Motl

Rád bych poděkoval vedoucímu práce Ing. Daliboru Gottwaldovi, PhD., za jeho vstřícný přístup a cenné rady, které byly klíčové při zpracování bakalářské práce. Dále děkuji společnosti Škoda Auto a.s. za poskytnutí dat potřebných k vypracování práce. V poslední řadě děkuji své rodině a přátelům za jejich podporu, díky níž se mi podařilo dosáhnout tohoto významného úspěchu.

ANOTACE

Práce se zaměřuje na přístup k Big Data v oddělení logistiky Škoda Auto a.s. V první kapitole jsou teoreticky vymezena Big Data a způsoby jejich zpracování. Druhá kapitola se zabývá analýzou současného stavu přístupu k Big Data v oddělení logistiky Škoda Auto a.s. Na základě této analýzy budou ve třetí kapitole navržena opatření zlepšující přístup ke zpracování Big Data v oddělení logistiky Škoda Auto a.s.

KLÍČOVÁ SLOVA

Big Data, interní logistika, Škoda Auto a.s., datová architektura, vizualizace dat, Power BI

TITLE

BIG DATA strategy in the logistics department of Škoda Auto a.s.

ANNOTATION

The work focuses on the Big Data strategy in the logistics department of Škoda Auto a.s. The first chapter provides a theoretical framework of Big Data and its processing methods. The second chapter analyses the current approach to Big Data within the logistics department of Škoda Auto a.s. Based on this analysis, the third chapter proposes measures to improve Big Data processing within the logistics department of Škoda Auto a.s.

KEYWORDS

Big Data, internal logistics, Škoda Auto a.s., data architecture, data visualization, Power BI

OBSAH

ÚVOD	9
1 TEORETICKÉ VYMEZENÍ PROBLEMATIKY BIG DATA	11
1.1 Charakteristika Big Data	11
1.2 Typologie struktury Big Data.....	14
1.3 Přístupy ke zpracování Big Data.....	17
1.4 Cloud Computing.....	19
1.5 Zabezpečení Big Data	21
1.6 Iot v kontextu Big Data	22
2 ANALÝZA PŘÍSTUPU K BIG DATA V ODDĚLENÍ LOGISTIKY ŠKODA AUTO A.S....	23
2.1 Škoda Auto a.s.	23
2.2 Současný stav přístupu k Big Data v oddělení interní logistiky svařovny závodu v Kvasinách	24
2.2.1 Distribuce materiálu ze skladu svařovny	25
2.2.2 Dílčí rozvoz materiálu ze skladu M7	26
2.2.3 Mezioperační převozy dílů.....	27
2.2.4 Automatické objednávání a kontrola návozu sudů ze skladu barev.....	27
2.2.5 Objednávání Gitterboxů.....	28
2.2.6 Proces hlášení a zpracování poruch ve svařovně	28
2.3 Rozhovor s experty	29
2.4 Zhodnocení aktuálního stavu	34
3 NÁVRHY NA ZLEPŠENÍ PŘÍSTUPU K BIG DATA V ODDĚLENÍ LOGISTIKY ŠKODA AUTO A.S.	36
3.1 Kombinace Data Lake a Data Warehouse.....	36
3.2 Data Lakehouse.....	38
3.3 Power BI reporting.....	40
3.4 Shrnutí.....	42
ZÁVĚR	44
POUŽITÁ LITERATURA.....	45
SEZNAM TABULEK.....	49
SEZNAM OBRÁZKŮ.....	50

SEZNAM ZKRATEK.....	51
---------------------	----

ÚVOD

Téma bakalářské práce se zabývá problematikou přístupu k Big Data v oddělení logistiky Škoda Auto. Toto téma práce je zvoleno z důvodu neustále rostoucího trendu rozsáhlé datové analýzy. Ta je klíčovým prvkem pro budoucí technologický rozvoj podniku a udržení tržní konkurenceschopnosti.

Podniky, které chtějí v současném tržním prostředí uspět, staví své úspěchy primárně na kvalitě svých výrobků a služeb. Za dosažením těchto atributů stojí mnoho aspektů, z nichž jeden je efektivita výrobního procesu. Tu v dnešním tržním prostředí podniky zvyšují rozsáhlou digitalizací většiny výrobních procesů. Digitalizace sebou však nese rozsáhlé datové objemy, v dnešní době nazývaná jako Big Data, které vyžadují rozsáhlé a komplikované architektury určené pro jejich další zpracování. Pro správné fungování oddělení interní logistiky svařovny je tedy zásadní, aby se s daty, jejichž objemy neustále narůstají, nakládalo vhodným způsobem, který dále umožní vyšší úroveň optimalizace procesů, přesnější plánování a řízení v rámci podniku. Tento přístup dále otevírá nové možnosti ve formě přesných predikčních nástrojů a umělé inteligence.

Bakalářská práce je členěna do tří kapitol. První kapitola se věnuje teoretickému vymezení problematiky Big Data. V úvodu této kapitoly Big Data charakterizována a rozdělena dle typu datové struktury. Obsahem úvodní kapitoly je dále způsob přístupu k Big Data a jejich zabezpečení. V závěru teoretické části práce jsou zmíněna cloudová úložiště a Internet of Things.

Druhá kapitola v úvodní části stručně popisuje Škoda Auto a.s., její historii, vývoj, současnou nabídku a poslední inovace v rámci podniku. Dále se věnuje analýze současného přístupu k Big Data. V této části jsou popsány procesy, při kterých vzniká většina dat v oddělení interní logistiky svařovny a způsob, jakým se s nimi nakládá. Součástí této kapitoly je dále rozhovor s experty, který cílí na vyzdvižení slabých míst v kontextu nakládání s daty

Třetí kapitola poskytuje, na základě zjištěných nedostatků, návrhy datových architektur, které by mohl podnik v budoucnu využít. Jsou zde navrženy dvě možné alternativy, které by zásadním způsobem změnily přístup k datům v oddělení interní logistiky Škoda Auto a.s. v Kvasínách.

Cílem bakalářské práce je na základě teoretického vymezení problematiky Big Data provést analýzu současného stavu přístupu k Big Data v oddělení interní logistiky svařovny Škoda Auto a.s. v závodu v Kvasínách. Na základě výstupů provedené analýzy budou navržena

opatření pro zlepšení přístupu k Big Data v oddělení interní logistiky svařovny ve Škoda Auto a.s.

1 TEORETICKÉ VYMEZENÍ PROBLEMATIKY BIG DATA

Teoretická část práce se zaměřuje na vymezení pojmu Big Data z různých úhlů pohledu, jeho charakteristiky a strukturu. Dále se zabývá způsoby zpracování a analýzy dat. V poslední řadě se soustředí i na tematiku Cloud computingu a Internet of things, se kterými jsou Big Data úzce svázaná.

1.1 Charakteristika Big Data

„Big Data odkazuje na schopnost realizovat ve velkém měřítku činnosti, které by v menším rozsahu nebyly možné. Umožňuje získávat nové poznatky nebo vytvářet hodnotu způsobem, který zásadně ovlivňuje trhy, organizace, vztahy mezi občany a vládami i další oblasti.“

(Mayer-Schönberger a Cukier, 2014, s. 6)

„Big Data mají obrovský objem a v porovnání s tradičními daty je jejich analýza a správa komplikovanější. Jejich ukládání vyžaduje škálovatelnou architekturu, efektivní způsoby ukládání a manipulace.“

(Sandhu, 2022, s. 33)

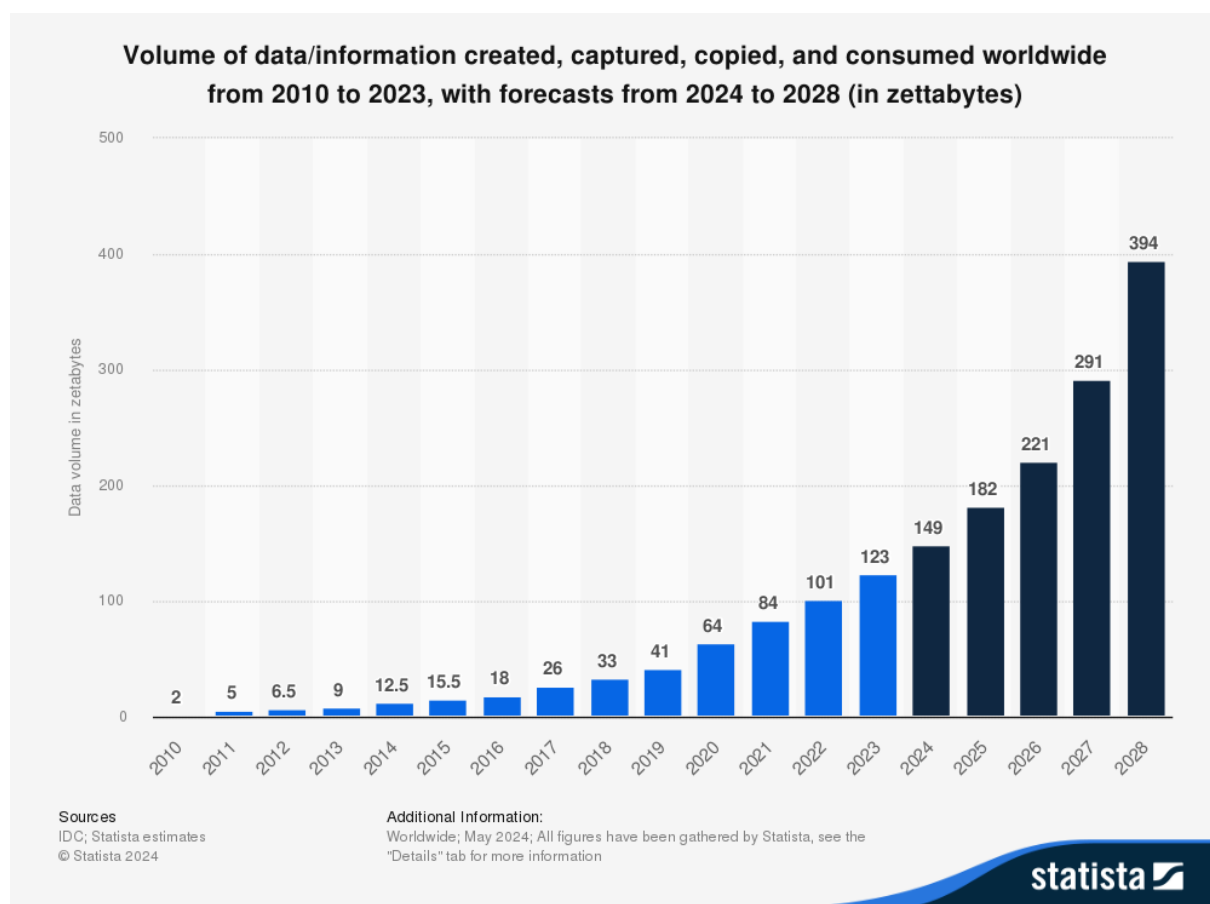
„Velká data představují informační zdroje s vysokým objemem, rychlostí a/nebo rozmanitostí, které vyžadují nákladově efektivní a inovativní způsoby zpracování. Tyto postupy přinášejí lepší nadhled, schopnost rozhodování a automatizaci procesů.“

(Gartner, 2024)

„Big Data jsou oceánem informací, ve kterém „plaveme“ každý den – obrovské objemy dat v řádech Zettabajtů proudících z našich počítačů, mobilních zařízení a strojních senzorů. Tato data organizace používají k podpoře rozhodování, zlepšování procesů a zásad, dále k vytváření produktů, služeb a zážitků zaměřených na zákazníka. Big Data jsou definována jako „velká“ nejen kvůli svému objemu, ale také kvůli rozmanitosti a složitosti své povahy. Typicky přesahují kapacity tradičních databázových systémů, pokud jde o jejich zachycení, správu a zpracování. Big Data mohou pocházet odkudkoli a mohou být cokoliv na zemi, co jsme schopni digitálně monitorovat. Meteorologické družice, Internet of Things (IoT), dopravní kamery, trendy sociálních sítí – to jsou jen některé z datových zdrojů, které vytěžujeme a analyzujeme proto, aby se podniky staly odolnějšími a konkurenceschopnějšími.“

(SAP, 2025)

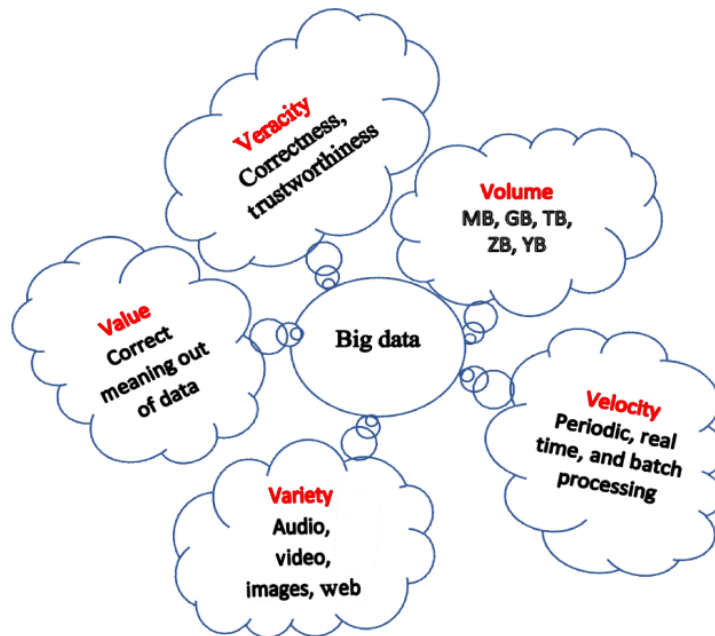
Koncept Big Data, nebo česky Velká data, má mnoho od sebe různých definic. Navzdory jeho všudypřítomnosti, Arbesman (2013) uvádí, že žádná stručná definice neexistuje. Obecně lze říci, že se jedná o nekonkrétně velký objem dat, u kterého vlivem jeho současné velikosti a trvalého nárůstu vznikají komplikace, týkající se jejich skladování a analýzy. Dle webu Statista.com (2024) objem dat, která byla vytvořena, zachycena, kopírována a konzumována mezi roky 2010 a 2023 vzrostl ze 2 Zettabajtů (ZB)¹, na 123 ZB. V letošním roce (2024) se odhaduje, že by objem dat měl dosáhnout 149 ZB. Tímto tempem tedy lze očekávat v roce 2028 2,64 násobek (394 ZB) současného objemu dat.



Obrázek 1 Objem dat/informací vytvořených, zachycených, zkopírovaných a spotřebovaných celosvětově mezi roky 2010 a 2023, včetně prognózy pro období mezi roky 2024 a 2028 (v zettabajtech), (Statista, 2024)

¹10²¹ bajtů

Doug Laney (2001) charakterizoval Big Data pomocí tzv. 3V, která měla specifikovat výzvy, kterým v této problematice čelíme. Jednalo se o Volume, Velocity a Variety. S příchodem nových datových formátů a způsobů zpracování dat, bylo třeba charakteristiku rozšířit. Sandhu (2022) uvádí, že k původním „V“ bylo dále přidáno Value a Veracity.



Obrázek 2 Pět V v Big Data (Sandhu, 2022)

V současné době můžeme Big Data charakterizovat až pomocí 14V (Artera, 2023):

- **Volume** (Objem) – množství zachycených a skladovaných dat, která jsou generována z různých zdrojů (sociální sítě, Internet of Things (IoT)² zařízení, průmyslová zařízení, apod.), které neustále roste,
- **Velocity** (Rychlost) – rychlost, kterou nová data vznikají, což úzce souvisí s rostoucí rychlostí zpracování, ukládání a analyzování dat,
- **Variety** (Různorodost) – data mohou mít různé zdroje (interní/externí), formáty (textové dokumenty, číselná data, informace ze senzorů) a v poslední řadě struktury (nestrukturovaná, semistrukturovaná a strukturovaná),
- **Value** (Hodnota) – jak cenné a užitečné informace je možné z dat získat a zda je možné se na základě těchto dat rozhodovat

² Síť propojených zařízení, vybavena senzory, které jim umožňují přijímat a přenášet data z a do systémů

- **Veracity** (Pravdivost) – kvalita a přesnost dat, což je jeden z méně kontrolovatelných parametrů. Pokud jsou dostupná data z různých zdrojů a formátu, analýza je v tuto chvíli nespolehlivá. Zajištění pravdivosti dat je tedy nutností (Sandhu, 2022),
- **Variability** (Variabilita) – vzhledem k tomu, že získaná data mohou pocházet z různých zdrojů (Různorodost), je důležité si oddělovat data podle tohoto, zda je možné z nich získat informace, nebo zda se jedná o data sloužící k predikci, která budou pro naše účely reprezentativní,
- **Venue** (Prostředí) – prostředí, systémy nebo platformy, kde jsou data zpracovávána a skladována,
- **Vocabulary** (Slovník) – potřeba sdílet terminologii a sémantiku, pro popis a definování datových modelů a jejich struktur,
- **Vagueness** (Nejasnost) – data mohou být částečná, nekompletní, nebo nejasná, v tu chvíli jsou nehodná pro tradiční analýzu. Nejasnost může být způsobena nepřesností zdrojů, samotnou různorodostí dat, nebo komplexitou procesu získávání dat,
- **Volatility** (Nestálost) – hodnota dat se může rychle měnit. Důvodem může být neustálá tvorba nových dat, například data z IoT senzorů, která se mění v reálném čase,
- **Visualization** (Vizualizace) – reprezentace dat tak, aby byla lépe pochopitelná a přehledná. Toto napomáhá k rozpoznávání vzorů, trendů a vzájemných vztahů, které by jinak byly hůře dohledatelné,
- **Virality** (Viralita) – rychlost, kterou se data šíří, přenášejí a využívají,
- **Viscosity** (Viskozita) – rozpor mezi reálnou událostí a jejím popisem. Například zpoždění, které může vést ke ztrátě pro uživatele klíčové informace, což vede k snížení spolehlivosti dat

1.2 Typologie struktury Big Data

Big Data strukturujeme podle jednoho z 5 (základních) „V“, Variety, tedy různorodost. Martinez-Mosquera, Navarrete a Lujan-Mora (2020) dělí Big Data dle struktury na:

- **Strukturovaná data** – jsou zpravidla reprezentována v tabulkové formě, nebo v relačních databázích. Jedná se nejorganizovanější formu dat, která je jednoduše přístupná, skladovatelná a analyzovatelná. Každý prvek strukturovaných dat je adresovatelný, je tedy možné ho přesně definovat a je možné ho dělit do různých skupin. Fakt, že strukturovaná data jsou uchovávána v relačních databázích, umožňuje

analyzovat a komplexně dotazovat pomocí Structured Query Language (SQL)³. Vzhledem k tomu, že data mají jasně definovanou strukturu, v praxi se jedná například o datech týkající se finančních transakcí (data z prodejů), demografické informace našich zákazníků v e-commerce⁴ (jméno, adresa, věk, pohlaví), nebo poloha Global Positioning System (GPS), zachycená v konkrétních časových oknech. Daní za jednoduchost analýzy strukturovaných dat může být limitovaná flexibilita, vyplývající z jejich tabulkového charakteru. Což vede také k nutnosti plánování struktur a jejich definování předem. To může vést ke zpomalování projektů (Sedgwick, 2024). Je důležité zmínit, že dle reportu od společnosti Cisco (2014), je odhadem pouhých 10% dat strukturovaných. Zbylých 90 % zastupují data semistrukturovaná a nestrukturovaná.

- **Semistrukturovaná data** – spojují dohromady vlastnosti a prvky dat strukturovaných a nestrukturovaných. Nedrží se formální struktury jako strukturovaná data (relační databáze) a zároveň přinášejí částečnou flexibilitu, jako nestrukturovaná data. Pro účely interní organizace těchto dat používáme určité značky, nebo metadata. Například formáty jako je Comma-separated values (CSV)⁵, eXtensible Markup Language (XML) a JavaScript Object Notation (JSON). Ty mají hierarchickou strukturu, ale nejsou vždy konzistentně uspořádané (Microsoft Azure, 2024). Jsou tedy ideální pro využití ve webových aplikacích a při zpracovávání dat z IoT zařízení. Tato data jsou uchovávána v databázích NoSQL⁶, které umožňují flexibilní zpracování dat a možnost provozu v libovolném měřítku (Microsoft Azure, 2024). S flexibilitou, který tento typ dat přináší, vyvstávají dva potenciální problémy. Zajištění konzistentnosti s rostoucím množstvím dat a jejich komplexitou. Dále problémy s kompatibilitou, při případné integraci jiných systémů, jelikož data tohoto typu nejsou nijak standardizovaná, (Matillion, 2024),
- **Nestrukturovaná data** – se nedrží žádného předem definovaného schématu nebo nemají nijak organizovanou strukturu. Do této kategorie řadíme textové dokumenty, emaily, sensorová data, obrázky a videa, nebo například data ze sociálních sítí. Tento typ dat může být velmi bohatý na informace (např.: zpětná vazba zákazníků), ale na druhou stranu je velmi komplikované jej zpracovat a uchovávat. V současné době můžeme k analýze takových dat využívat Artificial Intelligence (AI)⁷, což může vést

³ Jazyk sloužící k manipulaci s daty a správě databází.

⁴ Souhrnné označení pro obchodní činnosti prováděné na internetu a pomocí dalších elektronických prostředků.

⁵ Hodnoty oddělené čárkou.

⁶ „Není použito SQL“

⁷ Umělá inteligence.

k nárůstu nákladů. Nestrukturovaná data mohou v mnohých případech zachycovat skutečné scénáře z reálného světa, což přináší velkou míru flexibility zpracování. Je třeba však dávat pozor na jejich nekonzistentní formát a kvalitu, které mohou přinášet nepřesné výsledky analýzy (Matillion, 2024),

- **Metadata** – jsou takzvaná „data o datech“. Jedná se o data, která jsou oddělená od obsahu samotných dat (Badman a Kosinski, 2024). Může se jednat o jméno autora, velikost souboru, datum jeho vytvoření, strukturu tabulky, datový typ, nebo vzájemný vztah mezi tabulkami. Rozdělení metadata včetně jejich popisu je uvedeno v tabulce 1.

Tabulka 1 Typy Metadata

Metadata	Popis
• Deskriptivní	• Základní informace o datech, jako je jejich název, autor nebo klíčová slova,
• Strukturální	• Definují vzájemnou organizaci a vztahy mezi jednotlivými datovými prvky,
• Administrativní	• Poskytují informace o přístupech a právech o nakládání s daty. Dále, kdo je jejich vlastníkem,
• Technická	• Obsahují informace jako rozlišení, umístění v úložišti a formát. V případě cloudových řešení zajišťují, že jsou ve všech úložištích uložena správným způsobem,
• Archivační	• Zajišťují možnost uchovávat data po dlouhou dobu. Strategie jejich zálohování, nebo přechod na nové datové formáty.

Zdroj: Badman a Kosinski (2024), upraveno autorem

Metadata zajišťují vysokou kvalitu dat a jejich integritu. V případě databází jsou využívána k organizaci, třídění a vyhledávání dat, což umožňuje jejich další použití v rámci podniku.

1.3 Přístupy ke zpracování Big Data

Mucci a Stryker (2024) uvádí, že se analytika Big Data zabývá systematickým zpracováním a analýzou velkých souborů komplexních dat, ze kterých je možno získávat cenné poznatky. Jedná se o odhalování trendů, vzorců a možných korelací ve velkém množství surových (raw) dat. Na rozdíl od tradiční datové analytiky, Big Data analytika se nezabývá jen daty strukturovanými, ale všemi typy. Komplexita těchto dat vyžaduje pokročilé analytické techniky, jako například machine learning (ML)⁸ a data mining⁹.

Využíváme čtyři základní analytické metody:

- **Deskriptivní** – soustředí se na popis a shrnutí starších dat, aby došlo k pochopení jejich základů a charakteristik. Takzvaná „Co se stalo?“ část analýzy,
- **Diagnostická** – neboli „Proč se to stalo?“ fáze analýzy. Prostoupení hlouběji do dat nám umožňuje identifikovat kořeny různých trendů a vzorců, které byly pozorovány v deskriptivní části,
- **Prediktivní** – využívá statistického modelování, historických dat a ML k předpovídání budoucích trendů, tzn. „Co se stane?“,
- **Preskriptivní** – na základě získaných poznatků poskytuje doporučení sahající za hranice prediktivní metody. Využívá se k poskytování doporučení pro budoucí akce. Jedná se o tzv. „Co dělat?“ fázi.

Samotné analýze datových sad předchází několik fází, pomocí kterých se ze surových dat, stanou data připravená k analyzování. Microsoft (2024) tento proces dělí na:

- **Shromažďování** – tedy získávání dat z různých zdrojů, jako je web, mobilní zařízení, nebo cloudy a jejich ukládání. Jedním z prvních rámců sloužící pro zpracování Big Data je Apache Hadoop. Tento rámec umožňuje distribuované ukládání a paralelní zpracování dat. Dále můžeme využívat NoSQL, které podporuje všechny typy datových modelů. V praxi je možno je najít v podobě databází jako MongoDB, Apache Couch DB a Azure Cosmos DB. Kosinski (2024) uvádí, že nejmodernějšími řešeními jsou takzvaně datová jezera (Data lakes) a sklady (warehouses), kdy datové jezero uchovává nezpracovaná i nestrukturovaná data, která jsou pak připravená k použití. Jsou vhodná pro zpracování velkých datových objemů a machine learning. Datový sklad je systém, který získává strukturovaná data z různých zdrojů, která pak zpracovává pro provozní

⁸ Strojové učení, soustředí se na užívání dat a algoritmů tak, aby umožnili AI imitovat způsoby, jakými se člověk učí a postupně zlepšovala svou přesnost.

⁹ Dolování dat, využívá ML a statistickou analýzu k odhalení vzorců a cenných informací v rozsáhlých datových sadách.

použití, jako jsou přehledy v Power Business Intelligence (BI). Obě možnosti mají různé funkce, ale společně tvoří dobře uspořádaný systém. Nejnovějším přístupem k této problematice je kombinace obou systémů v tzv. Data Lakehouse¹⁰. V tomto případě není nutné předávat data mezi dvěma strukturami, protože k veškerému zpracování a analyzování dochází na jednom místě.

- **Integrace** – pro správnost v budoucnu zpracovaných dat je třeba je vždy vhodným způsobem integrovat. IBM Cloud Education Team (2021) dělí integrační procesy na dva možné. Extract – Transform – Load (ETL) a modernější způsob Extract – Load – Transform. Pomocí těchto principů přesouváme tzv. raw data ze zdrojového systému do cílové databáze. První z nich (ETL) postupuje přesně tak, jak napovídá jeho název. Data, která jsme získali ze zdrojových systémů jsou pročištěna a upravena (transformována) pro účely cílové databáze. Strukturovaná data dále putují do cílového úložiště. Nevýhodou tohoto systému může být časová náročnost přesunu dat. Na druhou stranu jedná se vždy o čistá a upravená data. Je tedy vhodnější pro systémy, které pracují s menšími objemy dat a nejsou tak pravidelně aktualizovány. Datové sklady se bez tohoto principu neobejdou. Novější z integračních principů (ELT) přesouvá zdrojová data přímo do cílové databáze. K jejich zpracování a čištění tedy dochází až v ní, nebo až v případě vzniku požadavku na konkrétní data. Celý proces je tedy rychlejší. Jedná se o zásadní výhodu datových jezer.
- **Zpracování** – v této fázi jsou uložená data ověřována, řazena a filtrována, čímž jsou připravena pro další použití a dochází i ke zlepšování výkonu v případě budoucích dotazů. Využívá se dvou způsobů zpracování, buď software pro integraci dat, který propojuje a konsoliduje data z různých platforem do sjednocených center, jako je datový sklad. Tím mají naši uživatelé přístup k informacím, které potřebují pro dolování dat, generování sestav business intelligence (BI) a provozní účely. Nebo systém zpracovávání dat v paměti RAM¹¹, což umožňuje získávat přehledy v reálném čase. Například Apache Spark,
- **Čištění** - konflikty, redundance, neplatná nebo neúplná pole a případné chyby ve formátování v již zpracovaných datech se opravují a čistí,
- **Analýza** – provádí se, jak bylo zmíněno, pomocí machine learning, data mining a dalších nástrojů.

¹⁰ Kombinace mezi lake a warehouse (jezero a sklad).

¹¹ Paměť počítače, do které se ukládají informace, které program potřebuje, když je spuštěný.

1.4 Cloud Computing

„Cloud Computing je založen na placeném přístupu ke službám, které nám umožňují pohodlný přístup k síti a sdílenému fondu konfigurovatelných výpočetních zdrojů na vyžádání. Jedná se o servery, sítě a služby, které mohou být pohotově zřízeny s minimálním úsilím na jejich správu nebo jakoukoliv interakcí s jejich poskytovatelem.“

(Mell a Grance, 2011)

Dle Google Cloud (2024) umožňuje využití Cloud Computingu předejít nákladnému a časově náročnému nákupu serverů a jiných prvků podobných infrastruktur. Takovéto systémy potřebují pro svou správnou funkci dostatečně výkonný zdroj energie a teplotně řízené místnosti. Dále je nutný odborně proškolený personál, který se dokáže o systém starat. Je také velmi komplikované systém dimenzovat pro případy prudkého nárůstu poptávky, nebo růstu podniku. V těchto případech společnost může skončit s větším množstvím výpočetní síly, než je sama schopna využít. Cloud Computing, jako jednoduše škálovatelné a cenově výhodné řešení, které takovýmito problémům předchází.

Cloud Computing využívá síť, nejčastěji internet, pomocí které se připojuje ke cloudové platformě. Tam je možné si vyžadovat přístup k uživateli placeným službám. Uživatel platí pouze za to, co sám využívá. Centrální server má na starosti komunikaci mezi klientovým zařízením a ostatními servery, kde dochází k výměně dat. Bezpečnost této datové komunikace je tedy důležitým aspektem. V případě firem není možno využívat pouze jeden typ Cloud Computingové architektury. Vždy je třeba upravovat systémy dle specifických požadavků a potřeb různých podniků.

Microsoft Azure (2024) dělí cloudy do třech modelů, pomocí kterých jsme schopni poskytovat Cloud Computingové služby:

- **Veřejný cloud** – veškeré výpočetní prostředky jsou dodávány přes internet. Jde o hardware (servery, úložiště), software a další infrastrukturu podporující jejich chod, které jsou vlastněny a spravovány poskytovateli cloudových služeb (třetími stranami),
- **Privátní cloud** – prostředky Cloud Computingového procesu využívá jedna společnost. Služby a infrastruktura cloudu se spravují v privátní síti. Může být tedy umístěn i v její budově. Společnosti si mohou nechávat platit za hostování jejich privátního cloudu externími poskytovateli služeb,
- **Hybridní cloud** – je kombinací mezi výše zmíněnými cloudy, kdy umožňujeme organizacím sdílet data a aplikace mezi veřejným a privátním cloudem. To přispívá ke

zvyšování flexibility a zabezpečení. Dále je organizacím umožněno lepší optimalizace a jednodušší dodržování předpisů.

Cloud Computingové služby jsou dle Microsoft Azure (2024) rozděleny do čtyřech distribučních modelů. Které se liší tím, co je v rámci služby nabízeno. Může jít o hardware, software, nebo kombinaci obojího. Někdy se jim říká „stack“ Cloud Computingu, jelikož jsou postaveny jedna na druhé. Není nutné vždy používat pouze jeden typ, je možno je různě kombinovat, klidně všechny najednou. Jedná se o:

- **Infrastructure as a Service (IaaS)** – kde si uživatelé, jak plyne z názvu, pronajímají IT infrastrukturu, jako jsou servery, virtuální počítače (VM), úložiště, sítě a operační systémy od poskytovatele cloudu. To vše za poplatek. Sandhu, (2022 s. 35) uvádí jako příklad Google Compute Engine, Microsoft Azure a HP Cloud,
- **Platform as a Service (PaaS)** – neboli „Platforma jako služba“ uživatelům nabízí prostředí pro vývoj, testování, doručování a správu softwarových aplikací. Umožňuje tak vývojářům rychle vytvářet webové, nebo mobilní aplikace, bez toho, aniž by museli spravovat, nebo nastavovat infrastrukturu serverů, úložiště, sítě a databáze, nutné pro vývoj. Jedná se opět o platformy od společností Google a Microsoft (Sandhu, 2022, s. 35),
- **Software as a Service (SaaS)** – nabízí na základě předplatného doručování softwarových aplikací přes internet. Poskytovatel cloudu se stará o podkladovou infrastrukturu, hostuje a spravuje aplikaci, ke které se uživatel pouze vzdáleně připojí pomocí běžného webového prohlížeče. Nemusí se tedy starat o softwarové upgrady, nebo opravy zabezpečení. Pro připojení k aplikacím je možno užívat běžné prohlížeče, jako je Google Chrome (Sandhu, 2022, s. 35),
- **Bezserverová architektura** – je architektura překrývající se s PaaS, kdy se uživatel vyhýbá nutnosti spravování infrastruktury anebo plánování kapacity a nastavování serverů. Může se tedy plně věnovat vytváření aplikačních funkcí. Mezi charakteristiky můžeme zařadit vysokou škálovatelnost, řízení událostmi a využívání prostředků jenom tehdy, dojde-li k využití konkrétní funkce (Microsoft Azure, 2024).

V praxi lze Cloud Computing využívat různě, buď pro analýzu dat, jejich ukládání a zálohování. Dále se zde nabízí dodávání software v podobě aktualizací, testování a vytváření aplikací, nebo vkládání inteligentních funkcí, modelů, sloužící k oslovování zákazníků a zachycování dat o nich v případě e-commerce.

1.5 Zabezpečení Big Data

S rostoucím objemem zpracovaných dat narůstá i nátlak na zabezpečení procesů, jako je zpracovávání, přenášení a ukládání dat. Velmi tedy závisí na zabezpečení cloudových a přenosových služeb, či přístupů.

Matturdi et al. (2014) vyzdvihuje hned několik výzev týkající se zabezpečení a ochrany soukromí Big Data,

- **Volume a Variety** – z charakteristiky vyplývající velký objem a různorodost dat mohou vytvářet problémy při nastavování vhodných zabezpečovacích prostředků,
- **Citlivost dat** – data mohou obsahovat citlivé informace, jejich špatná správa může vést k jejich úniku, tedy k narušení soukromí,
- **Riziko re-identifikace** – i přes to, že data mohla být již anonymizovaná, při jejich úniku může dojít ke zpětné identifikaci a k přiřazení ke konkrétním jednotlivcům
- **Neoprávněný přístup** – riziko neoprávněného přístupu se zvyšuje distribucí dat více uzly.

Společnost Microsoft (2024) zmiňuje několik hlavních aspektů cloudového zabezpečení. Jde o omezení přístupu pouze pověřeným osobám, ochranu samotných dat a infrastruktury, obnovení dat v případě jejich narušení, obvykle pomocí záloh. Dále je důležité mít plán reakce, pomocí kterého snižujeme dopady útoku a bráníme další systémy. V poslední řadě je třeba mít zajištěný přehled o stavu zabezpečení, aby došlo k eliminaci slepých míst, dále je nutné věnovat pozornost nově vznikajícím bezpečnostním hrozbám.

Zabezpečováním cloudu se předchází jak externím hrozbám, tak i ohrožování zabezpečení ze strany zaměstnanců. Microsoft (2024) nástroje na zabezpečení rozděluje na:

- **Správa stavu cloudového zabezpečení** – tyto nástroje, označované jako CSPM (Cloud Security Posture Management), analyzují prostředí cloudu, kontrolují chyby v konfiguracích a navrhuji zlepšující opatření,
- **Platforma ochrany cloudové pracovní zátěže** – neboli CWPP (Cloud Workload Protection Platforms), má na starosti bezpečnost aplikací (sítových, výpočetních a úložných). Dosahuje toho obdobným principem, tedy identifikací potenciálních hrozeb a navrhováním jejich řešení,
- **Zprostředkovatel zabezpečení přístupu ke cloudu** – monitoruje využívání cloudových aplikací, hodnotí jejich rizikovost, pomáhá splňovat požadavky na dodržování předpisů a ochraňuje data. Nazýváme ho CASB, z anglického překladu Cloud Access Security Brokers,

- **Správa identity a přístupu** – IAM (Identity and Access Management), poskytuje možnost ověřování identity a implementace více faktorového ověřování. Tím je značně omezen přístup neoprávněných osob k citlivým datům,
- **Správa nároků cloudové infrastruktury** – poskytuje přehled o tom, které identity mají přístup k různým prostředkům napříč všemi používanými platformami, například v rámci podniku, a pomáhají implementovat zásady minimálních oprávnění. Tento systém označujeme jako CIEM (Cloud Infrastructure Entitlement Management),
- **Platforma pro ochranu nativních cloudových aplikací** – přeloženo z anglického Cloud-Native Application Protection Platforms (CNAPP) pomáhá implementovat bezpečnostní prvky přímo během procesu vyvíjení. Aplikace jsou tedy zabezpečeny od jejich počátku.

1.6 Iot v kontextu Big Data

„Internet of Things (IoT) je rozsáhlá síť fyzických zařízení, které přesahují běžné počítačové sítě, a které budou v budoucnosti vytvářet v reálném čase obrovské toky Big Data.“

(De Francisci Morales et al., 2016, s. 1)

Iot a Big Data jsou navzájem úzce provázané pojmy. IoT, jako významný zdroj dat se neobejde bez technologií pro zpracování Big Data tak, aby bylo možné z nich získat chtěné informace. De Francisci Morales et al (2016) vyzdvihují tři výzvy v oblasti dat plynoucích z IoT. Jde o objem dat (Volume), který nepřehledně plyne a je třeba jej zpracovávat. Dále různorodost dat (Variety), kdy data mohou být strukturovaná, polostrukturovaná, nebo nestrukturovaná. Jako příklad můžeme uvést číselné údaje ze senzorů, nebo naopak právě třeba video soubor z kamery. Třetím aspektem je rychlost zpracování dat (Velocity). Rychlé zpracování dat umožňuje pohotově reagovat na aktuální situace.

IoT má různá užití. Velké množství lidí v současné době používá wearables, může jít například o chytré hodinky, které svým uživatelům sledují tepovou frekvenci. Toto zařízení se řadí do kategorie IoT, jelikož poskytuje konkrétní údaje a sleduje vývojové trendy, či anomálie s pomocí připojení k síti. Společnost Siemens (2024) využívá IoT v kombinaci s Big Data k monitorování výrobních linek, zvyšování efektivity, predikci údržby a zkracování prostojů. Microsoft Azure (2024) zmiňuje využívání IoT ve zdravotnictví. Kdy se pomocí senzorů a zařízení připojených k síti sleduje stav pacienta v reálném čase. Datová analytika může být lékařům dále nápomocná při identifikaci rizikových situací. Dalším odvětvím využití mohou být chytrá města (Cisco, 2024), kde by IoT senzory napomáhaly s efektivním způsobem řízení dopravy, veřejným osvětlením, nebo nakládáním s odpadem.

2 ANALÝZA PŘÍSTUPU K BIG DATA V ODDĚLENÍ LOGISTIKY ŠKODA AUTO A.S.

Oddělení logistiky ŠA generuje obrovská množství dat, která se týkají nejen interních procesů, jako jsou materiálové toky a skladování. Zahrnují dále i řízení dodavatelského řetězce, nebo na druhé straně distribuci hotových vozidel. Druhá část práce se bude na základě interních materiálů podniku zaměřovat na analýzu současného stavu přístupu k Big Data v oddělení interní logistiky svařovny v závodě v Kvasinách. Bude se soustředit na jednotlivé procesy, kde Big Data vznikají, způsob jejich zpracování, ukládání, zabezpečení a jejich formát. Dále bude obsahovat rozhovor s experty, jejichž poznatky budou společně se závěry z analýzy použity k celkovému zhodnocení současného přístupu k problematice.

2.1 Škoda Auto a.s.

Společnost Škoda Auto (ŠA) je jednou z nejdéle nepřetržitě vyrábějících automobilek na světě. Založena roku 1895 v Mladé Boleslavi, Václavem Laurinem a Václavem Klementem. Podnik se původně věnoval výrobě jízdních kol. K rozšíření sortimentu o osobní automobily došlo až roku 1905. Lídr místního trhu byl v roce 1991 začleněn do koncernu Volkswagen, což mu umožnilo se stát i mezinárodně konkurenceschopným hráčem. Společnost má hlavní sídlo a svůj největší závod v Mladé Boleslavi. Dále automobily vyrábí v závodech v Kvasinách a ve Vrchlabí. Produkce expandovala i do zahraničí a to konkrétně do Indie, Číny a na Slovensko.

V roce 2024 se společnost stala čtvrtou prodejně nejúspěšnější automobilkou na evropském trhu. Meziroční tržní podíl stoupl z 5,36% na 5,84%. I přes drobný meziroční pokles tržního podílu na domácímu trhu, stále zaujímá dominantní pozici s podílem 36,5%.

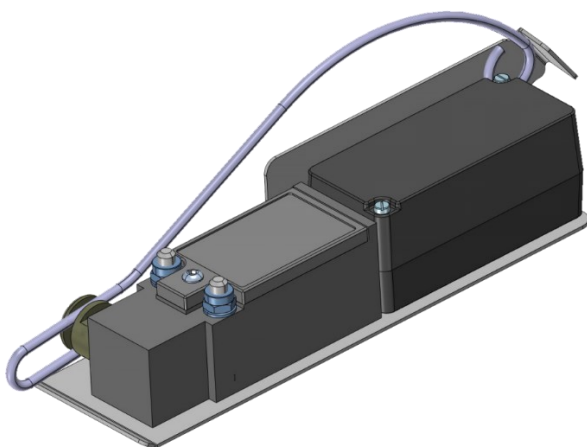
Rok 2024 se obdobně jako předcházející roky nesl v duchu investic do inovací a rozšiřování výrobních kapacit. Byla dokončena přestavba lakovny B, která zvýšila kapacitu lakovaných vozů ze 700 na 800 kusů za den. Do provozu byl dále uveden Digital Control Room (DCR), který má za úkol optimalizaci řízení výroby a logistiky.

Produktové portfolio automobilky bylo v roce 2024 rozšířeno o nový, plně elektrický model Elroq, který přinesl nový designový jazyk Modern Solid. Model navazuje na předchozí úspěchy Škody Enyaq. Dále byl představen model Škoda Kylaq, který je vyráběn pro indický trh. Společnost dále nabízí modely Kodiaq, Karoq, Kamiq, Scala, Fabia a Octavia, která si v loňském roce prošla faceliftem. Pro zahraniční trhy existují dále modely jako Kushaq, Slavia a jiné. Výroba tuzemsky prodávaných modelů probíhá převážně v závodech v Mladé Boleslavi a Kvasinách. Za rok 2024 bylo vyrobeno 925 164 vozů (Škoda Auto, 2024).

2.2 Současný stav přístupu k Big Data v oddělení interní logistiky svařovny závodu v Kvasinách

Ve svařovnách A a B se nachází více procesů, které můžeme nazvat kritickými, co se týká vzniku Big Data. Tyto procesy se opírají o tři principiálně odlišné systémy. Jedná se o systémy plně automatické, kde vznikají například data v podobě objednávek materiálu na linku jednoduchým odtížením senzoru. Dále se zde nacházejí systémy, které jsou částečně automatizované, kdy pracovník údržby provádí jistý úkon, který slouží jako impulz pro zahájení automatické sekvence. Opět může být v podobě objednávky materiálu. A posledním zdrojem Big Data jsou pracovníky manuálně zadávané informace. V tomto případě neprobíhá žádný z procesů automaticky. Pomocí automatizace jednotlivých procesů je cíleno na snížení chybovosti, zvýšení spolehlivosti a rychlosti zpracování požadavků, na druhou stranu však vznikají obrovské objemy dat, se kterými je třeba nakládat vhodným způsobem.

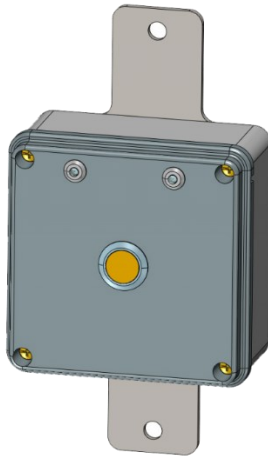
Pro správnou funkci jednotlivých procesů ve svařovně používá ŠA hned několik společně pracujících systémů. Pro plně automatizované systémy slouží radiová síť Smart Sensor Warehouse (SSW). Jedná se o vyměnitelnou baterii napájené senzory a vysílače, které společně komunikují na frekvenci 868 Mhz. Na této frekvenci jsou předávány informace o změně stavu senzorů a také tzv. „Heartbeat“, pomocí kterého ověřujeme, zda je senzor stále aktivní. Tím předcházíme situacím, kdy by nedocházelo k potřebným objednávkám materiálu. Stav nabití jejich baterie se odvozuje od celkového počtu odeslaných informací o změně stavu senzoru na základě odhadu, kolik senzory na jedno nabití vydrží. Sensorový systém je napojen na interní systém ŠA, Andon, který je přímo napojen na koncernový systém Logis¹².



Obrázek 3 Senzor SSW, (Galatech Engineering, 2023)

¹² Systém pro obecné řízení skladů a skladových zásob.

U částečně automatizovaných systémů jsou senzory nahrazeny tlačítky, případně táhly, pomocí kterých si pracovník vytvoří objednávku materiálu. Dále je zde opět vytvořeno propojení s interními systémy. Toto řešení je využíváno v případech, kdy je plně automatické řešení pomocí senzorů příliš komplikované a nákladné.



Obrázek 4 Tlačítko pro vytvoření objednávky, (Galatech Engineering, 2023)

V případě neautomatizovaných systémů, kde vznikají Big Data, je vhodné zmínit Aplikaci Mobilní Údržby (AMU). Jedná se o webové rozhraní, do kterého pracovníci údržby evidují poruchy, ke kterým ve svařovnách dojde. Aplikace slouží primárně pro evidenci oprav, ale také ke krátkodobým a dlouhodobým preventivním kontrolám zařízení. Jedná se o přívětivější reprezentaci Systemanalyse Programmentwicklung (SAP)¹³, se kterým oboustranně komunikuje.

2.2.1 Distribuce materiálu ze skladu svařovny

Pro rychlé a bezchybné zásobování výrobní linky se ve svařovně používá systém založený na principu automatických objednávek, které vznikají uvolněním pohybového senzoru. Tento senzor je umístěn na vozíku, který je opatřen skluzy, po kterých se posouvají tzv. Kleinladungsträger (KLT)¹⁴ tak, jak je pracovník spotřebovává. Každý senzor je vždy spárován s jedním typem materiálu, který má být navezen ze skladu. Každé odlehčení senzoru tedy znamená nový impuls do systému Andon, který vytvoří novou relaci v interním systému Logis. Na jejím základě je vytištěna objednávka do skladu svařovny. Každá z relací obsahuje čas, kdy vznikla, interní číslo a název potřebného materiálu a informaci o místě jeho spotřeby.

¹³ Systém pro plánování podnikových zdrojů.

¹⁴ Jedná se o standardizované plastové přepravky, určené pro manipulaci s malými díly.

Jedná se tedy o data strukturovaná, která můžeme pro účely dalšího zpracování exportovat z programu Logis do tabulky Excelu. Přístup k těmto datům je založen na principu správy uživatelů a jejich oprávnění. Distribuce materiálu ze skladu svařovny je jedním z výraznějších datových toků ve svařovně, data o objednávkách jsou tedy archivována přibližně po dva až tři týdny. Zálohují se v interní databázi ŠA, která je provozována společně se systémem Andon na databázovém serveru v Mladé Boleslavi.

2.2.2 Dílčí rozvoz materiálu ze skladu M7

V případě vysokoobrátkových dílů využíváme takzvaný dílčí rozvoz. Objednávka v tomto případě vzniká principiálně stejně, tedy odlehčením senzoru. V tu chvíli pracovník skladu dostává do svého Personal Digital Assistant (PDA)¹⁵ informaci o tom, jaký materiál je třeba na linku navézt. Automatizovaný systém mu tedy naplánuje optimalizovanou trasu tak, aby nabral potřebný materiál co nejefektivněji. V případě rozvozu materiálu systém opět poskytuje nejefektivnější trasu. Rozdíl oproti dílům, jejichž spotřeba není tak vysoká je, že pracovník skladu zboží nabírá přímo z palety. Materiál není ve skladě překládán do KLT, tím snižujeme počet manipulačních úkonů.

Pomocí odlehčení senzorů vnikají data o typu potřebného materiálu, místu jeho spotřeby a času vzniku objednávky. PDA přebírá základní informace o objednavce a navíc poskytuje data o stavu daných položek s časem změny stavu. Tedy že položka je aktuálně naložena, navezena, vykládána apod.



Obrázek 5 Personal Digital Assistant, (Zebra Technologies, 2025)

Jednotlivé relace lze exportovat do formy obdobně jako u běžně distribuovaného materiálu ze systému Logis do Excelové tabulky. Princip zálohování a přístupů je obdobný.

¹⁵ Odolná mobilní zařízení navržena pro efektivní správu skladových operací.

Vzhledem k tomu, že se jedná o vysokoobrátkový materiál, jsou tato data skladována po dva týdny.

2.2.3 Mezioperační převozy dílů

Mezioperační převozy dílů ve svařovnách můžeme rozdělit na dva typy. Prvním jsou díly, které vznikly v jedné ze svařoven v Kvasinách, a je potřeba je přesunout mezi dvěma místy ve svařovně (bočnice určená k dalšímu sváření). Druhým typem jsou díly, které byly původně vyrobeny v závodě v Mladé Boleslavi a dále budou zpracovány v kvasinské svařovně, jsou tedy uskladněny poblíž místa zpracování. Oba typy dílů se před zpracováním svařovacím automatem zakládají za posuvná vrata, vždy do dvou pozic nad sebou. Vrata tedy jednu z pozic uzavřou a druhá zůstává přístupná. Pro účely automatizace byla vrata byla opatřena senzory SSW, pomocí kterých snímáme jejich pozici. Toto řešení bylo technicky méně náročné a ekonomicky výhodnější, než vytvářet spojení mezi elektronickým ovládním vrat a systémem Andon. Posunem vrat do opačné pozice dochází k odlehčení senzoru a tedy k vytvoření objednávky v systému Logis (skrze Andon), podobně jako v případě skluzů. Jedná-li se o materiál ze skladu, putuje požadavek skrze Logis do PDA pracovníka skladu, který materiál následně naveze. Materiál, který byl předem zpracováván na jiném stanovišti ve svařovně je na základě požadavku, který putuje přímo do PDA převezen pracovníkem svařovny na místo další spotřeby.

Vzhledem k tomu, že se opět jedná o data, jejichž původ je sensorový systém SSW, jsou tedy opět přístupná v systému Andon a mají pevně stanovenou strukturu. Uchovávají se po dobu dvou až čtyř týdnů v interní databázi ŠA.

2.2.4 Automatické objednávání a kontrola návozu sudů ze skladu barev

V místech, kde je spotřebováván materiál, který se skladuje v sudech, jako například tmely a lepidla, je důležité nevytvářet prodlevy při výměně náplní a hlavně nesmí docházet k jejich záměnám. Za provozu je vždy jeden sud zpracováván a druhý, se stejným materiálem, je připraven k případné výměně poblíž místa spotřeby. Každý ze sudů je označen Quick Response (QR)¹⁶ kódem, který pracovníci údržby při výměně skenují, aby nedocházelo k záměnám materiálu. Kontrola správného materiálu určeného k záměně probíhá v rámci údržbářského systému Work Test Management (WTM)¹⁷, který posílá informaci, že došlo

¹⁶ Typ čárového kódu, který lze naskenovat fotoaparátem chytrého telefonu, lze jej použít k rychlému sdílení informací.

¹⁷ Aplikace pro potřeby údržby strojních zařízení

k záměně a je tedy třeba materiál doplnit, do Andonu. Andon dále generuje odvolávku do Logisu a ten upozorňuje pracovníka skladu skrze PDA.

Pracovník v tomto případě načtením čárového kódu vytváří podnět ke vzniku automatické objednávky. Každá z těchto relací obsahuje informace o jaký typ materiálu se jedná, čas vzniku objednávky a místo spotřeby, kam má být sud naveden. Přístup k datům je zajištěn skrze rozhraní Andon, jako v předchozích případech. Odvolávková data se skladují po jeden měsíc.

2.2.5 Objednávání Gitterboxů

V případech, kdy by instalace senzorů SSW byla nevhodná, nebo technicky neproveditelná, používáme tlačítka, nebo táhla, pomocí kterých vytváříme objednávku. Taková situace může nastat například v místě spotřeby drobného materiálu, který z praktických důvodů ponecháváme v Gitterboxech (GLT)¹⁸. V případě, kdy pracovník pracující v daném místě spotřeby vidí, že materiál dochází, stiskne tlačítko a vytvoří tím objednávku materiálu, se kterým je tlačítko spárováno. V místech, kde není možná instalace tlačítka, se používají táhla.

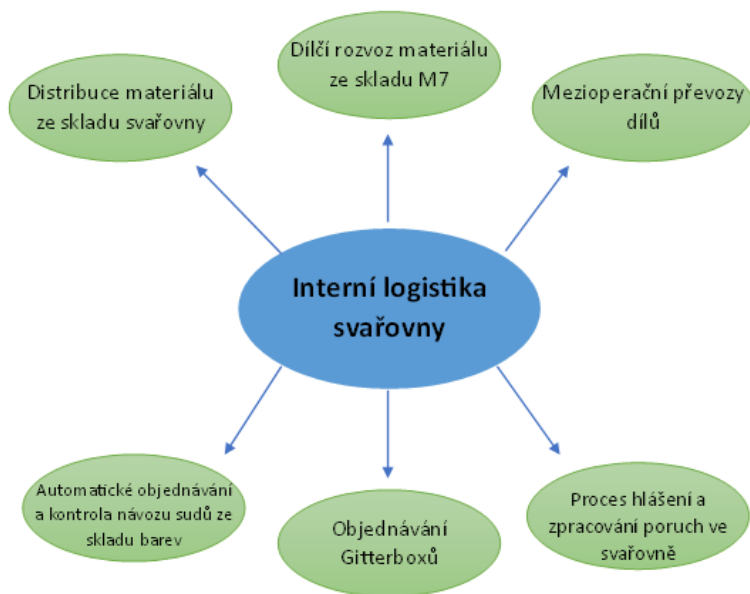
Datový tok je v tomto případě stejný jako u senzorů SSW, přístupy uživatelů a způsob uchování také. Tím, že objednávané velké množství materiálu najednou, nám nevznikají tak velká množství dat, jako u ostatních typů objednávek. Data zde tedy můžeme skladovat po dobu až 6 měsíců.

2.2.6 Proces hlášení a zpracování poruch ve svařovně

Big Data ve svařovně nevznikají pouze v rámci zásobování výrobního procesu. Velké datové toky existují i ve formě zpráv o údržbě. K tomu slouží webové rozhraní AMU. Jedná se o systém hlášení o poruchách a průběhu jejich řešení za posledních 24 hodin. Jsou vždy ve formě dokumentu PDF, který na svém začátku má stručný soupis všech závad, ke kterým v daném období došlo s časem kdy nastaly a kdy byly vyřešeny. Dále následuje výpis jednotlivých závad s potřebnými detaily. Jedná se o jednotlivé kroky, pomocí kterých pracovník údržby závadu opravoval, včetně doby, po kterou trvalo provést jednotlivé úkony. Pokud by to bylo užitečné, může být report opatřen fotografiemi závady. Dále je zde vždy uvedeno jméno pracovníka a místo, na kterém k opravě došlo. Na závěr se do reportu uvádí příčina vzniku dané závady, například běžné opotřebení zařízení, a případné opatření tak, aby se do budoucna takovým závadám předcházelo.

¹⁸ Paletové boxy z plastu.

Informace o průběhu opravy může pracovník zadávat buď formou předem připravených úkonů a příčin, ze kterých může vybírat. Existuje ale i možnost zadávat vlastní poznámky a úkony. Může se tedy jednat o nestrukturovaný typ dat. Přístup je řešen formou oprávnění uživatelů. Buď pro běžný přístup nebo pro úpravu dat v programu SAP. Databáze těchto reportů se uchovává po neomezenou dobu v SAP: Plant Maintenance¹⁹.



Obrázek 6 Kritické procesy pro vznik Big Data v oddělení interní logistiky svařovny, (autor)

2.3 Rozhovor s experty

Cílem tohoto rozhovoru bylo vyzdvihnout případné nedostatky jednotlivých systémů a řešení krizových situací, které mohou během výrobního procesu nastat. Na dotazy týkající se procesů využívající systém SSW, automatického objednávání a kontroly návozu sudů ze skladu barev a objednávání GLT odpovídal pan Jan Anderle, který se zabývá projekcí logistických projektů. Pro poslední část rozhovoru, proces hlášení a zpracování poruch ve svařovně, poskytnul své poznatky Mgr. Jan Krystuf. Rozhovor proběhl 27.3.2025 online formou. Jedná se o jeho doslovný přepis.

¹⁹ Modul údržby programu SAP

Existuje nějaká slabina systému objednávek pomocí SSW? Pokud ano, jak a jak často se projevuje? Řeší se aktuálně nějak?

„Největším problémem jsou mechanické závady, ať už na regálech, nebo přímo na senzorech. Pokud sensor zůstane odtížen i přes to, že je ve skluzu materiál, pak dochází k nadbytečným objednávkám materiálu a následným vratkám. Horší případ je, když sensor zůstane zatížen i přes to, že ve skluzu není dostatek materiálu, dochází tak k zastavení toku materiálu na linku, v krajním případě vedoucí k výrobě nekompletního vozu nebo zastavení výrobní linky. Pracovníci výroby a plánovačky pravidelně monitorují tok materiálu a v případě pochybností kontaktují náš SSW tým, který závadu prověří a případný problém vyřeší.“

- **Dochází k sekundárnímu využití dat o objednávání a navážení materiálu? Je současná struktura těchto dat vhodná pro další analýzu?**

„Data ze sensorů a objednávek si uchováváme a v případě problémů jsme schopni zpětně zjistit, co se na daném místě dělo a jak bylo s materiálem ve skluzu monitorováno. Stejně tak jsme z nich schopni historicky počítat obrátky pro konkrétní místo spotřeby a sledovat vývoj potřeby materiálu v čase.“

- **Jaká je spolehlivost přenosu dat mezi PDA a interními systémy? Co se děje v případě výpadků připojení, pokud k nim dochází?**

„K výpadkům pochopitelně dochází, ne často, ale dochází. Abychom škody způsobené možným výpadkem minimalizovali, pro všechny operace máme náhradní technologie. V případě výpadku objednávání Andonem plánovačky objednávají materiál pomocí systému Internes Transport-LeitSystem (iTLS)²⁰. V případě výpadků na dílčích rozvozech jsou skluzy na lince opatřeny kanban kartami²¹. Pracovník pak jezdí a hledá na lince, kde je nedostatek materiálu, tam si kartu vezme k sobě a ve skladu dle informací na kartě materiál vychystá a přiveze. Tyto výpadky pravidelně testujeme tak, aby pracovníci znali náhradní technologii a byli schopni ji v případě neplánovaného výpadku používat, pokud možno s nulovým nebo minimálním dopadem do výroby.“

- **Lze nějak rozeznat, že byla objednávka pomocí senzoru vytvořena omylem (objedávka vznikne podezřele brzy)? Obdobně v případě duplicitních objednávek. Dokáže systém do jisté míry tyto situace filtrovat?**

„Tyto situace neřešíme. Po odtížení senzoru systém čeká 7 minut a pokud během tohoto času nedojde k opětovnému zatížení senzoru, pak teprve systém vygeneruje odvolávku. Různé

²⁰ Systém pro interní transporty materiálu využívaný koncernem Volkswagen

²¹ Nástroj v rámci systému Kanban, slouží k předání impulzu k zahájení další činnosti

materiály mají v paletě různá množství dílů a jejich potřeba na lince se může u některých dílů výrazně měnit. Je tedy rizikové pokoušet se toto předvídat. Navíc vždy je lepší přivést jedno KLT navíc, než aby linka zůstala bez materiálu.“

- **Mají jednotlivé objednávky v systémech vlastní priority, nebo si jsou všechny rovny?**

„V systému pracujeme se dvěma úrovněmi priority, STANDARD a URGENT. V případě SAS senzorů ale nemáme možnost ani důvod dávat objednávkám vyšší prioritu. Priority využíváme u tlačítek, dlouhý stisk = URGENT.“

- **Jsou data, která získáváme z objednávek a rozvozu materiálu vhodná pro vyhodnocování efektivity zásobování linky?**

„Pro tyto účely se dají použít data z jednotlivých senzorů. Ukazatelem je čas od vygenerování odvolávky na materiál po doručení materiálu (ať už systémové, nebo fyzické zatížení senzoru).“

- **Jaká je poruchovost senzorů SSW?**

„Z více než 6500 senzorů, které nyní v Kvasinách máme, jsou to jednotky prověřených/opravených/vyměněných senzorů denně. Na svařovně nám občas způsobuje mechanické problémy se senzory prach ze svařování. Občas je tedy nutné senzory, které začnou vykazovat závadu, vyčistit.“

- **Uvažuje se o využití AI k predikci časů objednávek na základě historických dat, k optimalizaci tras apod?**

„Máme podobnou funkcionalitu na montáži pro robotický sklad AKL. V případě výpadku aplikace predikuje seznam objednávek, které by pravděpodobně přišly. Přesnost ale není pro dlouhodobé použití vysoká, protože spotřeba materiálu na lince se mění.“

- **Jak často dochází k nesouladu mezi tím, co bylo fyzicky navezeno, a co je zaznamenáno v systému (Proces automatického navážení a objednávání sudů)?
Jak se takové situace případně řeší?**

„Je to výjimečná situace, ale důsledky záměny mohou být značné. V současné době je záměna téměř nemožná. Kontrolu provádíme při doručení sudu v systému Andon. Pracovníci údržby evidují výměny sudu v systému WTM, který riziko záměny také hlídá. Posledním stupněm je skener na pumpě, který neumožní po výměně materiálu pumpu spustit, dokud pracovník nenačte štítek na sudu. Buď by tedy muselo dojít k fatálnímu selhání pracovníka logistiky i údržby nebo k záměně štítku na sudu u dodavatele.“

- **Dohlíží systém na situace, kdy došlo k duplicitní objednávce sudů, nebo naopak neobvykle dlouhou dobu k žádné nedošlo?**

„K duplicitní objednávce dojít nemůže, nová objednávka nelze vytvořit, dokud není ta předchozí korektně uzavřena.“

- **Jsou data o spotřebě sudů s materiálem sekundárně zpracovávána?**

„Jsou uchovávány a v případě nesrovnalostí v navázení analyzovány. Víme tedy, kdy a který pracovník údržby sud vyměnil, a také kdy a kdo objednávku zpracovával na straně logistiky.“

- **Existuje nějaké slabé místo systému automatického objednávání a návozu sudů, či jeho pravidelně opakovaná závada?**

„Jediným problémem, který tu občas vzniká je, že pracovník údržby provede výměnu sudu bez zadání do systému WTM. Logistika se tedy o výměně nedozví a nepřiveze další materiál. Většina pump má ale dvě pozice pro zpracování a jedno vychystávací. Pracovník tedy nemá žádný sud k výměně, ale máme čas jej doručit, než dojde ke zpracování druhého sudu.“

- **Jsou tendence k rozšíření systému o predikční model (AI), sloužící k odhadu blížící se potřeby vyměnit sud s materiálem?**

„Vzhledem k tomu, že máme vždy jeden sud ve vychystávací pozici, není k nějaké predikci důvod. Zjednodušením by bylo, kdyby skener na pumpě byl schopen v plném rozsahu nahradit WTM. Tedy evidovat kontroly konkrétním uživatelem, hlídat expirace materiálu a posílat Andonu informaci, že došlo k výměně sudu.“

- **Trpí systém objednávání pomocí tlačítek / táhel nějakou specifickou chybou, která se vyskytuje sporadicky?**

„Jediná chyba, která může nastat je, že uživatel včas nezmačkne tlačítko a nestihne si tak objednat včas další materiál ke zpracování.“

- **Je možné, aby došlo ke vzniku duplicitních objednávek, například dvojitým stisknutím tlačítka? Dokáže tuto chybu software vyhledit?**

„Nová objednávka nelze vytvořit bez toho, aniž by byla ta předchozí uzavřena.“

- **Jak časté jsou závady tlačítek a táhel sloužící k objednávkám?**

„Velmi zřídka, navíc diagnostika je jednoduchá. Pokud se u tlačítka po stisknutí nerozsvítí kontrolka, pak je někde chyba a je nutné materiál objednat náhradní technologií a zajistit opravu.“

- **Je možné data o objednávkách využít k predikci budoucích potřeb materiálu, nebo je jejich spotřeba příliš nepravidelná?**

„Spotřeba je spíše nepravidelná, vznikalo by množství varování, která by byla neodůvodněná a uživatel by na ně po čase přestal reagovat úplně.“

- **Dochází při vkladní informací do AMU k chybám nebo nejednotnému zápisu? V případech, kdy pracovník nevybírání z předdefinovaných závad. Popřípadě k jakým?**

„Ano, k chybám a nejednotnému zápisu dochází, protože každý pracovník má nějaké „svoje“ názvosloví a dochází k překlepům apod.“

- **Jaký je přibližný poměr mezi informacemi zadávanými pomocí předem definovaných závad a pomocí vlastních poznámek?**

„Přibližný poměr – 80:20 (definované : vlastní zápis)“

- **Poskytují záznamy dostatek informací pro analýzu příčin a možnost jim v budoucnu předcházet?**

„Záznamy poruch neslouží k analýze příčin. Pouze pro evidenci výpadků ve výrobě.“

- **Lze data vložená do AMU zpětně filtrovat a dále zpracovávat? Pokud ano, zpracovávají se data sekundárně a za jakým účelem?**

„Záznamy poruch neslouží k analýze příčin. Pouze pro evidenci výpadků ve výrobě.“

- **Lze data vložená do AMU zpětně filtrovat a dále zpracovávat? Pokud ano, zpracovávají se data sekundárně a za jakým účelem?**

„Ano, AMU je pouze nadstavba nad SAP PM databází, která následně umožňuje další filtrování a zpracování dat. Data se dále zpracovávají za účelem plánovaných údržeb, slouží k statistickým účelům a výpočtu využitelnosti linek.“

- **Používají se data v současné době jako podklad k plánování budoucí údržby, nebo slouží pouze jako dokumentace minulosti?**

„Ano, jak bylo zmíněno, podle výskytu častých poruch a dalších, se data využívají pro plánování budoucí údržby.“

- **Uvažuje se o užití AI k predikci závad na základě historie poruch? Dále k jejímu využití pro analyzování reportů a fotografií v nich obsažených?**

„Nevíme o žádném plánovaném projektu na podobné téma.“

2.4 Zhodnocení aktuálního stavu

Na základě popisu jednotlivých procesů a konzultace s experty jsme získali ucelený obraz o současném stavu jednotlivých procesů v rámci oddělení. Pomocí těchto poznatků jsme schopni vyzdvihnout jak silné, tak slabé stránky přístupu k datům a samotnému fungování procesů.

Každý ze systému stojí na principiálně jednoduché technice a softwaru, které jsou doplněny o mnoho preventivních řešení. Jako příklad by bylo vhodné uvést 7minutovou rezervu, která zabraňuje falešným odvolávkám materiálu, Heartbeat SSW senzorů, který hlídá aktivitu jednotlivých senzorových zařízení, nebo využití QR kódů, bránící záměně sudů. Všechny systémy mají i své zálohy, pro případ jejich poruchy. Současná architektura jednotlivých systémů fungujících ve svařovně může být označena za velmi robustní, spolehlivou a stabilní. Má však svá omezení

V oddělení interní logistiky svařovny vznikají velké datové objemy. Na základě teoretického vymezení, v kapitole 1.1 Charakteristika Big Data, je však za Big Data považovat nemůžeme. Společnost SAP (2025) ve své definici zmiňuje, že „Big Data jsou definována jako „velká“ nejen kvůli svému objemu, ale také kvůli rozmanitosti a složitosti své povahy.“ Tato charakteristika odpovídá uznávanému modelu 3V.

Přestože jsou v rámci výrobního procesu generovány velké objemy dat, nedochází v tomto případě k přesažení limitů tradičních databázových systémů (Volume), tak jak uvádí jedna z definic (SAP, 2025). Ačkoliv jsou data generována kontinuálně, skladování v rozmezí 2 až 6 týdnů nám neumožňuje aplikovat pokročilé analytické metody, jako při běžné práci s Big Data. Dalším rysem Big Data je jejich rychlost (Velocity), která se týká jejich vzniku, zpracování a uskladňování. V případě dat ve svařovně dále nedochází ke zpracování dat v reálném čase, čímž by byl tento aspekt naplněn. Data sice vznikají průběžně, k jejich zpracování dochází však se zpožděním. V některých případech nedochází k sekundární analýze dat vůbec. Příkladem mohou být reporty AMU, které slouží pouze jako evidence. Poslední ze základních dimenzí Big Data je různorodost (Variety). Ve svařovně je většina dat strukturovaná, což nám zjednodušuje jejich zpracování a způsob ukládání. Výjimku tvoří nestrukturovaná poznámky a popisy závad, které pracovníci vkládají do reportů AMU. Jedná se o přibližně 20% dat vložených do záznamů, které mohou obsahovat cenné informace, ale je komplikované s nimi dále pracovat.

I přes to, že procesy ve svařovně generují rozsáhlé objemy dat, nenaplnují však základní charakteristiky Big Data. Nevyužitý potenciál Big Data analytiky je zde hlavně v podobě absence dlouhodobého uchovávání dat a datové analýzy v reálném čase. Vhodným řešením by

byla koncepční změna přístupu k nakládání s daty. Tedy nástroje pro dlouhodobou archivaci a implementace pokročilých analytických nástrojů. Což by umožnilo z dat získávat informace, ke kterým není možné se v menším měřítku dostat, jak napovídá definice Mayer-Schönbergera a Cukiera (2014). Návrhy na změnu přístupu a nová řešení se bude zabývat následující část práce.

3 NÁVRHY NA ZLEPŠENÍ PŘÍSTUPU K BIG DATA V ODDĚLENÍ LOGISTIKY ŠKODA AUTO A.S.

Z analytické části této práce vyplývá, že současný stav přístupu k datům v oddělení interní logistiky svařovny nenaplnuje základní principy Big Data. Cílem této části práce je na základě zjištěných nedostatků navrhnout koncepční řešení přístupu k nakládání s daty. To v budoucnu umožní efektivní sběr a uchovávání všech datových formátů, jejich následné zpracování, vzájemnou integraci, analýzu, predikci a vizualizaci.

V následujících kapitolách bude navržena obecná struktura systému, vhodná pro nakládání s Big Data. Dále zde budou popsány možnosti aplikace těchto struktur ve výrobním procesu a jejich možný přínos. Cílem této části je poskytnutí dlouhodobě udržitelného konceptu, který bude podporovat rozhodování na základě dat a informací z nich získaných a také návrh možných technologických řešení, které v současné době nabízí trh.

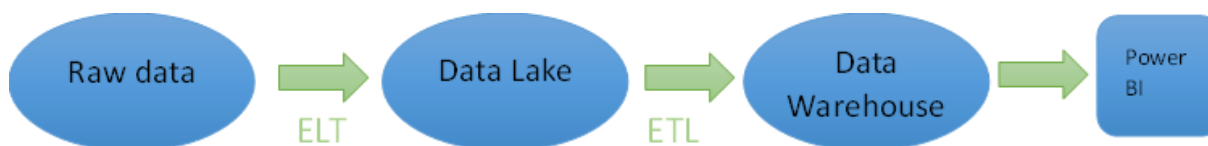
Pro uchovávání Big Data je v současné době běžnou praktikou využívat již zmíněná Data Lakes (DL), česky datová jezera. Jsou vhodná pro zpracovávání velkých datových objemů, které mohou být v libovolných formátech. Tedy strukturována, semistrukturovaná a nestrukturovaná. Data nejsou až do chvíle, kdy jsou potřeba k dalšímu zpracování, svázána, ani filtrována na základě pravidel tradičních databází. Jak napovídá teorie, datová jezera nemusí nutně nahrazovat tradiční datové sklady. Ba naopak je vhodné, aby se tyto dvě struktury navzájem doplňovaly, nebo abychom využívali Data Lakehouse.

3.1 Kombinace Data Lake a Data Warehouse

Jedním z možných řešení pro oddělení interní logistiky svařovny ve Škoda Auto a.s. by byla hybridní architektura kombinující Data Lake a Data Warehouse. Toto řešení by mohlo částečně využít současné databázové architektury, která se v podniku používá.

Toto technické řešení v prvním kroku zpracovává surová podniková data ve všech formátech do DL pomocí integračního principu ELT. Tento způsob integrace zajišťuje plynulý přesun dat, bez případného zpomalení vlivem možnosti vzniku chyby během procesu transformace dat. K té dochází až po kompletním přesunu dat do DL. Mimo možnosti nepřetržité manipulace s daty je u DL dále možnost skladovat velké datové objemy, které nám v budoucnu umožní jejich využití pro další analýzu (predikci). Data Lakes jsou dále dobře škálovatelná, což je zásadní výhodou v případě neustálého nárůstu objemů interních podnikových dat. Data, která jsou nyní uskladněna v DL jsou v dalším kroku transformována pro potřeby Data Warehouse, kam budou dále přesunuta. Jedná se o čištění, sjednocování

formátů, jazyků a jednotek, nebo odstraňování duplicit. Data nahraná do DW jsou v tuto chvíli dále dostupná pro další analýzu, nebo vizualizaci v podobě Power BI reportů. Mezi Data Lake a Data Warehouse tedy využíváme integračního principu ETL, který může proces zpracování dat zpomalovat.



Obrázek 7 Datový tok kombinovaného řešení Data Lake a Data Warehouse, (autor)

Tuto hybridní architekturu by bylo možné v oddělení interní logistiky svařovny ve Škoda Auto a.s. aplikovat ve více případech. První z možných aplikací je propojování dat z různých datových zdrojů. Konkrétně by se jednalo o informaci, že došlo k zatížení senzoru SSW pomocí KLT, tedy konkrétní čas, kdy byl materiál doručen na místo svého určení. Tato informace by byla uchováвана v DL, kvůli velkému množství dat vzniklých ze senzorů SSW. Dále by tento údaj byl propojen s časem, kdy pracovník pomocí PDA naskenuje materiál, který zrovna vykládá. Ten by byl skladován v Data Warehouse z důvodu častého dotazování a možnosti propojení s dalšími zdroji strukturovaných dat (jména pracovníků, směny). Propojením těchto dvou údajů by umožnilo detailněji sledovat efektivitu vykládek materiálu. Na základě toho by bylo dále možné rovnoměrněji plánovat pracovníky jednotlivých směny, sledovat ergonomii přepravních vozíků, nebo zda je současné rozložení míst, kde dochází k vykládání materiálu, vhodné. Obdobně by bylo možné aplikovat tento princip například v případě objednávek GLT. V tomto případě by byla propojována data o založení materiálu do automatického stroje, skladovaná v Data Lake, a informace o objednání materiálu pomocí tlačítka. Propojování dat z různých zdrojů by umožnilo odhalit místa, která jsou pracovně příliš vytížená, nebo naopak nedostatečně vytížená. Také by bylo jednodušší odhalit skryté slabiny logistických procesů ve svařovně. Další úroveň datové analýzy v oddělení interní logistiky svařovny ve Škoda Auto a.s. by bylo sledovat dopad mimořádných situací ve svařovně na materiálový tok. Jednalo by se opět o vhodné propojování dat z různých zdrojů. Šlo by o sledování doby jednotlivých výpadků, zpoždění jiného materiálu, nebo přesměrování materiálového toku. Na základě analýzy těchto dopadů by vedoucí směny a pracovníci logistiky věděli, čemu věnovat více pozornosti tak, aby předešli nekompletním vozům, nebo zastavení výrobní linky.

Pro plynulý přechod dat mezi DL a DW je třeba využívat metadatovou vrstvu, která zde slouží jako sjednocovací prvek. Pomocí této vrstvy jsme schopni data kategorizovat, což nám umožňuje je organizovat a vyhledávat v obou strukturách úložiště. Jsme schopni dohledat původ dat, historii úprav a dále řídit přístupy k datům. V případě této architektury nemusí být tato vrstva součástí technického řešení. Její implementace je však pro správné nakládání s daty klíčová. Na trhu existuje více řešení, například Apache Atlas, nebo Microsoft Purview.

V případech, kde je třeba mít okamžitý přístup k datům, je vhodné aby se data pohybovala v podobě nepřetržitého proudu (stream) přímo do datového jezera. Jsme tímto schopni okamžitě reagovat například na vzniklou závadu ve svařovně. Pro případy přehledů o výkonu za danou směnu, nebo daný měsíc se data mohou pohybovat v dávkách (batch), dle potřeb. Může se jednat o data ze systému SAP, reporty z aplikací a podobně.

Pro splnění požadavku škálovatelnosti je v tomto případě vhodné využít cloudových služeb v modelu PaaS (Platform as a Service). Tento model umožní soustředění na vývoj a správu datové architektury, bez nutnosti se zabývat infrastrukturou. Trh v současnosti nabízí mnoho možností, jde například o Azure Data Lake, Amazon S3, nebo BigQuery od společnosti Google.

Tato kombinace je velmi spolehlivým a současně flexibilním řešením. Přináší podporu jak v real-time situacích, tak pro dlouhodobou analytickou činnost. Pomocí té bude v budoucnu možné využívat velmi spolehlivých predikčních nástrojů.

3.2 Data Lakehouse

Jedná se o sjednocení dvou předchozích řešení, pomocí kterého dosáhneme centralizované a velmi spolehlivé struktury pro práci s daty. V případě Data Lakehouse (DLH) odpadá nutnost přesunu dat mezi DL a DW pomocí integračního procesu ETL, který může být zdrojem chyb a tedy prodlevy. Důvodem je, jak již bylo zmíněno, že k veškerému zpracování dat dochází na jednom místě (DLH) a současně nejsme omezeni na žádný datový typ. Tato struktura by tedy umožnila ukládat veškerá potřebná data z oddělení interní logistiky svařovny rychleji a spolehlivěji, než v případě prvního návrhu, bez nutnosti předchozí strukturalizace, nebo úprav.

Data jsou nahrávána do Data Lakehouse „ELT-like“ principem, transformace je však v tomto případě nedílnou součástí samotné platformy. Data jsou dále na základě požadavků zpracována a jsou přímo přístupná pro software jako je Power BI, po vzoru datových skladů, tak Machine Learning nástroje, jako v případě tradičního datového jezera. V praxi by tedy bylo možné vytvářet dashboardy v Power BI zobrazující pohyby materiálů, predikující výskyt

poruch, nebo monitorující využití manipulační techniky. Možnost rychlé odezvy a přístupu k datům, ze všech zdrojů, dále v praxi umožňuje například zpětnou analýzu problémových směn. To vše díky možnosti kombinování různých dat, bez jejich nutné přípravy, nebo rozsáhlého filtrování. Data Lakehouse také přináší možnost rozpoznávání nárůstu požadavku na materiál dopravovaný na linku, pomocí porovnávání současného požadavku, který máme okamžitě dostupný, a dlouhodobého trendu.

Toto řešení je spojeno s vysokou škálovatelností a nižšími provozními náklady, než v případě prvního návrhu. Dále se zde zjednodušuje řízení přístupu uživatelů, kvality dat a celkové datové integrity podniku. Důvodem je metadatová vrstva, která je vždy součástí Data Lakehouse. Pomocí této vrstvy, jsme schopni řídit přístupy k datům, sledovat kompletní datové toky, historické verze a úpravy dat. Dále nám je umožněno kategorizování dat, jak strukturovaných, tak nestruturovaných.

S rostoucím množstvím dat je důležité zachování jejich konzistence a integrity. Velkou výhodou tohoto řešení je tedy podpora transakcí²² ACID, která je v případě DW také možná. Datová jezera ji však bez dodatečných vrstev, jako je Delta Lake nebo Apache Iceberg, nepodporují. Podpora těchto transakcí je možná právě kvůli přítomnosti metadatové vrstvy. Tabulka 2 uvádí zkratky, které označují skupinu vlastností, které by každá transakce měla splňovat

²² Jakákoliv operace, která je považována za jednu jednotku práce, která proběhne celá, nebo neproběhne vůbec a zanechá úložiště v konzistentním stavu.

Tabulka 2 Transakce ACID

Zkratka	Popis
A - Atomicita	Transakce se provede buď celá, nebo vůbec. To v případě, kdy jakákoliv z dílčích částí selže
C - Konzistence	Data jsou po každé operaci v platném a konzistentním stavu.
I - Izolace	Pokud probíhá více transakcí zároveň, nezasahují si vzájemně do dat. Vždy se čeká, než první transakce skončí.
D - Trvalost	Jakmile je daná transakce ukončena a změny jsou zapsány do databáze, měly by být trvalé a tedy zůstat v databázi i v případě výpadku elektrické energie apod.

Zdroj: MongoDB (2025), upraveno autorem

Vhodným řešením, které nabízí trh je od společnosti Microsoft, Fabric. Jedná se o řešení formou SaaS, které využívá cloudového úložiště OneLake. Do systému jsou integrovány prvky AI a jeho součástí jsou i analytické prvky jako Power BI. Tato platforma splňuje všechny výše zmíněné požadavky na DLH. Od strojového učení, přes podporu transakcí ACID, až po reportingové nástroje. Fabric je vhodné doplnit o Microsoft Purview, pomocí kterého jsou spravována metadata.

Vzhledem k současnému vývoji digitalizace výrobních procesů, velkému nárůstu datových objemů a požadavkům směřující k real-time analytice a predikci trendů, je tato architektura vhodná pro nasazení ve Škoda Auto a.s. Spojuje možnosti efektivního reportingu, která je typická u datových skladů a flexibilitu datových jezer.

3.3 Power BI reporting

Oba předcházející návrhy vyžadují rozsáhlou infrastrukturu. Je však vhodné zmínit, že je možné získat centralizovaný a přehledný přístup k potřebným logistickým ukazatelům na základě několika zdrojů dat, bez nutnosti rozsáhlých změn interní datové architektury. Reporty v Power BI od společnosti Microsoft by mohly být vhodným prvním krokem k rozsáhlé datové analýze.

Software Power BI je SaaS platformou umožňující vytváření reportingových sad, dle vlastních potřeb a požadavků. Umožňuje spojovat různé datové zdroje, například ve formě tabulky excelu a vytvářet z nich jednotné informační panely, které jsou vhodným podkladem pro operativní rozhodování, nebo odhalení slabých míst.

V případě interní logistiky svařovny, by bylo možné vytvořit informační panely pro jednotlivé uživatelské role, vždy se zaměřením na konkrétní Key Performance Indicators (KPI)²³. Rozdělení informačních panelů do Power BI dle rolí, zaměření a KPI's je shrnuto v tabulce 3.

Tabulka 3 Informační panely v Power BI dle uživatelských rolí

Role	Zaměření	KPI
Pracovník logistiky	Stav objednávek, zásobování linky, vytížení tras	Průměrná doba mezi vytvořením objednávky a jejím doručením, počet objednávek za směnu/den
Mistr výroby	Stav zařízení, jejich výpadky a jejich vliv na výrobu	Počet poruch za danou směnu, průměrná doba výpadků
Pracovník údržby	Opakující se závady, efektivita oprav	Počet zásahů za směnu, průměrná délka opravy, seznam opakujících se poruch
Pracovník plánování výroby	Trendy spotřeby materiálu, predikce zpoždění	Průměrné zpoždění dodávek na linku, spotřeba materiálu za daný časový úsek

Zdroj: autor

System by v první řadě mohl být postaven na datech exportovaných ze systému Logis. Ten v sobě uchovává informace o jednotlivých objednávkách, jako je typ materiálu, čas vzniku objednávky, stav zpracování apod. Dalším ze zdrojů by mohly být reporty z AMU, respektive data ze SAP PM, vzhledem k tomu, že AMU je pouze nástavbou tohoto software. I přes to, že část dat v reportech je nestrukturovaná, informace jako čas vzniku poruchy a typ poruchy z nich

²³ Klíčové ukazatele výkonnosti

však je možné extrahovat. Dalším zdrojem pro vizualizaci potřebných ukazatelů je možné využít různé tabulky excelu, které pracovníci daných úseků využívají pro vlastní účely.

Aplikace tohoto vizualizačního analytického nástroje s sebou nese, mimo nízké vstupní náklady a možnost rychlého zavedení, hned několik přínosů. Díky přehlednosti a možnosti personalizace jednotlivých dashboardů jsme schopni lépe identifikovat opakující se problémy ve výrobním procesu. Možnost rychle vizualizovat data dále hraje důležitou roli v případě operativního rozhodování. Celkově je tento první krok vhodný pro podporu zavedení budoucího systému zpracování dat tím, že bude možné v praxi otestovat, jaké ukazatele je vhodné do budoucna sledovat a bylo by tedy přínosné je implementovat do budoucí analytické architektury. Při přípravě těchto reportů je však třeba věnovat pozornost rozdílům ve struktuře a sémantice vstupních dat.

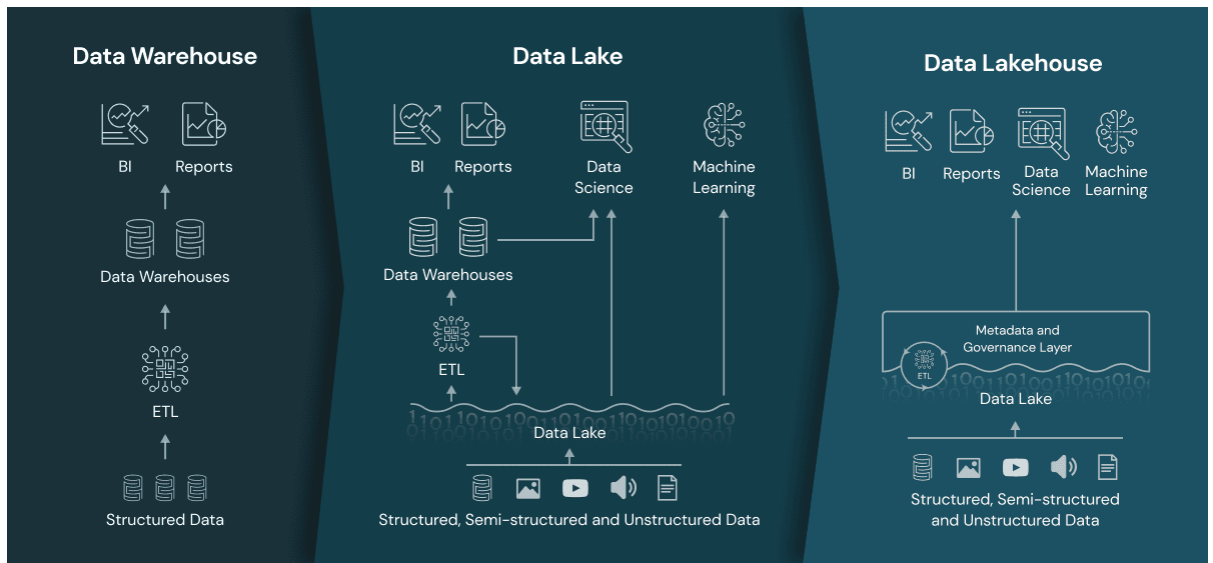
3.4 Shrnutí

V rámci této kapitoly byla navržena tři opatření, která cílila na zlepšení přístupu a nakládání s daty, která vznikají v rámci výrobního procesu. První kombinované řešení, které využívá Data Lake a Data Warehouse, považujeme za velmi robustní a spolehlivé. Vyznačuje se však nižší rychlostí práce s daty. Důvodem je přesun dat mezi dvěma typy úložišť. Data jsou ve všech formátech sbírána nejdříve do datového jezera, následně jsou přetransformována a přesunuta do datového skladu, kde se strukturalizují a jsou připravena na další použití například pro účely reportingu v Power BI.

Jako druhá alternativa bylo navrženo řešení využívající Data Lakehouse. Jedná se o modernější řešení, které kombinuje výhody datových jezer i skladů. Hlavní výhodou je absence nutnosti přesunu dat z jedné struktury do druhé, k veškerému zpracování dochází na jednom místě. Toto řešení je dále lépe škálovatelné, méně komplikované. Vzniká tu tedy menší prostor pro chyby, nebo duplicity. Jeho implementace je komplikovanější než v předchozím případě, přináší však řadu výhod.

Poslední z navržených řešení není považováno za jednu z alternativ. Jedná se spíše o první krok směrem k rozsáhlé datové analytice ve svařovně. Reporting pomocí Power BI by umožnil vizualizaci ukazatelů potřebných pro operativní rozhodování, nebo by dále pomohl odhalit kritická místa ve výrobním procesu. To vše by bylo možné dosáhnout za nízké náklady a téměř bez zásahu do současné datové infrastruktury.

Každé z řešení můžeme považovat za jednu z vývojových fází datové infrastruktury podniku. Výběr vhodného řešení by měl být zvážen na základě potřeb podniku.



Obrázek 8 Proces integrace dat (DataBricks,2020)

ZÁVĚR

Správné nakládání s podnikovými daty je v dnešní době klíčovým aspektem pro další rozvoj podniku a zajištění tržní konkurenceschopnosti. Cílem práce tedy bylo na základě teoretického vymezení problematiky Big Data provést analýzu současného stavu přístupu k Big Data v oddělení interní logistiky svařovny Škoda Auto a.s. v závodě v Kvasínách. Na základě výsledků této analýzy byly navrženy datové architektury vhodné pro budoucí využití v rámci podniku. Ve vztahu k dosažení cíle práce a zachování logické struktury, byla práce rozdělena do třech hlavních kapitol.

V rámci první kapitoly byla prvně vymezena problematika Big Data. Ta byla hned v úvodu této kapitoly charakterizována a rozdělena na základě typu datové struktury. Dále zde bylo zmíněno zabezpečení dat a způsob přistupování k Big Data. Na konec se kapitola věnovala Internet of Things a cloudovým úložištím.

Druhá kapitola se ve svém úvodu krátce věnovala historii a současnému vývoji Škoda Auto a.s. Hlavní náplní této analytické části práce byl však popis procesů v oddělení interní logistiky svařovny, které jsou klíčové pro vznik velkých datových objemů. Dále následoval rozhovor s experty, cílený na vyzdvižení slabých míst v kontextu nakládání s daty. Na základě výsledků této analýzy a provedeného rozhovoru jsme došli k závěru, že ač v podniku vznikají velké datové objemy, nemůžeme je však nazývat Big Data. Hlavním důvodem je absence hloubkové datové analýzy a příliš krátká doba archivace dat.

Ve třetí kapitole jsou proto navrženy dvě alternativní řešení datové architektury, z nichž každá má své výhody a nevýhody. Dále je v této kapitole navrženo řešení jednoduchého reportingu pomocí Power BI, které by sloužilo jako pilotní řešení pro budoucí softwarové změny, bez větších zásahů do současné datové architektury.

POUŽITÁ LITERATURA

- APACHE SOFTWARE FOUNDATION, 2018. **Apache Atlas – Data Governance and Metadata framework for Hadoop.** Online. Dostupné z: <https://atlas.apache.org/#/>. [cit. 2025-04-14].
- Badman, Annie a Kosinski Matthew, 2024. **What is Metadata?** Online. Dostupné z: <https://www.ibm.com/think/topics/metadata>. [cit. 2025-04-14].
- CISCO, 2014. **Not Big, Different: How Big Data Is Changing the Way We Do Business.** Online. Dostupné z: https://www.cisco.com/c/dam/en_us/about/ciscoitwork/enterprise-networks/docs/i-bd-04212014-not-just-big-different.pdf. [cit. 2024-11-24].
- CISCO, 2024. **Cities and Communities.** Online. Dostupné z: <https://www.cisco.com/c/en/us/solutions/industries/smart-connected-communities.html>. [cit. 2024-12-09].
- DATABRICKS, 2020. **What is a Data Lakehouse?** Online. Dostupné z: <https://www.databricks.com/blog/2020/01/30/what-is-a-data-lakehouse.html>. [cit. 2025-04-14].
- DATABRICKS, 2025. **ACID Transactions in Databases.** Online. Dostupné z: <https://www.databricks.com/glossary/acid-transactions>. [cit. 2025-04-10].
- DE FRANCISCI MORALES, Gianmarco; BIFET, Albert; KHAN, Latifur; GAMA, Joao a FAN, Wei. **IoT Big Data Stream Mining.** Online. *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. 2016, s. 2119-2120. ISBN 9781450342322. Dostupné z: <https://doi.org/10.1145/2939672.2945385>. [cit. 2024-12-09].
- DELL, 2024. **Co je systémová paměť (RAM).** Online. Dostupné z: <https://www.dell.com/support/kbdoc/cs-cz/000148441/co-je-systemova-pamet-ram>. [cit. 2024-11-25].
- DEPOTO, 2024. **Skladové čtečky (PDA / skenery / mobilní skladník).** Online. Dostupné z: <https://www.depoto.cz/cs/skladove-ctecky>. [cit. 2025-03-18]
- ESP HOLDING A.S., 2024. **Kanban – jak výroba tahem optimalizuje stav zásob a přispívá k efektivitě ve výrobě.** Online. Dostupné z: <https://esp.cz/cs/blog/kanban-vyroba-tahem-optimalizuje-stav-zasob-prispiva-efektivite-vyrobe>. [cit. 2025-04-09].
- GALATECH ENGINEERING, 2023. **Sensor Signal Warehouse.** Online. Dostupné z: <https://ssw.galatech-engineering.cz/>. [cit. 2025-04-30].

GARTNER, 2024. **Information Technology Glossary: Big Data**. Online. Dostupné z: <https://www.gartner.com/en/information-technology/glossary/big-data>. [cit. 2024-11-19].

GOOGLE CLOUD, 2024. **What is Cloud Computing?** Online. Dostupné z: <https://cloud.google.com/learn/what-is-cloud-computing>. [cit. 2024-12-05].

IBM Cloud Education Team, 2021. **ELT vs. ETL: What's the Difference?** Online. Dostupné z: <https://www.ibm.com/think/topics/elt-vs-etl>. [cit. 2025-04-09].

IBM, 2024. **Data Mining**. Online. Dostupné z: <https://www.ibm.com/topics/data-mining>. [cit. 2024-11-24].

IBM, 2024. **Machine Learning**. Online. Dostupné z: <https://www.ibm.com/topics/machine-learning>. [cit. 2024-11-24].

KOSINSKI, Matthew, 2024. **Data Warehouses vs. Data Lakes vs. Data Lakehouses**. Online. Dostupné z: <https://www.ibm.com/think/topics/data-warehouse-vs-data-lake-vs-data-lakehouse>. [cit. 2025-04-14].

MARTINEZ-MOSQUERA, Diana; NAVARRETE, Rosa a LUJAN-MORA, Sergio, 2020. **Modeling and Management Big Data in Databases—A Systematic Literature Review**. *Sustainability* Online. roč. 12, č. 2. ISSN 2071-1050. Dostupné z: <https://doi.org/10.3390/su12020634>. [cit. 2024-11-24].

MATTURDI, Bardi; ZHOU, Xianwei; LI, Shuai a LIN, Fuhong, 2014. **Big Data security and privacy: A review**. Online. *China Communications*., roč. 11, č. 14, s. 135-145. ISSN 1673-5447. Dostupné z: <https://doi.org/10.1109/CC.2014.7085614>. [cit. 2024-12-07].

MAYER-SCHÖNBERGER, Viktor a CUKIER, Kenneth. **Letting the data speak**. In: *Big Data: a revolution that will transform how we live, work, and think*. Boston: Houghton Mifflin Harcourt, 2014, s. 6. ISBN 978-0-544-00269-2. [cit. 2024-11-21].

MELL, Peter a GRANCE, Timothy, 2011. **The NIST Definition of Cloud Computing**. Online. *NIST Special Publication 800-145*, 2011, Dostupné z: <https://nvlpubs.nist.gov/nistpubs/legacy/sp/nistspecialpublication800-145.pdf>. [cit. 2024-12-05].

MICROSOFT AZURE, 2024. **Healthcare Solutions**. Online. Dostupné z: <https://azure.microsoft.com/en-us/solutions/healthcare/>. [cit. 2024-12-09].

MICROSOFT AZURE, 2024. **What is Big Data Analytics?** Online. Dostupné z: <https://azure.microsoft.com/cs-cz/resources/cloud-computing-dictionary/what-is-big-data-analytics>. [cit. 2024-11-24].

- MICROSOFT AZURE, 2024. **What is Cloud Computing?** Online. Dostupné z: <https://azure.microsoft.com/cs-cz/resources/cloud-computing-dictionary/what-is-cloud-computing/>. [cit. 2024-12-05]
- MICROSOFT AZURE, 2024. **What is NoSQL Database?** Online. Dostupné z: <https://azure.microsoft.com/cs-cz/resources/cloud-computing-dictionary/what-is-nosql-database/>. [cit. 2024-11-24].
- MICROSOFT AZURE, 2024. **What is SQL Database?** Online. Dostupné z: <https://azure.microsoft.com/cs-cz/resources/cloud-computing-dictionary/what-is-sql-database/>. [cit. 2024-11-24].
- MICROSOFT EDGE, 2025. **Generátor QR kódů.** Online. Dostupné z: <https://www.microsoft.com/cs-cz/edge/features/qr-code-generator>. [cit. 2025-03-22]
- MICROSOFT, 2023. **Use Microsoft Purview to govern Microsoft Fabric.** Online. Dostupné z: <https://learn.microsoft.com/en-us/fabric/governance/microsoft-purview-fabric>. [cit. 2025-04-14].
- MICROSOFT, 2024. **Co je zabezpečení cloudu?** Online. Dostupné z: <https://www.microsoft.com/cs-cz/security/business/security-101/what-is-cloud-security>. [cit. 2024-12-07].
- MICROSOFT, 2025. **Co je Microsoft Fabric?** Online. Dostupné z: <https://learn.microsoft.com/cs-cz/fabric/fundamentals/microsoft-fabric-overview>. [cit. 2025-04-10].
- MICROSOFT, 2025. **Power BI – vizualizace dat.** Online. Dostupné z: <https://www.microsoft.com/cs-cz/power-platform/products/power-bi>. [cit. 2025-04-22].
- MongoDB, 2025. **ACID Properties in DBMS Explained.** Online. Dostupné z: <https://www.mongodb.com/resources/basics/databases/acid-transactions>. [cit. 2025-04-10].
- MUCCI, Tim a STRYKER, Cole, 2024. **Big Data Analytics.** Online. Dostupné z: <https://www.ibm.com/topics/big-data-analytics>. [cit. 2024-11-24].
- OBALOVÉ MATERIÁLY, 2021. **KLT přepravy – k čemu jsou vhodné?** Online. Dostupné z: <https://www.obalove-materialy.cz/o-nas/clanky/klt-prepravky-k-cemu-jsou-vhodne>. [cit. 2025-03-17].
- PATGIRI, Ripon a AHMED, Arif. **Big Data: The V's of the Game Changer Paradigm.** 2016 *IEEE 18th International Conference on High Performance Computing and Communications; IEEE 14th International Conference on Smart City; IEEE 2nd International Conference on Data Science and Systems (HPCC/SmartCity/DSS)*, 2016, s. 17–24. ISBN 978-1-5090-4297-5. Dostupné z: <https://doi.org/10.1109/HPCC-SmartCity-DSS.2016.0014>. [cit. 2024-11-23].

RAVERA Sergio, 2023. **Big Data Characteristics (3V, 5V, 10V, 14V)**. Online. Dostupné z: <https://www.artera.net/en/data-science-en/big-data-characteristics-3v-5v-10v-14v/>. [cit. 2024-11-23].

SANDHU, Amanpreet Kaur. **Big Data with Cloud Computing: Discussions and Challenges**. *Big Data Mining and Analytics*, 2022, roč. 5, č. 1, s. 32–40. ISSN 2096-0654. Dostupné z: <https://doi.org/10.26599/Big DataMA.2021.9020016> [cit. 2024-11-21].

SAP, 2025. **Co je internet věcí?** Online. Dostupné z: <https://www.sap.com/cz/products/technology-platform/what-is-iot.html>. [cit. 2025-11-23].

SAP, 2025. **Co jsou big data?** Online. Dostupné z: <https://www.sap.com/cz/products/technology-platform/what-is-big-data.html>. [cit. 2025-04-03].

SEDGWICK, Niamh, 2024. **What Are Unstructured, Structured, and Semi-Structured Data Types** Online. Dostupné z: <https://www.matillion.com/blog/what-are-unstructured-structured-and-semi-structured-data-types>. [cit. 2024-11-24].

SIEMENS, 2024. **Digitální podnik** Online. Dostupné z: <https://www.siemens.com/cz/cs/reseni/digitalni-podnik.html>. [cit. 2024-12-09].

SÖHNER KUNSTSTOFFTECHNIK GMBH, 2025. **Skládací velkoobjemové boxy**. Online. Dostupné z: <https://www.soehner.de/cz/vyrobky/skladaci-velkoobjemove-boxy>. [cit. 2025-03-23].

ŠKODA AUTO, 2024. **Výroční zpráva 2024**. Online. Dostupné z: <https://reporting.skoda-auto.cz>. [cit. 2024-03-13].

ZEBRA TECHNOLOGIES, 2025. **TC72/TC77 Touch Computer Specification Sheet**. Online. Dostupné z: <https://www.zebra.com/us/en/products/spec-sheets/mobile-computers/handheld/tc72-tc77.html>. [cit. 2025-04-30].

SEZNAM TABULEK

Tabulka 1	Typy Metadata	16
Tabulka 2	Transakce ACID	40
Tabulka 3	Informační panely v Power BI dle uživatelských rolí.....	41

SEZNAM OBRÁZKŮ

Obrázek 1	Objem dat/informací vytvořených, zachycených, zkopírovaných a spotřebovaných celosvětově mezi roky 2010 a 2023, včetně prognózy pro období mezi roky 2024 a 2028 (v zettabajtech).....	12
Obrázek 2	Pět V v Big Data	13
Obrázek 3	Senzor SSW	24
Obrázek 4	Tlačítko pro vytvoření objednávky	25
Obrázek 5	Personal Digital Assistant	26
Obrázek 6	Kritické procesy pro vznik Big Data v oddělení interní logistiky svařovny.....	29
Obrázek 7	Datový tok kombinovaného řešení Data Lake a Data Warehouse.....	37
Obrázek 8	Proces integrace dat	43

SEZNAM ZKRATEK

AI	Artificial Intelligence Umělá Inteligence
AMU	Aplikace Mobilní Údržby
BI	Business intelligence Podniková Inteligence
CASB	Cloud Access Security Brokers Zprostředkovatel zabezpečení přístupu ke cloudu
CIEM	Cloud Infrastructure Entitlement Management Správa nároků cloudové infrastruktury
CNAPP	Cloud-Native Application Protection Platforms Platforma pro ochranu nativních cloudových aplikací
CSPM	Cloud Security Posture Management Správa stavu cloudového zabezpečení
CSV	Comma-separated values Hodnoty oddělené čárkou
CWPP	Cloud Workload Protection Platforms Platforma ochrany cloudové pracovní zátěže
DCR	Digital Control Room Digitální řídicí místnost
DL	Data Lake Datové jezero
DLH	Data Lakehouse Datové jezero/sklad
DW	Data Warehouse Datový sklad
GLT	Gitterbox Paletový box
GPS	Global Positioning System Globální polohový systém
IAM	Identity and Access Management

	Správa identity a přístupu
IoT	Internet of Things Internet věcí
iTLS	Internes Transport-LeitSystem Interní systém řízení dopravy
JSON	JavaScript Object Notation Zápis objektů v JavaScriptu
KLТ	Kleinladungsträger Malý přepravní box
ML	Machine learning Strojové učení
NoSQL	Není použito SQL
PDA	Personal Digital Assistant Osobní digitální asistent
QR	Quick Response Rychlá odpověď (QR kód)
RAM	Random Access Memory Operační paměť počítače
SAP	Systemanalyse Programmentwicklung Systémová analýza vývoj programu
SQL	Structured Query Language Strukturovaný dotazovací jazyk
SSW	Smart Sensor Warehouse Sklad s chytrými senzory
ŠA	Škoda Auto a.s.
VM	Virtual Machine Virtuální stroj
WTM	Work Test Management Řízení pracovních testů
XML	eXtensible Markup Language Rozšiřitelný značkovací jazyk
ZB	Zettabaj