

IT a anatomie firmy

(Microsoft Fabric)

(pracovní dokument)



Lukáš Kuthan

VŠE Praha, 2024

Obsah

1.	Úvod	3
2.	Microsoft Fabric	4
2.1	Unifikace	4
2.1.1	OneLake	4
2.1.2	Data Governance	6
2.1.3	Integrace umělé inteligence.....	6
2.2	Komponenty Microsoft Fabric	6
2.2.1	Data Factory	6
2.2.2	Synapse Data Engineering	7
2.2.3	Synapse Data Science.....	7
2.2.4	Synapse Data Warehousing.....	7
2.2.5	Synapse Real-Time Analytics	7
2.2.6	Power BI	7
2.2.7	Data Activator	7
2.3	Opodstatnění použití Microsoft Fabric	7
3.	Zdroje	9

1. Úvod

Společnost Microsoft v loňském roce (2023) představila unikátní platformu **Microsoft Fabric sloužící pro end-to-end datově analytická řešení**. Microsoft uvedením této platformy reaguje na stále zvyšující se nároky na komplexitu a různorodost v datové sféře.

U většiny firem v kontextu Business Intelligence zůstává základním kamenem klasický přístup **budování datových skladů spolu s reportingovým řešením**, který je vhodný pro práci **se strukturovanými daty** a pomáhá tradičně businessu při plánovacích, rozhodovacích a optimalizačních činnostech. Avšak jedním z hlavních trendů poslední doby je například **analýza nestrukturovaných dat**, která vyžaduje jiné přístupy už k samotnému ukládání a celkovému zpracování pro smysluplné výstupy (Antunes et al. 2022).

Podíváme-li se na problematiku dat optikou úrovní Business analytiky, zjistíme, že analytika může být buď deskriptivní, diagnostická, prediktivní anebo preskriptivní. Každá úroveň, stejně jako forma dat, vyžaduje jiné přístupy, znalosti a technologie pro její ovládnutí. Například **pokročilé úrovně analytiky** typicky vyžadují datovou architekturu a technologie, které zvládnou ukládat a zpracovávat velké množství dat různé formy a umožní aplikovat **algoritmy strojového učení a umělé inteligence**. Všechny úrovně analytiky a s nimi spojené přístupy mají však z pohledu firem jeden společný cíl – ovlivňovat pozitivním způsobem kvalitu a výkonnost firem (Mallam et al. 2021).

Výše zmíněné situace často vyžadují **vícero nástrojů a technologií** od různých dodavatelů, což má za důsledek například duplicitní uložení těch samých dat, rozdílné licenční podmínky, odlišné ekosystémy, a především **vysoké nároky** (znalostní, časové, finanční) **na samotnou integraci** všech komponent. Tuto situaci se snaží řešit právě **Microsoft Fabric** skrze unifikovanou analytickou platformu pokrývající **veškeré datové potřeby firem od úvodní integrace a uložení dat až po samotnou prezentační vrstvu** spolu s celkovou správou. Platforma nebyla vytvořena na zelené louce, ale tvoří ji například již známé Microsoft produkty pro práci s daty jako **Azure Data Factory, Azure Synapse Analytics a Power BI**. Hlavní výhody této platformy představují široké low-code i no-code možnosti integrace rozmanitých datových zdrojů, podpora pokročilé analytiky, škálovatelnost a flexibilita pro aktuální potřebu, podpora umělé inteligence a celková unifikace všech komponent do jednoho prostředí (Anon. 2023).

2. Microsoft Fabric

V květnu minulého roku (2023) společnost Microsoft představila svou novou unifikovanou analytickou platformu a spustila s tímto představením její fázi „public preview“. Od této chvíle se Microsoft Fabric dostal ke zkoumání širší veřejnosti a **zaujal především svou komplexností**, která je dostupná pouze v rámci jedné jediné platformy. V kontextu různorodých rolí a přístupů, které se podílejí na budování datové analytiky je ve většině případů zapotřebí vícero produktů a služeb a mnohdy i od jiných dodavatelů, aby došlo ke kompletnímu pokrytí všech potřeb. Na tuto situaci se Microsoft Fabric snaží reagovat **integrací několika komponent do jedné platformy**, která dokáže pokrýt potřeby všech datových rolí od datového inženýra a vědce až po datového / business analytika spolu se samotnými business uživateli a konzumenty datové analytických výstupů (Ghoshal 2023).

Microsoft Fabric je tedy **tvořen jednotlivými komponentami**, které mají vždy specifické zaměření a účel. Jinými slovy **každá komponenta** nebo sada komponent je **určena pro jinou úlohu** při práci s daty, respektive pro jinou roli datového charakteru. Podíváme-li se však na názvy jednotlivých komponent, v ideálním případě očima osoby, která do dnešního dne využívala Microsoft datové nástroje a služby, zjistíme že jsou velmi podobné. Microsoft Fabric **nepředstavuje sadu zbrusu nových komponent**, ale z velké části **integruje a unifikuje již dobře známé produkty** z datové rodiny Azure a Power Platform jako je Azure Data Factory, Azure Synapse Analytics nebo Power BI (Ghoshal 2023).

V základním pojetí je Microsoft Fabric **cloudová sada SaaS** nástrojů, která umožňuje zpracovávat velké množství různorodých dat s vysokou mírou škálovatelnosti. Celá platforma má za cíl pokrýt **požadavky jak vývojářských rolí, tak i ty ze strany business uživatelů**. Microsoft Fabric integruje klíčové nástroje a produkty z rodiny Azure a Power Platform, spolu s jejich optimalizací, aby tvořily unifikovanou platformu a jedno místo pravdy pro zpracování, ukládání a prezentování podnikových dat (Bisson 2023).

2.1 Unifikace

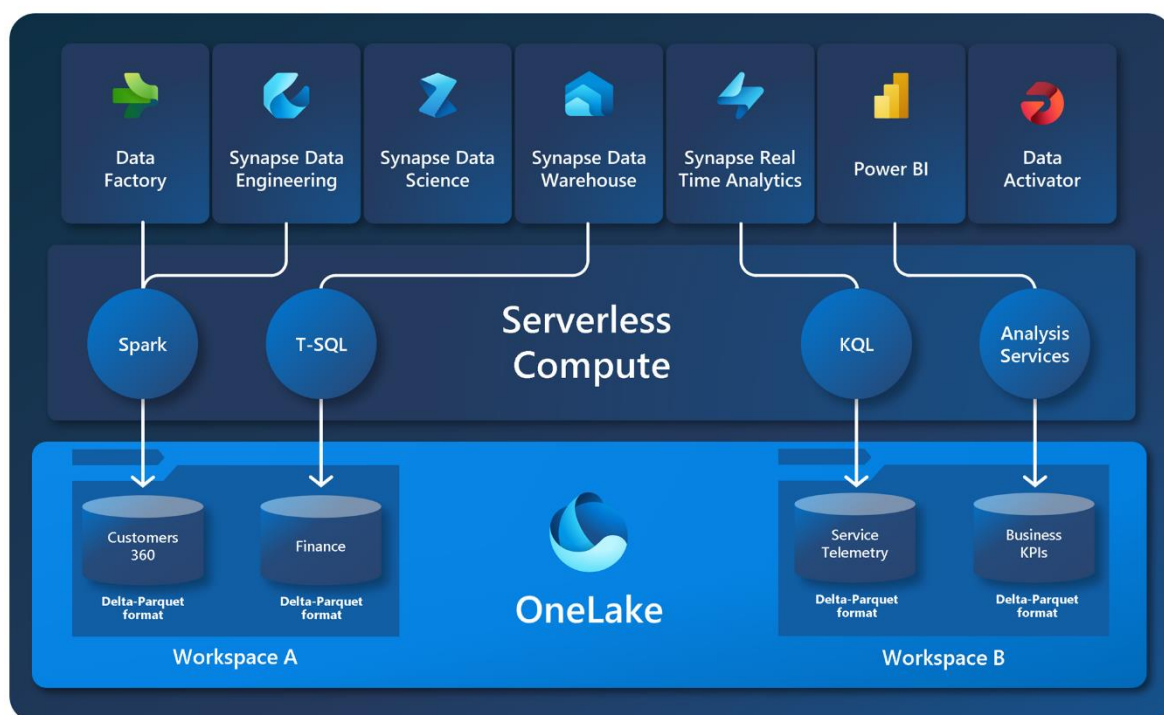
Unifikace a jednotnost nově vzniklé platformy by se dala popsat z vícero úhlů pohledu. Nalezneme jak technologické aspekty, tak i ty, které souvisí již s výše zmiňovanou komplexitou pro různé role a scénáře. K těmto aspektům by se určitě měla i řadit **oblast Data Governance**, která se podílí na logice vybudování a užívání platformy Microsoft Fabric (Bisson 2023).

2.1.1 OneLake

Microsoft Fabric by měl být standardně implementován **na úrovni jednoho Microsoft tenantu**, který typicky představuje jednu firmu či společnost. **Jednomu tenantu** pak přísluší vždy právě **jedno centrální úložiště všech dat – OneLake**. Jedná se o úložiště **typu data lake**, které přijímá různorodé formáty dat a často bývá označován jako „OneDrive pro data“. OneLake se naskytuje jako elegantní řešení v situacích, kdy firmy a společnosti typicky disponují několika datovými silami. Ty si opět můžeme představit jako data lake úložiště či relační databáze, které vznikají například na úrovni jednotlivých oddělení. Ve výsledku vzniká situace, která může být poměrně **náročná na správu a řízení**. Tato náročnost se stává realitou v situacích, kdy jsou přítomny různé nástroje, technologie a přístupy, které jednotlivá úložiště aplikují. S tím často mohou souviset i **specifické požadavky na sdílení těchto dat** mezi jednotlivými odděleními (úložišti) nebo požadavky na celkové zabezpečení a správu dat. Při samotném sdílení dat dochází ke kopírování dat a ta samá data se nám ukládají na vícero místech na jednu a zvyšují nároky na datovou infrastrukturu a potřeby datových kontrol (Bisson 2023).

OneLake se snaží reagovat na výše zmíněné situace tím, že **v kontextu jednoho tenantu (firmy) existuje striktně právě jednou a vzniká tak centrální úložiště veškerých dat**. Důležité je však zmínit, že následná práce s daty už může být distribuována **do jednotlivých domén** a firmy tak mohou jít naproti **konceptu doménového vlastnictví dat**. Ba naopak tento scénář je optimální v kontextu Microsoft Fabric, který přímo **nabízí doménové uspořádání architektury**. K tomuto účelu slouží takzvané **pracovní prostory**, které shlukují datové artefakty například podle jejich nadřazené domény či obecné funkcionality. Pracovní prostory poté poskytují možnost vlastní správy veškerého obsahu a řízení přístupu k němu. Aby se však data od svého úvodního načtení dostala do požadované lokace a podoby, **Microsoft Fabric podporuje čtyři typy enginů**, které umožňují s daty pracovat. Dle Obrázek 1 se jedná o **T-SQL, Spark, Analysis Services a KQL**. Jejich použití závisí na požadovaném výstupu

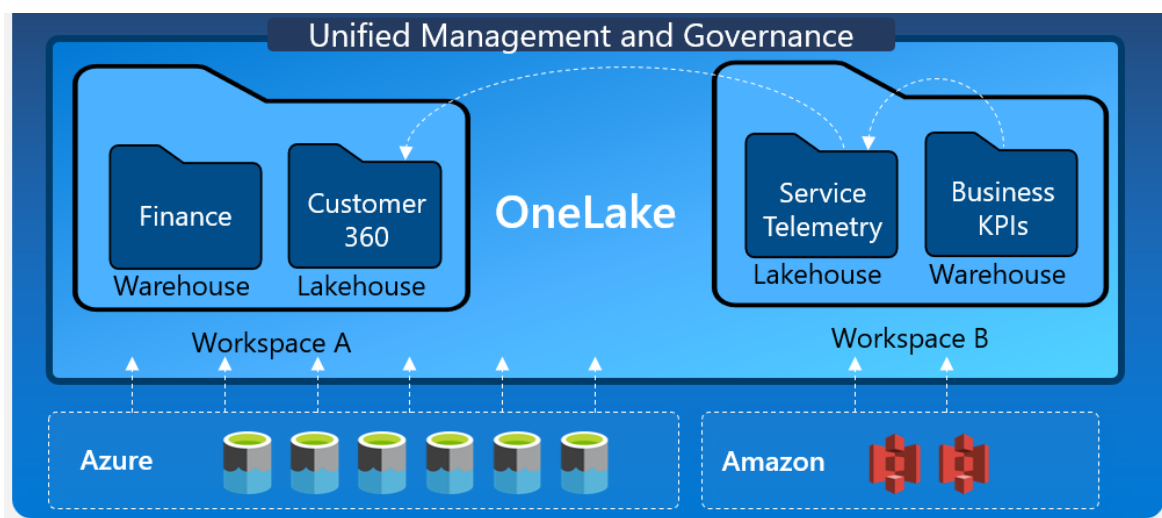
či roli, která s daty pracuje. Těmito enginy jsou pokryty nejenom potřeby pro klasickou analýzu dat pomocí jazyka SQL s výstupem v Power BI, ale scénáře v oblastech strojového učení nebo real-time analytiky, které generují specifické požadavky na výkon a možnost práce s velkými objemy dat (Microsoft 2023).



Obrázek 1: Operace s daty v Microsoft Fabric pomocí různých enginů. Zdroj: (Microsoft 2023)

Zmiňované enginy podporují práci s **Delta Parquet formátem**, který je zde nativním standardem pro uložení a práci s daty. Jinými slovy, pokud transformujeme data uvnitř platformy Microsoft Fabric, dřív nebo později se nám do tohoto formátu data dostanou. Typicky se tomu tak bude dít například při transformaci dat z originálních formátů jako jsou například csv nebo json soubory do optimální tabulkové struktury právě typu Delta **v kontextu Lakehouse a Warehouse** uložit. Delta Parquet nebo taky někdy Delta Lake představuje formát, který je tvořen parquet soubory s tím, že je k nim **přidána i transakční historie, důležitá metadata** (například o schématu dat) a **optimalizační mechanismus**, díky kterému dochází k významné komprimaci dat (Armbrust et al. 2020).

Dalším klíčovým konceptem platformy Microsoft Fabric jsou takzvané „Shortcuts“. **Shortcuts** by se daly doslovně přeložit jako **zkratky**, ale výstižnější by bylo použít spíše slovo **odkaz nebo pointer**. Shortcuts řeší problém několikanásobného kopírování a ukládání těch samých dat. Dle níže přiloženého obrázku (Obrázek 2) můžeme vidět jejich princip fungování. V zásadě nám umožňují získat přístup k datům bez nutnosti je přesouvat nebo kopírovat, což **šetří úložiště a redukuje počet operací** při manipulaci s daty. Shortcuts umožňují odkazovat se na data uvnitř prostředí Microsoft Fabric i mimo něj. První varianta tedy například umožňuje sdílení dat mezi doménami a jejich datovými sklady neboli Warehousey (podle originálního názvosloví Microsoft Fabric). Druhá varianta umožňuje získávat data z externího prostředí, a to konkrétně z úložišť Azure Data Lake Storage Gen2 a Amazon S3 (Bisson 2023).



Obrázek 2: Shortcuts v prostředí Microsoft Fabric. Zdroj: (Witcombe 2023)

2.1.2 Data Governance

Z dosavadních zjištění a charakteristik platformy Microsoft Fabric vyplývá, že **pro oblast Data Governance jsou vytvořeny velmi solidní podmínky**. Základem je **centrální úložiště dat**, které umožňuje jeho globální správu, jednotné technologické přístupy a snazší kontrolu oproti decentralizovaným celkům s rozdílnými specifiky. Centrální úložiště však reálně figuruje spíše na pozadí, protože vzniklé datové artefakty spadají do pracovních prostorů, se kterými přicházejí do styku jak vývojáři, tak konzumenti obsahu. **Pracovní prostory disponují vlastní správou a nastavením**. Hierarchicky je důležité zmínit, že Microsoft Fabric zavádí i **domény neboli Domains**, které mohou shlukovat vícero pracovních prostorů a taktéž nabízí možnost správy a vlastního řízení. Z pohledu Data Governance tak vzniká **víceúrovňová správa**, která na té nejvyšší úrovni definuje naprostý základ, ale například vlastnictví dat je delegováno jednotlivým doménám. Oblast kvality dat nám podporují výše definované Shortcuts, které redukují počet potřebných kopií a extrakčních operací nad daty. Tím se **snižuje prostor pro vznik chyb** vedoucích k nekvalitě a zároveň eliminujeme situace, kdy vícero domén / oddělení potřebuje část stejných dat, které se tak musí několikrát namnožit a na výstupu už nemusí být vůbec stejná a validní dle původního zdroje poskytnutí, kvůli jiným transformacím a přístupům. Pro plnohodnotné zavedení Data Governance je vhodné využít i **dedikovaný produkt Microsoft Pureview** pro tuto oblast (Microsoft 2024).

2.1.3 Integrace umělé inteligence

Microsoft Fabric aspiruje na globální integraci umělé inteligence do valné většiny svých komponent. Skrze zabudování služeb **Azure OpenAI a Copilot** cílí na širší a rychlejší adopci celé platformy, kde bude například možné na základě přirozeného jazyka **vygenerovat potřebný kód pro transformaci dat** či celý ETL mechanismus a nebo vytvořit **kompletní Power BI report** (Ghoshal 2023).

2.2 Komponenty Microsoft Fabric

Microsoft Fabric nabízí několik komponent / oblastí, které **slouží specifickým úlohám či datovým rolím**. Každá komponenta pak nabízí možnost vytvářet různé artefakty a používat jiné technologie pro co možná nejlepší pokrytí daných potřeb.

2.2.1 Data Factory

První komponentou je Data Factory, která nabízí více než **150 konektorů pro extrakci dat** z různorodých datových zdrojů. Primárním účelem je pak orchestrace a tvorba datových pipeline spolu s mechanismy pro error handling. Data Factory umožňuje aplikovat i jednodušší **low-code transformační kroky** a integrovat DataFlow gen2 pro Power Query transformace či Notebooks využívající Apache Spark engine (Berge 2023).

2.2.2 Synapse Data Engineering

V této komponentě je umožněno provádět složitější a více rozmanité transformace za použití **Spark Jobs a Notebooků, které využívají pro práci s daty například rozhraní Pythonu pro Apache Spark – PySpark**. Klíčovým artefaktem pro uložení strukturovaných a nestrukturovaných dat je **Lakehouse**, který kombinuje principy data lake a datového skladu. Zde je možné ukládat data jako jednotlivé soubory a převádět je poté do tabulkové struktury a formátu **Delta Lake**. **Lakehouse** také umožňuje využití SQL endpointu pro read-only operace spolu s možností vytváření view objektů (Berge 2023).

2.2.3 Synapse Data Science

Tato komponenta utváří prostředí pro vývoj, testování a **nasazování modelů strojového učení**. Umožňuje bezproblémovou integraci populárních knihoven pro strojové učení a s využitím engine **Apache Spark** je optimalizován pro práci s velkými objemy dat. Nejčastěji se pro vývoj modelů využívá již zmíněný PySpark anebo SparkR (Berge 2023).

2.2.4 Synapse Data Warehousing

V této komponentě nacházíme plnohodnotné řešení na bázi datového skladu. Jedná se zde především o možnost podporující **DDL a DML dotazy jazyka SQL** nad klíčovým artefaktem této komponenty – Warehouse, což není možné realizovat v Lakehouse artefaktu z komponenty **Synapse Data Engineering**. Klíčovým jazykem pro zpracování a manipulaci s daty je **T-SQL** a Warehouse typicky umožňuje ukládat již a pouze plně strukturovaná data (Berge 2023).

2.2.5 Synapse Real-Time Analytics

Synapse Real-Time Analytics řeší v Microsoft Fabricu **analýzu a zpracování streamovaných dat** s poskytnutím vhodných mechanismů pro optimalizaci a škálování těchto úloh. Klasickým scénářem zde může být zpracování dat, která jsou vysokofrekvenčně **generována z IoT** čidel. Zásadním nástrojem je zde **jazyk KQL** uzpůsobený pro **dotazování se na streamovaná data**.

2.2.6 Power BI

Power BI představuje velmi **komplexní BI nástroj**, který v kontextu Microsoft Fabric řešení figuruje primárně jako **vizualizační prvek**. Nicméně Power BI ve své desktopové verzi **umožňuje pokrývat ETL** scénáře pomocí **Power Query** a následně tvořit vybrané metriky a analyzovat data pomocí jazyka DAX. V Power BI je možné vytvořit nový datový model nebo se připojit na již vytvořený. S příchodem Microsoft Fabric se rozšířila nabídka možností pro připojení na data o **Direct Lake mód**, který je ve své podstatě přímo napojen na datové struktury platformy Microsoft Fabric a dokáže data přenášet do vizualizační části ve skoro reálném čase. Na pozadí funguje optimalizační mechanismus, který loguje často dotazovaná data a načítá je rychleji a rychleji. Power BI existuje jak v cloudové verzi, kde je možno výsledný obsah sdílet a efektivně spravovat, tak i v desktopové, která disponuje širší funkcionalitou a slouží primárně pro hlavní vývoj reportů (Berge 2023).

2.2.7 Data Activator

Poslední komponentou je Data Activator, který slouží pro **automatizované úkony**, které se mají provést při splnění definovaných podmínek. Jedná se o **low-code nástroj**, který například umožňuje kontrolovat uživatelem určená data a při splnění jistých kritérií posílat **oznámení do Microsoft Teams** nebo Outlooku. Komponenta také nabízí **integraci Power Automate** pro vytváření vlastních mechanismů podmínek a následných úkonů (Berge 2023).

2.3 Opodstatnění použití Microsoft Fabric

Microsoft Fabric se využívá jako platforma pro vytvoření datové architektury a reportingových řešení z několika hlavních důvodů:

- Dosavadní architektura a reportingová řešení využívají produkty z rodiny Microsoft Azure a Power Platform, což znamená jednodušší integraci a implementaci do prostředí Microsoft Fabric.
- Jednoduchá škálovatelnost je nespornou výhodou této platformy. Bezproblémová škálovatelnost platformy eliminuje případné starosti s konfigurací potřebné infrastruktury.
- Jednotná Data Governance nad celým datovým spektrem je dalším validním důvodem pro volbu této platformy. Redukuje se tak počet platformů či systémů, které je třeba spravovat a efektivně řídit.
- Budování best-practise přístupů v ekosystému Microsoft Fabric představuje další důvod, který má za cíl prakticky zdokonalit firemní znalosti a zkušenosti s touto platformou. Tento fakt povede k možnosti nabízet kvalitnější, osvědčená a stabilní řešení. S tímto bodem souvisí i udržení kroku s nekonečným rozvojem informačních technologií na poli datových nástrojů a produktů.
- Posledním bodem je soulad s doménovým pohledem na vlastnictví a správu dat, kdy firma bude těžit z definované struktury domén uvnitř platformy Microsoft Fabric a bude tak možné reflektovat v této platformě odraz reality organizační struktury firmy.

3. Zdroje

- Anon., 2023. Build 2023: Microsoft launches Fabric, a new data and analytics platform [GADGETS NEWS]. *The Times of India* [online]. [vid. 2024-02-10]. Dostupné z: <https://www.proquest.com/docview/2817793506/citation/B9343615FA2D4168PQ/1>
- ANTUNES, António Lorvão, Elsa CARDOSO a José BARATEIRO, 2022. Incorporation of Ontologies in Data Warehouse/Business Intelligence Systems – A Systematic Literature Review. *International Journal of Information Management Data Insights* [online]. 2(2), 100131. ISSN 2667-0968. Dostupné z: doi:10.1016/j.jjimei.2022.100131
- ARMBRUST, Michael, Tathagata DAS, Liwen SUN, Burak YAVUZ, Shixiong ZHU, Mukul MURTHY, Joseph TORRES, Herman VAN HOVELL, Adrian IONESCU, Alicja ŁUSZCZAK, Michał ŚWITAKOWSKI, Michał SZAFRAŃSKI, Xiao LI, Takuya UESHIN, Mostafa MOKHTAR, Peter BONCZ, Ali GHODSI, Sameer PARANJPYE, Pieter SENSTER, Reynold XIN a Matei ZAHARIA, 2020. Delta lake: high-performance ACID table storage over cloud object stores. *Proceedings of the VLDB Endowment* [online]. 13(12), 3411–3424. ISSN 2150-8097. Dostupné z: doi:10.14778/3415478.3415560
- BERGE, Eirik, 2023. An Overview of Microsoft Fabric Going Into 2024. *Medium* [online] [vid. 2024-03-24]. Dostupné z: <https://towardsdatascience.com/an-overview-of-microsoft-fabric-going-into-2024-cd51a18c07b5>
- BISSON, Simon, 2023. Understanding OneLake and lakehouses in Microsoft Fabric. *InfoWorld.com* [online]. [vid. 2024-03-24]. Dostupné z: <https://www.proquest.com/docview/2848449805/abstract/6DCAD662D62A4DA1PQ/1?sourcetype=Trade%20Journals>
- BUTTE, Vijay Kumar a Sujata BUTTE, 2022. Enterprise Data Strategy: A Decentralized Data Mesh Approach. In: *2022 International Conference on Data Analytics for Business and Industry (ICDABI): 2022 International Conference on Data Analytics for Business and Industry (ICDABI)* [online]. s. 62–66 [vid. 2024-03-26]. Dostupné z: doi:10.1109/ICDABI56818.2022.10041672
- CIRKOVSKÝ, 2018. Co je to strategické řízení nákupu? – I. část. *Benefico* [online] [vid. 2024-02-24]. Dostupné z: <https://benefico.cz/co-je-to-strategicke-rizeni-nakupu-i/>
- GARTNER a GUPTA, 2021. Using Data Fabric Architecture to Modernize Data Integration. *Gartner* [online] [vid. 2024-03-26]. Dostupné z: <https://www.gartner.com/smarterwithgartner/data-fabric-architecture-is-key-to-modernizing-data-management-and-integration>
- GHOSHAL, Anirban, 2023. Why Microsoft is combining all its data analytics products into Fabric. *InfoWorld.com* [online]. Dostupné z: <http://ezproxy.techlib.cz/login?url=https://www.proquest.com/trade-journals/why-microsoft-is-combining-all-data-analytics/docview/2817696905/se-2?accountid=119841>
- HAELEN, Bennie a Dan DAVIS, 2023. *Delta Lake: Up and Running*. B.m.: O'Reilly Media, Inc. ISBN 978-1-09-813968-1.
- HALKJÆR, Mathias, 2023. *Microsoft Fabric deep dive – Data Factory* [online]. 2023. Dostupné z: <https://www.fellowmind.com/globalassets/denmark/downloadable-content/microsoft-fabric-deep-dive---data-factory-v.2.pdf>
- HARIHARAN, 2023. *Adopt Data Mesh in Microsoft Fabric – Hari BI Academy* [online] [vid. 2024-03-29]. Dostupné z: <https://haribiacademy.com/2023/07/adopt-data-mesh-in-microsoft-fabric/>
- HROUDA, Adam, 2023. Import SharePoint data to Fabric. *Adam's Bubble* [online] [vid. 2024-04-08]. Dostupné z: <https://adamsbubble.com//posts/import-sharepoint-data-to-fabric>
- HRŮZA, Tomáš, 2007. *Jak poznat vyspělé procesy?* [online] [vid. 2024-02-25]. Dostupné z: <https://www.systemonline.cz/sprava-it/cmmi-model-hodnoceni-vyspelosti-procesu-1.htm>
- ILIJASON, Robert, 2020. *Beginning Apache Spark Using Azure Databricks: Unleashing Large Cluster Analytics in the Cloud*. B.m.: Apress. ISBN 978-1-4842-5781-4.
- JOHNSON, Aimee, 2023. Data Lakehouse vs Data Warehouse in Microsoft Fabric. *Advancing Analytics* [online] [vid. 2024-04-06]. Dostupné z: <https://www.advancinganalytics.co.uk/blog/2023/6/13/data-lakehouse-vs-data-warehouse-in-microsoft-fabric>
- KLOPFENSTEIN, 2024. *Data Factory in Microsoft Fabric documentation – Microsoft Fabric* [online] [vid. 2024-04-06]. Dostupné z: <https://learn.microsoft.com/en-us/fabric/data-factory/>

- LAPLANTE, A. a an O'Reilly Media Company SAFARI, 2020. *The Modern Cloud Data Platform* [online]. B.m.: O'Reilly Media, Incorporated. Dostupné z: <https://books.google.cz/books?id=0D1GzgEA-CAAJ>
- L'ESTEVE, Ron, 2022. *The Azure Data Lakehouse Toolkit: Building and Scaling Data Lakehouses on Azure with Delta Lake, Apache Spark, Databricks, Synapse Analytics, and Snowflake* [online]. Berkeley, CA: Apress [vid. 2024-04-02]. ISBN 978-1-4842-8232-8. Dostupné z: doi:10.1007/978-1-4842-8233-5
- MALLAM, Pooja, Ashu ASHU a Baljeet SINGH, 2021. Business Intelligence Techniques Using Data Analytics: An Overview. In: *2021 International Conference on Computing Sciences (ICCS): 2021 International Conference on Computing Sciences (ICCS)* [online]. Phagwara, India: IEEE, s. 265–267 [vid. 2024-02-10]. ISBN 978-1-66549-445-8. Dostupné z: doi:10.1109/ICCS54944.2021.00059
- MAZUMDAR, Dipankar, Jason HUGHES a J. B. ONOFRE, 2023. *The Data Lakehouse: Data Warehousing and More* [online]. 12. říjen 2023. B.m.: arXiv. [vid. 2023-11-25]. Dostupné z: <http://arxiv.org/abs/2310.08697>
- MICROSOFT, 2023. *What is OneLake? - Microsoft Fabric* [online] [vid. 2024-03-24]. Dostupné z: <https://learn.microsoft.com/en-us/fabric/onelake/onelake-overview>
- MICROSOFT, 2024. *Governance and compliance in Microsoft Fabric – Microsoft Fabric* [online] [vid. 2024-03-24]. Dostupné z: <https://learn.microsoft.com/en-us/fabric/governance/governance-compliance-overview>
- REŠL, Štěpán, 2023. *Copy Activity, Dataflows Gen2, and Notebooks vs. SharePoint Lists – Data Meerkat* [online] [vid. 2024-04-04]. Dostupné z: <http://localhost:4000/copy-activity-dataflows-gen2-and-notebooks-vs-sharepoint-lists>
- RUSSO, Marco, 2024. *Direct Lake vs. Import mode in Power BI – SQLBI* [online] [vid. 2024-04-20]. Dostupné z: <https://www.sqlbi.com/blog/marco/2024/04/06/direct-lake-vs-import-mode-in-power-bi/>
- SPARK DOCUMENTATION, 2024. *Cluster Mode Overview – Spark 3.5.1 Documentation* [online] [vid. 2024-04-02]. Dostupné z: <https://spark.apache.org/docs/latest/cluster-overview.html>
- STRAKOVÁ, Jarmila a Jan VÁCHAL, 2018. Situační Analýza, Strategické Řízení a Rozhodování Podniků. *Mladá Věda*. 6(2), 205–215.
- ŠVIRK, Martin, 2017. *Strategie podniku* [online]. Jindřichův Hradec. Diplomová práce. Vysoká škola ekonomická v Praze. Dostupné z: https://vskp.vse.cz/69512_strategie-podniku?title=Strategie+podniku&page=1
- WITCOMBE, Jordan, 2023. Microsoft Fabric: An introduction to OneLake. *Advancing Analytics* [online] [vid. 2024-03-24]. Dostupné z: <https://www.advancinganalytics.co.uk/blog/2023/5/24/microsoft-fabric-an-introduction-to-onelake>