

Výběr a vyhodnocení SW nástroje a jeho funkcionalit

metodický dokument

autoři: Ing. Jiří Marek, Ing. Martina Hábová, Ph.D.

verze: 1.0 final

datum: 09.10.2020



Evropská unie
Evropský sociální fond
Operační program Zaměstnanost



Dokument je vytvořen v rámci projektu
„Optimalizace postupů při integraci prvků elektronického zdravotnictví do procesů ÚZIS a
tvorba dokumentace k informačnímu systému vzácných onemocněních v ČR“
registrační číslo CZ.03.4.74/0.0/0.0/15_025/0015811



Obsah

1	Úvod.....	4
1.1	Popis aktivity	4
1.2	Základní požadavky na navržené řešení.....	5
1.2.1	Logické součásti systému.....	5
1.2.2	Základní požadavky na navržené řešení.....	6
2	Průzkum trhu.....	8
2.1	Vybrání dodavatelé.....	8
2.2	Způsob hodnocení.....	8
2.3	InterSystems	9
2.3.1	Hodnocení	10
2.4	CloverDX.....	19
2.4.1	Hodnocení	20
2.5	Teradata Big Data Analytics.....	21
2.5.1	Hodnocení	22
2.6	Microsoft	30
2.6.1	Hodnocení	34
2.7	Informatica.....	43
2.7.1	Hodnocení	44
2.8	SAS.....	45
2.8.1	Hodnocení	46
2.9	ORACLE	54
2.9.1	Hodnocení	54
2.10	IBM.....	63
2.10.1	Hodnocení	63
3	Vyhodnocení průzkumu trhu	65
3.1	Odůvodnění řešení MICROSOFT (Microsoft SQL Server Enterprise):.....	66
3.2	Popis řešení / komponent.....	67
4	Dodatečný nástroj	68
4.1	Popis – HP Vertica	69
5	Závěr	70
5.1	Průzkum trhu	70
5.2	Dodatečný nástroj.....	71



Zkratky

ISVO	Informační systém vzácných onemocnění
ÚZIS ČR	Ústav zdravotnických informací a statistiky ČR
NZIS	Národní zdravotní informační systém
DG	Diagnóza z definice MKN10
NRHZS	Národní registr hrazených zdravotních služeb
DI	Datová integrace
DS	Datový sklad
BI	Business Intelligence
KDD	Knowledge Data Discovery
DM	Datamart, Datamodeling
QA	Quality Assurance
DG	Data Governance



1 Úvod

Tento dokument popisuje proces „Výběru SW nástrojů“ v rámci projektu „Optimalizace postupů při integraci prvků elektronického zdravotnictví do procesů ÚZIS a tvorba dokumentace k informačnímu systému vzácných onemocnění v ČR“, registrační číslo CZ.03.4.74/0.0/0.0/15_025/0015811, a jeho klíčové aktivity č. 3 Pořízení, implementace a pilotní ověření využití softwarových nástrojů pro analýzy dat státní správy.

1.1 Popis aktivity

Projekt: *Optimalizace postupů při integraci prvků elektronického zdravotnictví do procesů ÚZIS a tvorba dokumentace k informačnímu systému vzácných onemocnění v ČR eHealth a vzácná onemocnění.*

KA03 – Pořízení, implementace a pilotní ověření využití softwarových nástrojů pro analýzy dat státní správy

Cílem aktivity je, vzhledem ke vzrůstající poptávce po rozsáhlých analytických hodnoceních dat z nových zdravotních registrů NZIS (zejména NRHZS), pořízení softwarového nástroje, sloužícího k efektivnímu zpracování a vytěžování datových zdrojů ÚZIS ČR, k jejich propojování a provádění transformací nad daty tak, aby byla zajištěna požadovaná validita a kvalita dat. Klíčovou vlastností je také rychlost zpracování (půjde o big data v rozsahu jednotek TB)

Komplexní jednotné technologické řešení bude podporovat následující funkcionality:

- Nástroj pro zpracování a datovou integraci vstupních dat
- Databázi pro uložení výsledků z předchozího kroku, jejich správu a zpřístupnění (rozšíření stávající licence)
- Nástroj pro zajištění bezpečnosti dat

Detailní požadavky na nástroje a funkcionality uvedené výše budou nastaveny v rámci průzkumu trhu a bude rozhodnuto, zda budou nástroje soutěženy jako celek, nebo se bude jednat o více samostatných výběrových řízení pokrývajících jednotlivé nástroje a logické celky. Dodavatel bude vybrán ve výběrovém řízení.

Zdůvodnění výběru SW nástrojů.

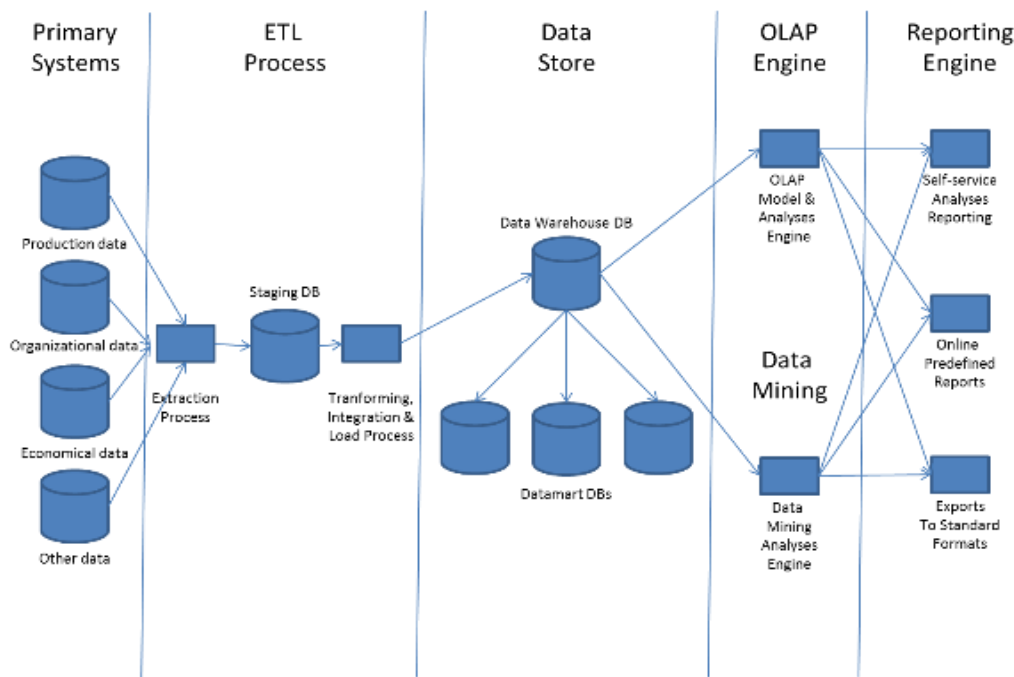
Pro naplnění výše uvedené klíčové aktivity byl proveden průzkum trhu na základě top produktů v oblastech:

- Datová integraci (DI)
- Datový sklad (DS)
- Business Intelligence (BI)
- Knowledge Data Discovery (KDD)
- Další/ostatní funkcionality (Metadata management, Data Governance & Data Quality tool (DG))



Níže uvedený obrázek popisuje logické celky a jejich vazby, které je nutno v rámci projektu řešit SW nástroji.

Obrázek 1: Schéma logických celků



V rámci projektu, resp. jeho následném využití se předpokládá zpracování velkých objemů dat, a to z více zdrojů. Průzkum trhu byl orientován na sloupcově orientované databázové systémy na rozdíl od klasických řádkových relačních databází.

Pro průzkum trhu byli vybráni klíčoví dodavatelé produktů v oblastech poskytujících komponenty pro datovou integraci, datový sklad, business intelligence a KDD (knowledge data Discovery).

1.2 Základní požadavky na navržené řešení

1.2.1 Logické součásti systému

Navržený systém musí být složen z následujících logických součástí, které zajišťují požadované schopnosti a funkce:

Komponenty	Schopnosti a funkce
Data subsystému	integrace Vizualní transformace, zpracování dat ze souborů, databází, webových služeb, dalších komponent tohoto systému
Obchod DWH	Databáze datového skladu pro efektivní ukládání a získávání (meta)dat
BI OLAP	Rozměrové modelování, OLAP kostky, kontingenční tabulky, multidimenzionální dotazování, samoobslužné analýzy
Pokročilé DM	Nástroje pro dolování dat pro pokročilé analýzy (statistický motor, modul obchodních pravidel, regresní modely a klasifikace, shlukování,



asociativní modely a porovnávání vzorů, ...)

Vykazování	Online předdefinované nebo samoobslužné analýzy, export na text, pdf, slovo, excel, powerpoint, obrázky, ...
Specializované komponenty	Nástroje pro správu systému, nástroje pro správu metadat, nástroje QA, ...

Skutečné technické řešení nemusí přesně odpovídat těmto komponentám, ale musí plnit své schopnosti.

1.2.2 Základní požadavky na navržené řešení

Následující seznam shrnuje obecné požadavky v systému bez přesné specifikace.

- 1) Složitost: Všechny základní prvky procesu BI&DM jsou zahrnuty – získávání, zpracování a integrace dat, jejich ukládání, základní i pokročilé analýzy, generování sestav
- 2) Škálovatelnost: Systém by mohl být jednoduše vylepšen, aby se zvýšil výkon novým (nejlépe nezávislým) hardwarem a licencemi za rozumnou cenu
- 3) Modularita, otevřenost a interoperabilita: Konkrétní procesní komponenty by mohly být řešeny pomocí různých specializovaných nástrojů, které by mohly být implementovány různými dodavateli. Tyto komponenty budou integrovány přes specifikovaná rozhraní dle přesných požadavků. Musí spolupracovat na několika úrovních – například některé pokročilé výsledky úlohy ukládání dat by mohly být použity jako vstup pro proces integrace dat; metadata s popisem procesu integrace dat budou uložena v DWH a dále analyzována atd.
- 4) Vyměnitelnost: Díky předchozím požadavkům mohly být některé součásti řešení vyměněny s některými alternativami během životního cyklu systému, což by mohlo lépe vyřešit měnící se očekávání. Náklady na samotnou změnu (migrace dat a metadat, integrace s jinými součástmi atd.) nesmí být příliš vysoké. Zvláštním požadavkem je umožnit použití řešení založeného na open-source a/nebo bezplatných nástrojích pro neziskové a vzdělávací účely, které bude co nejvíce kompatibilní se standardním systémem *DWH*.
- 5) Rozšiřitelnost: Systém lze snadno rozšířit o další nástroje a komponenty, které uživatelům poskytnou nové funkce a možnosti, které nejsou zahrnuty na začátku.
- 6) Zajištění kvality: Je třeba podporovat nástroj pro navrhování, ověřování a zlepšování procesů kvality dat a metadat. Je nezbytné podporovat validaci a zajistit úplnost, konzistentnost a aktuálnost zpracovávaných dat a umožnit validaci celého procesu (metadat) transformací dat od jejich získání až do konečného vykazování.
- 7) Bezpečnost: Systém bude použit pro zpracování velmi citlivých osobních zdravotních údajů, takže musí být zabezpečen proti všem vnějším nebo vnitřním hrozbám. Má podporu zabezpečené autentizace a autorizace, ukládání a komunikace. Musí umožnit protokolování a auditování operací provádění a čtení. Vzhledem k tomu, že tento systém bude k dispozici omezenému počtu pokročilých uživatelů, budou přístupová práva uživatelů nastavena na úroveň databáze, tabulky nebo sloupce. Přístupová práva na úrovni řádků nebo buněk



nemusí být podporována. Systém musí být provozován na místních serverech předkladatele. Cloudová nebo externě zajišťovaná řešení jsou zakázána z důvodu zákonných omezení.

- 8) Jednoduchost: Konkrétní komponenty a celý systém stejně, musí být velmi jednoduché na používání a správu. Stabilita řešení je zásadní.
- 9) Správa metadat a dat, archivace a zálohování: Musí existovat podpora nástrojů, které umožňují správu verzí a vývojový cyklus (GIT nebo jiné) a paralelní týmová spolupráce na všech procesech, úlohách (datových tocích), db schématech atd. Systém musí umožnit potvrzení a vrácení změn v úkolech, porovnání změn v datech a metadatech atd.
- 10) Požadavky na výkon: Systém by měl být navržen pro několik souběžných uživatelů, dávkové zpracování zdrojových dat a komplexní analýzy dolování dat. Kompletní proces integrace dat čtvrtletních přírůstků dat by neměl trvat déle než desítky nebo několik stovek minut. OLAP analýzy sestavy musí reagovat v maximálně několika sekundách. Očekává se vylepšená paralelizace procesu – hardwarová architektura bude nyní založena na serveru s více členy a více vlákny (až do mnoha desítek vláken). Očekává se také další vylepšení distribuované platformy.



2 Průzkum trhu

2.1 Vybraní dodavatelé

Pro průzkum trhu byli vybráni klíčoví dodavatelé produktů v oblastech poskytujících komponenty pro datovou integraci, datový sklad, business intelligence a KDD (knowledge data Discovery).

Tabulka 1: Seznam zvolených dodavatelů

Dodavatel	Nabízené řešení
InterSystems	HealthShare, DeepSee
CloverDX	CloverDX
Teradata Big Data Analytics	Teradata Aster Analytics
Microsoft	MSSQL, Azure, Power BI
Informatica	Informatica Data Integration Platform
SAS	SAS Data Analysis Platform Product Suite
Oracle	Oracle Dedicated Region Cloud @ Customer
IBM	IBM Analytics Platform

2.2 Způsob hodnocení

Požadavky uvedené v kapitole 1.2 Základní požadavky na navržené řešení byly rozčleněny na základní oblasti funkcionalit, které musí vybrané řešení obsahovat. Jedná se o:

- Data Integration (DI)
- Data Storage (DS)
- Business Intelligence (BI)
- Knowledge Data Discovery (KDD)
- Ostatní: Metadata management, Data Governance & Data Quality tool (DG)

Základním předpokladem pro hodnocení je skutečnost, že řešení zahrnuje veškeré výše zmíněné oblasti. **Dodavatelé, kteří tento předpoklad nenaplní, byli z podrobného hodnocení vyřazeni.**

Dále byly výše uvedené oblasti rozděleny do podrobnějších požadavků/požadovaných funkcionalit (viz dále v textu u každého z dodavatelů), a bylo hodnoceno, zda řešení těmito požadavkům vyhovuje (ANO/NE, popř. ČÁSTEČNĚ). Na základě toho byly jednotlivým oblastem přiřazeny body (ANO – 1 bod, ČÁSTEČNĚ – 0,5 bodu, NE – 0 bodů).

Následuje stručný popis vybraných dodavatelů a jejich komponent.



2.3 InterSystems

HealthShare je platforma standardně využívaná pro zdravotnickou informatiku, která zajišťuje strategickou interoperabilitu a aktivní analýzu informací zejména v nemocniční síti a v místních, regionálních či celostátních sítích. Je postavena na objektové databázi Caché – celosvětově nepoužívanější databázový systém v nemocničních aplikacích. InterSystems Ensemble platforma pro rychlou integraci a vývoj propojitelných aplikací.

InterSystems DeepSee je řešení Business Intelligence (BI), které umožňuje analýzu v reálném čase.

Je navržen tak, aby byl optimální pro vkládání do transakčních aplikací.

DeepSee má čtyři moduly (Architect, Connector, Designer, Analyzer), které poskytují možnosti vytváření datových modelů, analytických prvků, řídicích panelů a připojení k externím datům. Datový model DeepSee je postaven na základě transakčních dat z Caché, a proto nevyžadují tradiční datový sklad.

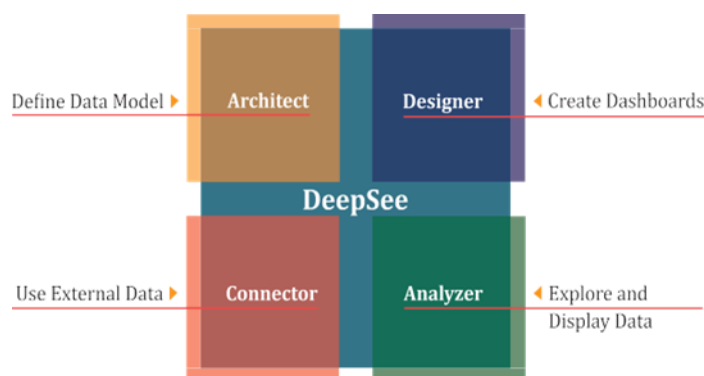
DeepSee Architect je určen k definici datového modelu, dimenzi datového modelu a jaká surová data budou v modelu zahrnuta. Pracuje se s transakčními daty, jejichž organizace a indexace právě tímto modelem v reálném čase.

DeepSee Connector poskytuje „snímek“ dat, která lze transformovat a odpovídala by tak datovému modelu. Snímek může být jednorázovým importem, nebo plánovanou úlohou.

DeepSee Designer slouží k vytváření řídicích panelů, grafů a kontingenčních tabulek, včetně odkazů a dalších součástí uživatelského rozhraní vytvořených DeepSee Analyzerem pro koncové uživatele.

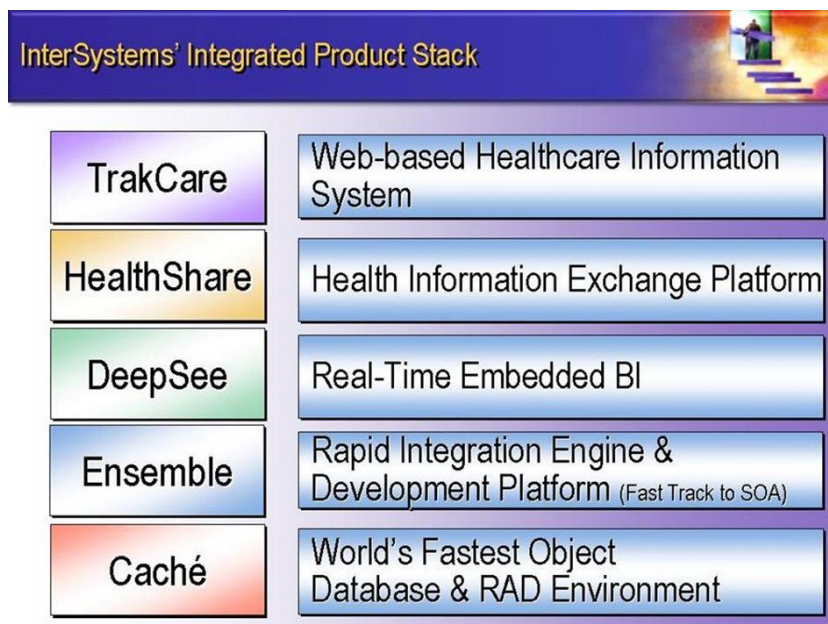
DeepSee Analyzer umožňuje vytvářet kontingenční tabulky, grafy vytvořené a definované DeepSee Architect metodou point-click nebo drag-and-drop. Komponenty jsou dynamické a uživatelům umožňují procházet data až k detailu.

Obrázek 2: Moduly DeepSEE





Obrázek 3: Komponenty produktů InterSystems'



2.3.1 Hodnocení

Z pohledu kompletnosti systémů InterSystems, vyhovuje řešení všem obecným požadavkům.

Oblasti	Ano/Ne	Produkt
Data Integration (DI)	ANO	HealthShare
Data Storage (DS)	ANO	HealthShare
Business Intelligence (BI)	ANO	HealthShare
Knowledge Data Discovery (KDD)	ANO	HealthShare + KNIME
Ostatní: Metadata management, Data Governance & Data Quality tool (DG)	ANO	HealthShare

Dále bylo provedeno hodnocení v jednotlivých oblastech:

2.3.1.1 Data Integration (DI)

Job Development

Vlastnosti	Ano/Ne
dataflow programming (https://en.wikipedia.org/wiki/Dataflow_programming)	ČÁSTEČNĚ
separate data transformation flow and process (OK/error) flow	ČÁSTEČNĚ
error (or exceptions) handling as a global trigger (callback)	ANO
data and error flow debugging - data preview at any point of a flow (+	ANO



breakpoints)	
multilevel node grouping / container	NE
include external jobs (subjobs)	ANO
iterations / loops	ANO

Engine type

Vlastnosti	Ano/Ne
ETL	ANO
manual in-database built-in function execution	ANO
automatic in-database built-in function execution	ANO
in-database UDF execution	ANO
ELT	ANO
Map&Reduce	NE

Build-in features

Vlastnosti	Ano/Ne
visual data mapping (src to dest columns)	ANO
SQL commit / rollback	ANO
data normalization	ANO
data denormalization	ANO
data pivoting	ANO
data (tables) comparison	NE
complex data (rows) generator	ANO
SCD level 1 generation	NE
SCD level 2 generation	NE
SCD level 3 generation	NE
regular expressions	ANO
GPG signing, en/de-crypting, validating	NE
X509 signing, en/de-crypting, validating	ANO
symmetrical data encryption	ANO
sensitive data processing support; sensitive data masking during job execution	ANO



multiline EBNF or other grammar (parse input data using given grammar)	NE
conditional data transformation operators/expressions (if, switch, etc.) inside nodes	ANO
business rules engine / rule based decisions	ANO

Celkem v oblasti Data Integration (DI) = 22 bodů

2.3.1.2 Data Storage (DS)

Storage type

Vlastnosti	Ano/Ne
row-oriented	ANO
column-oriented	ANO
document-oriented	ANO
multidimensional structure (OLAP)	ANO
HDFS (Hadoop)	NE
in-memory column-oriented RDBMS	ANO

Data types

Vlastnosti	Ano/Ne
SQL standard (int, varchar, date, etc)	ANO
binary	ANO
JSON	ANO
XML	ANO

Backup, restore & replication

Vlastnosti	Ano/Ne
fast / bulk load	ANO
dump to SQL insert syntax	ANO
hot-copy snapshots	ANO
hot-copy snapshots w/o query/update lock	ANO
cold-copy snapshots	ANO
CLI backup tool	ANO



client GUI backup tool	ANO
WebGUI backup tool	ANO
journal	ANO
master/slave replication	ANO
multi-master replication	ANO

Security

Vlastnosti	Ano/Ne
User permissions	
on actions - Standard SQL (select, update, ...)	ANO
to schema	ANO
to table / document	ANO
to row	ANO
to column / attribute	ANO
Encryption	
encryption support; please fill used methods too (e.g. AES256-CBC)	ANO
table / document encryption	ANO
column / attribute encryption	ANO

Celkem v oblasti Data Storage (DS) = 30 bodů

2.3.1.3 Business Intelligence (BI)

BI components

Vlastnosti	Ano/Ne
Server components	
OLAP server	ANO
operational reports	ANO
Editing components	
OLAP cubes editor	ANO
visual OLAP cubes editor for business users	ANO
visual report editor for business users	ANO
Viewing components	



OLAP viewer	ANO
report viewer	ANO

OLAP engine

Vlastnosti	Ano/Ne
OLAP type	
MOLAP	NE
ROLAP	NE
HOLAP	ANO
OLAP actions	
customizable dimensions and attributes hierarchy	ANO
data drill up/down	ANO
data drill through	ANO
data drill across	ANO
time-series auto generation	ANO

Results visualization

Vlastnosti	Ano/Ne
customizable multi-page tables	ANO
customizable pivot tables	ANO
grouped and stacked charts (bar, column, pie, donut, area, line)	ANO
XY / scatter plot	ANO
timeseries	ANO
box plots	ANO
bubble chart	ANO
calendar heatmap	NE
dendrogram	NE
parallel coordinates	ANO
net chart	NE
stock chart	ANO
gauge	ANO
Voronoi diagram	NE



Sankey diagram	NE
Gantt chart	NE
map, geo and custom SVG ROI charts	ANO
external GIS support	ANO
customizable JavaScript library	ANO
D3js library support	ANO
other custom actions support	ANO

Reporting engine

Vlastnosti	Ano/Ne
General	
multi-page templates	NE
web based interface	ANO
thin client	ANO
thick client	NE
prepared scheduled reports	ANO
self-serve / ad-hoc reports	ANO
report customization	ANO
conditional formatting	ANO
form input controls positionable anywhere in a report	NE
mobile device access (responsive design)	ANO
Importable templates and graphics	
templates in HTML + CSS	ANO
templates in DOC(X)	ANO
templates and graphics in SVG	ANO
templates and graphics in PDF or EPS	NE
other templates and graphics	NE
Internationalization	
multiple languages - i18n (translation)	ANO
multiple locales - l10n (format date, etc)	ANO

Security

Vlastnosti	Ano/Ne
------------	--------



facts permissions	ANO
facts permissions	ANO

Celkem v oblasti Business Intelligence (BI) = 42 bodů

2.3.1.4 Knowledge Data Discovery (KDD)

KDD components

Vlastnosti	Ano/Ne
Server components	
KDD server	ANO
Edifing components	
KDD console / script editor for expert users, e.g. SPSS Statistics (sKDD)	NE
KDD visual tool for business users, e.g. SAS Visual Analytics (vKDD)	ANO
KDD visual tool with dataflow programming, e.g. SPSS Modeller (mKDD)	ANO
Viewing components	
analysis viewer	ANO

KDD features

Vlastnosti	Ano/Ne
KDD techniques	
descriptive statistics	NE
associative rules	NE
clustering (w/ custom metrics)	NE
dimension reduction	NE
classification trees	NE
fuzzy classification	NE
neural networks (details)	NE
regression types	NE
survival analysis	NE
forecasting	NE
neural networks (details)	NE
threat / fraud detection	NE



time series	NE
text analytics	NE
others - write your own	NE
KDD standards support	NE
CRISP-DM principles support	NE

Results visualization

Vlastnosti	Ano/Ne
customizable multi-page tables	NE
customizable pivot tables	NE
grouped and stacked charts (bar, column, pie, donut, area, line)	NE
XY / scatter plot	NE
timeseries	NE
box plots	NE
bubble chart	NE
calendar heatmap	NE
dendrogram	NE
parallel coordinates	NE
net chart	NE
stock chart	NE
gauge	NE
Voronoi diagram	NE
Sankey diagram	NE
Gantt chart	NE
map, geo and custom SVG ROI charts	NE
external GIS support	NE
customizable JavaScript library	NE
D3js library support	NE
other custom actions support	NE

Celkem v oblasti Knowledge Data Discovery = 4 bodů.



CELKOVÉ HODNOCENÍ

Oblast	Body
Data Integration (DI)	22
Data Storage (DS)	30
Business Intelligence (BI)	42
Knowledge Data Discovery (KDD)	4
CELKEM bodů	98

Řešení Intersystems získalo v podrobném hodnocení 98 bodů.



2.4 CloverDX

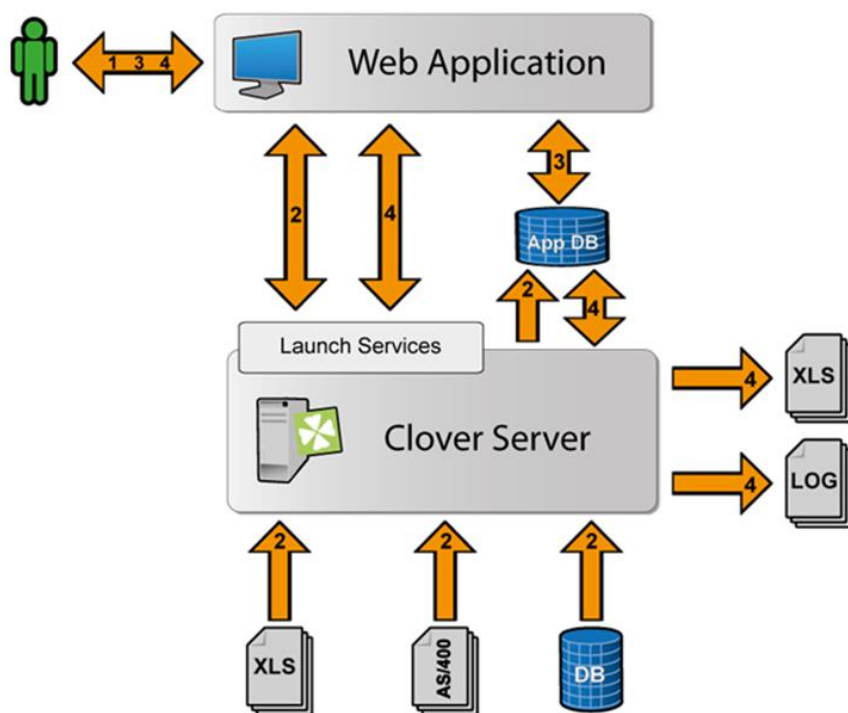
CloverDX Data Management Platform je Java integrační platforma pro integraci, transformaci, mapování a manipulaci s daty. CloverDX disponuje množstvím out-of-the-box konektorů, čímž se může připojit k celé řadě zdrojů dat včetně on-premise relačních databází (Oracle, MS SQL, Maria, PostgreSQL a mnoho dalších), databází NoSQL, cloudového úložiště (ADLS, CosmosDB, AWS S3, AWS Aurora, Snowflake, Redshift), aplikací (Salesforce, NetSuite, ...) a obecných API REST / SOAP. Cloudových služeb využívá MS Azure, Amazon Web Services a Google Cloud.

CloverETL je nástroj pro transformaci a integraci dat, používá se k čištění, standardizaci, transformaci a distribuci dat do aplikací, databází a skladů. Je to program založený na Javě a díky jeho struktuře založené na komponentách je možné přizpůsobení a vložení vlastních komponent. Může být použit samostatně jako aplikace příkazového řádku, serverová aplikace nebo dokonce zabudována do jiných aplikací.

CloverDX Designer je samostatná aplikace pro tvorbu a spuštění grafů. Zajišťuje připojení a komunikaci se serverem CloverDX, k vytváření projektů, grafů a všech dalších zdrojů na serveru CloverDX stejným způsobem, jako při práci s CloverDX Designer pouze nad lokálními daty.

CloverDX Server je serverová aplikace pro spuštění, plánování a monitorování grafů. Cluster CloverDX se skládá z několika instancí serverů CloverDX, které běží paralelně na různých uzlech, poskytuje funkce navíc vysokou dostupnost a škálovatelnost vyplývající z distribuovaného produkčního prostředí.

Obrázek 4: Schéma komunikace produktů CloverDX





2.4.1 Hodnocení

Z pohledu kompletnosti systémů CloverDX, NEvyhovuje řešení všem obecným požadavkům.

Oblasti	Ano/Ne	Produkt
Data Integration (DI)	ANO	CloverETL
Data Storage (DS)	NE	
Business Intelligence (BI)	NE	
Knowledge Data Discovery (KDD)	NE	
Ostatní: Metadata management, Data Governance & Data Quality tool (DG)	NE	

Vzhledem k tomu, že nástroje CloverDX nezahrnují veškeré oblasti zpracování dat, nebude řešení CloverDX dále hodnoceno.



2.5 Teradata Big Data Analytics

On-premise řešení disponuje komponentami, jako jsou Advanced Analytics, Teradata Integrated Data Ware House, Teradata Aster Analytics. Existuje mimo on-premise řešení i varianty v cloudu na platformě Vantage, která poskytuje sjednocenou analytiku, datová jezera a datové sklady - vše v cloudu.

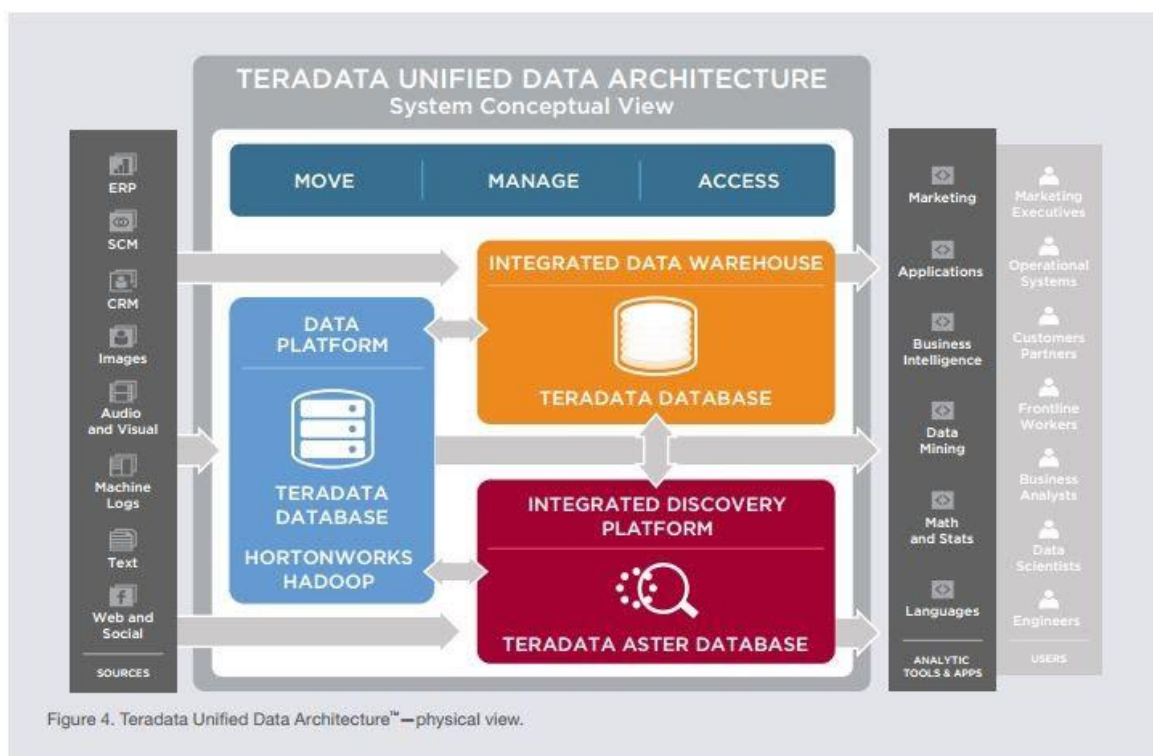
Vantage Analyst je komponenta Teradata Vantage poskytující kromě integrovaného datového a analytického prostředí, jež obsahuje mezifunkční analytiku ve velkém měřítku, rychlé samoobslužné vyhledávání a strojové učení napříč několika zdroji a typy dat, což vede k rychlejšímu a bohatšímu porozumění a lepším výsledkům analýz.

Vantage Analyst zajišťuje bezproblémovou integraci mezi uživatelským rozhraním, funkcemi strojového učení a vizualizačními technikami specifickými pro danou doménu. Umožňuje analytikům provádět komplexní analýzu bez nutnosti znalostí kódování.

Vantage Analyst Workflow zrychluje proces vývoje analytiky tím, že poskytuje point-and-click automatizaci opakovatelných úkolů.

Vantage Customer Experience je komponenta pro prediktivní analýzy pro organizace zaměřené na zákazníka, kde je snaha získat úplný přehled o interakcích se zákazníky, které pokrývají více kontaktních bodů a oddělení.

Obrázek 5: Schéma Teradata Unified Data Architecture





2.5.1 Hodnocení

Z pohledu kompletnosti systémů Teradata, vyhovuje řešení všem obecným požadavkům.

Oblasti	Ano/Ne	Produkt
Data Integration (DI)	ANO	Informatica PowerCenter with PowerExchange
Data Storage (DS)	ANO	Teradata Aster
Business Intelligence (BI)	ANO	IBM Cognos (alternatives are Tableau, Microstrategy,...)
Knowledge Data Discovery (KDD)	ANO	Teradata Aster with Teradata Aster AppCenter
Ostatní: Metadata management, Data Governance & Data Quality tool (DG)	ANO	Semanta, SAP PowerDesigner, Informatica Metadata Manager

Dále bylo provedeno hodnocení v jednotlivých oblastech:

2.5.1.1 Data Integration (DI)

Job Development

Vlastnosti	Ano/Ne
dataflow programming (https://en.wikipedia.org/wiki/Dataflow_programming)	ANO
separate data transformation flow and process (OK/error) flow	ANO
error (or exceptions) handling as a global trigger (callback)	ANO
data and error flow debugging - data preview at any point of a flow (+ breakpoints)	ANO
multilevel node grouping / container	ANO
include external jobs (subjobs)	ANO
iterations / loops	NE

Engine type

Vlastnosti	Ano/Ne
ETL	ANO
manual in-database built-in function execution	ANO
automatic in-database built-in function execution	ANO
in-database UDF execution	ANO
ELT	ANO



Map&Reduce	ANO
------------	-----

Build-in features

Vlastnosti	Ano/Ne
visual data mapping (src to dest columns)	ANO
SQL commit / rollback	ANO
data normalization	ANO
data denormalization	ANO
data pivoting	ANO
data (tables) comparison	ANO
complex data (rows) generator	ANO
SCD level 1 generation	ANO
SCD level 2 generation	ANO
SCD level 3 generation	ANO
regular expressions	ANO
GPG signing, en/de-crypting, validating	NE
X509 signing, en/de-crypting, validating	NE
symmetrical data encryption	ANO
sensitive data processing support; sensitive data masking during job execution	ANO
multiline EBNF or other grammar (parse input data using given grammar)	NE
conditional data transformation operators/expressions (if, switch, etc.) inside nodes	ANO
business rules engine / rule based decisions	ANO

Celkem v oblasti Data Integration (DI) = 27 bodů

2.5.1.2 Data Storage (DS)

Storage type

Vlastnosti	Ano/Ne
row-oriented	ANO
column-oriented	ANO
document-oriented	NE
multidimensional structure (OLAP)	NE



HDFS (Hadoop)	ANO
in-memory column-oriented RDBMS	NE

Data types

Vlastnosti	Ano/Ne
SQL standard (int, varchar, date, etc)	ANO
binary	NE
JSON	ANO
XML	ANO

Backup, restore & replication

Vlastnosti	Ano/Ne
fast / bulk load	ANO
dump to SQL insert syntax	ANO
hot-copy snapshots	NE
hot-copy snapshots w/o query/update lock	NE
cold-copy snapshots	NE
CLI backup tool	ANO
client GUI backup tool	NE
WebGUI backup tool	ANO
journal	ANO
master/slave replication	ANO
multi-master replication	ANO

Security

Vlastnosti	Ano/Ne
User permissions	
on actions - Standard SQL (select, update, ...)	ANO
to schema	ANO
to table / document	ANO
to row	ANO



to column / attribute	ANO
Encryption	
encryption support; please fill used methods too (e.g. AES256-CBC)	ANO
table / document encryption	ANO
column / attribute encryption	ANO

Celkem v oblasti Data Storage (DS) = 21 bodů

2.5.1.3 Business Intelligence (BI)

BI components

Vlastnosti	Ano/Ne
Server components	
OLAP server	ANO
operational reports	ANO
Editing components	
OLAP cubes editor	ANO
visual OLAP cubes editor for business users	ANO
visual report editor for business users	ANO
Viewing components	
OLAP viewer	ANO
report viewer	ANO

OLAP engine

Vlastnosti	Ano/Ne
OLAP type	
MOLAP	ANO
ROLAP	ANO
HOLAP	ANO
OLAP actions	
customizable dimensions and attributes hierarchy	ANO
data drill up/down	ANO
data drill through	ANO



data drill across	ANO
time-series auto generation	ANO

Results visualization

Vlastnosti	Ano/Ne
customizable multi-page tables	ANO
customizable pivot tables	ANO
grouped and stacked charts (bar, column, pie, donut, area, line)	ANO
XY / scatter plot	ANO
timeseries	ANO
box plots	ANO
bubble chart	ANO
calendar heatmap	ANO
dendrogram	ANO
parallel coordinates	ANO
net chart	ANO
stock chart	ANO
gauge	ANO
Voronoi diagram	ANO
Sankey diagram	ANO
Gantt chart	ANO
map, geo and custom SVG ROI charts	ANO
external GIS support	ANO
customizable JavaScript library	ANO
D3js library support	ANO
other custom actions support	NE

Reporting engine

Vlastnosti	Ano/Ne
General	
multi-page templates	ANO



web based interface	ANO
thin client	ANO
thick client	ANO
prepared scheduled reports	ANO
self-serve / ad-hoc reports	ANO
report customization	ANO
conditional formatting	ANO
form input controls positionable anywhere in a report	ANO
mobile device access (responsive design)	ANO
Importable templates and graphics	
templates in HTML + CSS	ANO
templates in DOC(X)	NE
templates and graphics in SVG	NE
templates and graphics in PDF or EPS	NE
other templates and graphics	ANO
Internationalization	
multiple languages - i18n (translation)	ANO
multiple locales - l10n (format date, etc)	ANO

Security

Vlastnosti	Ano/Ne
facts permissions	ANO
facts permissions	ANO

Celkem v oblasti Business Intelligence (BI) = 51 bodů

2.5.1.4 Knowledge Data Discovery (KDD)

KDD components

Vlastnosti	Ano/Ne
Server components	
KDD server	ANO
Editing components	



KDD console / script editor for expert users, e.g. SPSS Statistics (sKDD)	ANO
KDD visual tool for business users, e.g. SAS Visual Analytics (vKDD)	ANO
KDD visual tool with dataflow programming, e.g. SPSS Modeller (mKDD)	NE
Viewing components	
analysis viewer	ANO

KDD features

Vlastnosti	Ano/Ne
KDD techniques	
descriptive statistics	ANO
associative rules	ANO
clustering (w/ custom metrics)	ANO
dimension reduction	ANO
classification trees	ANO
fuzzy classification	ANO
neural networks (details)	ANO
regression types	ANO
survival analysis	ANO
forecasting	ANO
neural networks (details)	ANO
threat / fraud detection	ANO
time series	ANO
text analytics	ANO
others - write your own	ANO
KDD standards support	
CRISP-DM principles support	ANO

Results visualization

Vlastnosti	Ano/Ne
customizable multi-page tables	NE
customizable pivot tables	NE
grouped and stacked charts (bar, column, pie, donut, area, line)	ANO



XY / scatter plot	ANO
timeseries	ANO
box plots	ANO
bubble chart	ANO
calendar heatmap	ANO
dendrogram	ANO
parallel coordinates	ANO
net chart	ANO
stock chart	ANO
gauge	ANO
Voronoi diagram	ANO
Sankey diagram	ANO
Gantt chart	ANO
map, geo and custom SVG ROI charts	ANO
external GIS support	ANO
customizable JavaScript library	ANO
D3js library support	ANO
other custom actions support	NE

Celkem v oblasti Knowledge Data Discovery = 39 bodů.

CELKOVÉ HODNOCENÍ

Oblast	Body
Data Integration (DI)	27
Data Storage (DS)	21
Business Intelligence (BI)	51
Knowledge Data Discovery (KDD)	39
CELKEM bodů	138

Řešení TERADATA BIG DATA ANALYTICS získalo v podrobném hodnocení 138 bodů.



2.6 Microsoft

MSSQL je výkonný relační databázový systém založený na architektuře klient/server s integrovanou podporou jazyka XML a JSON, je škálovatelný a bezpečný.

Enterprise

Komplexní klíčová výkonná technologie uchování dat v paměti s vysokou dostupností, pro komplexní business řešení BI s integrovaným systémem BI pro mobilní zařízení a neomezená virtualizace díky programu Software Assurance. Edice Enterprise nabízí nejvyšší úroveň škálovatelnosti a výkonnosti pro nejdůležitější podnikové úlohy.

Standard

Vydání SQL Server Standard přináší základní správu dat a databázi business intelligence pro oddělení a malé organizace pro provozování jejich aplikací a podporuje běžné vývojové nástroje pro místní a cloudové aplikace – umožňuje efektivní správu databází s minimálními prostředky IT.

Web

Vydání SQL Server Web Edition pro webové hostitele a webové VAPs, které poskytují škálovatelnost, dostupnost a možnosti správy pro malé až velké webové vlastnosti.

Express Edition

Express Edition je bezplatná databáze na základní úrovni a je ideální pro výuku a vytváření aplikací založených na datových a stolních počítačích a malých serverech. Je to nejlepší volba pro nezávislé dodavatele softwaru, vývojáře a nadšence vytvářející klientské aplikace. Pro potřebu pokročilejších databázových funkcí, lze SQL Server Express bezproblémově upgradovat na jiné vyšší verze SQL Serveru.

SQL Server Express LocalDB je odlehčená verze Expressu, která má všechny své funkce programovatelnosti, běží v uživatelském režimu a má rychlou instalaci s nulovou konfigurací a krátký seznam předpokladů.



Tabulka 2: Porovnání jednotlivých edicí MS SQL včetně jejich limitů

Vlastnosti	Enterprise	Standard	Web	Express s pokročilými službami	Express
Maximální výpočetní kapacita použitá v jedné instanci - SQL Server Database Engine 1	Maximální operační systém	Omezeno na menší ze 4 soketů nebo 24 jader	Omezeno na menší ze 4 soketů nebo 16 jader	Omezeno na menší 1 soket nebo 4 jádra	Omezeno na menší 1 soket nebo 4 jádra
Maximální výpočetní kapacita použitá v jedné instanci - Analysis Services nebo Reporting Services	Maximální operační systém	Omezeno na menší ze 4 soketů nebo 24 jader	Omezeno na menší ze 4 soketů nebo 16 jader	Omezeno na menší 1 soket nebo 4 jádra	Omezeno na menší 1 soket nebo 4 jádra
Maximální paměť pro fond vyrovnávacích pamětí na instanci databázového stroje SQL Server	Maximální operační systém	128 GB	64 GB	1410 MB	1410 MB
Maximální paměť pro mezipaměť segmentu sloupců na instanci databázového stroje SQL Server	Neomezená paměť	32 GB	16 GB	352 MB	352 MB
Maximální velikost dat optimalizovaných pro paměť na databázi v databázovém stroji SQL Server	Neomezená paměť	32 GB	16 GB	352 MB	352 MB
Maximální využití paměti na instanci Analysis Services	Maximální operační systém	16 GB 2 64GB 3	N / A	N / A	N / A
Maximální využití paměti na instanci Reporting Services	Maximální operační systém	64 GB	64 GB	4 GB	N / A
Maximální velikost relační databáze	524 PB	524 PB	524 PB	10 GB	10 GB



Tabulka 3: Dostupnost k RDBMS

Vlastnosti	Enterprise	Standard	Web	Express s pokročilými funkcemi	Express
Podpora jádra serveru 1	Ano	Ano	Ano	Ano	Ano
Doručení protokolu	Ano	Ano	Ano	Ne	Ne
Zrcadlení databáze	Ano	Ano 2	Ano 3	Ano 3	Ano 3
Záložní komprese	Ano	Ano	Ne	Ne	Ne
Snímek databáze	Ano	Ano	Ano	Ano	Ano
Vždy v selhání clusteru instance 4	Ano	Ano	Ne	Ne	Ne
Skupiny vždy k dispozici 5	Ano	Ne	Ne	Ne	Ne
Skupiny základní dostupnosti 6	Ne	Ano	Ne	Ne	Ne
Automatické přesměrování připojení pro čtení a zápis	Ano	Ne	Ne	Ne	Ne
Obnovení online stránky a souboru	Ano	Ne	Ne	Ne	Ne
Vytvoření a přestavba online indexu	Ano	Ne	Ne	Ne	Ne
Obnoví se obnovitelný index online	Ano	Ne	Ne	Ne	Ne
Změna schématu online	Ano	Ne	Ne	Ne	Ne
Rychlá obnova	Ano	Ne	Ne	Ne	Ne
Zrychlené obnovení databáze	Ano	Ano	Ano	Ne	Ne
Zrcadlené zálohy	Ano	Ne	Ne	Ne	Ne
Hot přidejte paměť a CPU	Ano	Ne	Ne	Ne	Ne
Poradce pro obnovu databáze	Ano	Ano	Ano	Ano	Ano
Šifrovaná záloha	Ano	Ano	Ne	Ne	Ne



Vlastnosti	Enterprise	Standard	Web	Express s pokročilými	Express
Hybridní záloha do Windows Azure (záloha na URL)	Ano	Ano	Ne	Ne	Ne
Skupina bez clusterů 5,6	Ano	Ano	Ne	Ne	Ne
Failover servery pro zotavení po havárii 7	Ano	Ano	Ne	Ne	Ne
Failover servery pro vysokou dostupnost 7	Ano	Ano	Ne	Ne	Ne
Failover servery pro zotavení po havárii v Azure 7	Ano	Ano	Ne	Ne	Ne

Microsoft Power BI je analytická a reportingová platforma importuje data z různých zdrojů dat, jako jsou soubory, vlastní databáze, online cloudové zdroje.

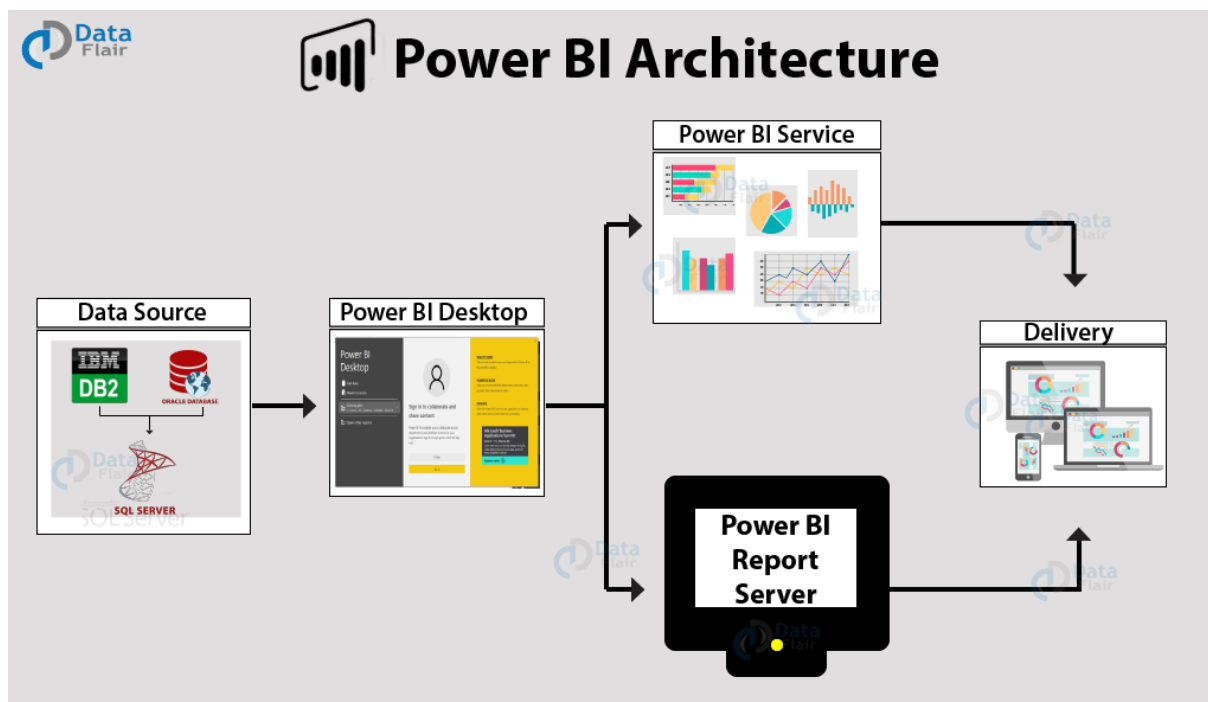
Power BI je jak cloudová platforma, tak i on-premise řešení ve verzi Power BI Pro. Prostřednictvím jednotlivých modulů Power BI jsou data transformována a formátována. Komponenta Power BI Desktop je vývojový nástroj pro zpracování dat pro prezentační nástroje a dostupný na platformách Windows, Android i iOS.

Disponuje formáty zdrojů dat jako Excel, text/CSV, XML, JSON, Oracle Database, IBM DB2 Database, MySQL, PostgreSQL, Sybase Database, Teradata Database a další.

Microsoft Azure je flexibilní a otevřený cloudový nástroj, který umožňuje rychlé vytváření, nasazení a zpracování aplikací napříč globální sítí datových center spravovaných společností Microsoft. Aplikace lze vytvářet pomocí jakéhokoli jazyka, rámce nebo platformy a dále integrovat do veřejných cloudových aplikací v požadovaném IT prostředí.



Obrázek 6: Schéma Power BI Architecture



2.6.1 Hodnocení

Z pohledu kompletnosti systémů Microsoft, vyhovuje řešení všem obecným požadavkům.

(Hodnocení je prováděno na úrovni Microsoft SQL Server Enterprise)

Oblasti	Ano/Ne	Produkt
Data Integration (DI)	ANO	SQL Server Integration Services (SSIS)
Data Storage (DS)	ANO	SQL Server - Database Engine, Analytics Platform Systém
Business Intelligence (BI)	ANO	SQL Server Reporting Services (SSRS), SQL Server Analysis Services (SSAS), Power BI, Datazen
Knowledge Data Discovery (KDD)	ANO	SQL Server Analysis Services (SSAS), Microsoft R Server, Azure Machine Learning
Ostatní: Metadata management, Data Governance & Data Quality tool (DG)	ANO	Data Quality Services (DQS), Master Data Services (MDS)

Dále bylo provedeno hodnocení v jednotlivých oblastech:



2.6.1.1 Data Integration (DI)

Job Development

Vlastnosti	Ano/Ne
dataflow programming (https://en.wikipedia.org/wiki/Dataflow_programming)	ANO
separate data transformation flow and process (OK/error) flow	ANO
error (or exceptions) handling as a global trigger (callback)	ANO
data and error flow debugging - data preview at any point of a flow (+ breakpoints)	ANO
multilevel node grouping / container	ANO
include external jobs (subjobs)	ANO
iterations / loops	ANO

Engine type

Vlastnosti	Ano/Ne
ETL	ANO
manual in-database built-in function execution	ANO
automatic in-database built-in function execution	ANO
in-database UDF execution	ANO
ELT	ANO
Map&Reduce	ANO

Build-in features

Vlastnosti	Ano/Ne
visual data mapping (src to dest columns)	ANO
SQL commit / rollback	ANO
data normalization	ANO
data denormalization	ANO
data pivoting	ANO
data (tables) comparison	ANO
complex data (rows) generator	ANO
SCD level 1 generation	ANO



SCD level 2 generation	ANO
SCD level 3 generation	ANO
regular expressions	ANO
GPG signing, en/de-crypting, validating	ANO
X509 signing, en/de-crypting, validating	ANO
symmetrical data encryption	ANO
sensitive data processing support; sensitive data masking during job execution	ANO
multiline EBNF or other grammar (parse input data using given grammar)	ANO
conditional data transformation operators/expressions (if, switch, etc.) inside nodes	ANO
business rules engine / rule based decisions	ANO

Celkem v oblasti Data Integration (DI) = 31 bodů

2.6.1.2 Data Storage (DS)

Storage type

Vlastnosti	Ano/Ne
row-oriented	ANO
column-oriented	ANO
document-oriented	ANO
multidimensional structure (OLAP)	ANO
HDFS (Hadoop)	ANO
in-memory column-oriented RDBMS	ANO

Data types

Vlastnosti	Ano/Ne
SQL standard (int, varchar, date, etc)	ANO
binary	ANO
JSON	ANO
XML	ANO



Backup, restore & replication

Vlastnosti	Ano/Ne
fast / bulk load	ANO
dump to SQL insert syntax	ANO
hot-copy snapshots	ANO
hot-copy snapshots w/o query/update lock	ANO
cold-copy snapshots	ANO
CLI backup tool	ANO
client GUI backup tool	ANO
WebGUI backup tool	ANO
journal	ANO
master/slave replication	ANO
multi-master replication	ANO

Security

Vlastnosti	Ano/Ne
User permissions	
on actions - Standard SQL (select, update, ...)	ANO
to schema	ANO
to table / document	ANO
to row	ANO
to column / attribute	ANO
Encryption	
encryption support; please fill used methods too (e.g. AES256-CBC)	ANO
table / document encryption	NE
column / attribute encryption	ANO

Celkem v oblasti Data Storage (DS) = 28 bodů

2.6.1.3 Business Intelligence (BI)

BI components

Vlastnosti	Ano/Ne
------------	--------



Server components	
OLAP server	ANO
operational reports	ANO
Editing components	
OLAP cubes editor	ANO
visual OLAP cubes editor for business users	ANO
visual report editor for business users	ANO
Viewing components	
OLAP viewer	ANO
report viewer	ANO

OLAP engine

Vlastnosti	Ano/Ne
OLAP type	
MOLAP	ANO
ROLAP	ANO
HOLAP	ANO
OLAP actions	
customizable dimensions and attributes hierarchy	ANO
data drill up/down	ANO
data drill through	ANO
data drill across	ANO
time-series auto generation	ANO

Results visualization

Vlastnosti	Ano/Ne
customizable multi-page tables	ANO
customizable pivot tables	ANO
grouped and stacked charts (bar, column, pie, donut, area, line)	ANO
XY / scatter plot	ANO
timeseries	ANO
box plots	ANO



bubble chart	ANO
calendar heatmap	ANO
dendrogram	ANO
parallel coordinates	ANO
net chart	ANO
stock chart	ANO
gauge	ANO
Voronoi diagram	ANO
Sankey diagram	ANO
Gantt chart	ANO
map, geo and custom SVG ROI charts	ANO
external GIS support	ANO
customizable JavaScript library	ANO
D3js library support	ANO
other custom actions support	ANO

Reporting engine

Vlastnosti	Ano/Ne
General	
multi-page templates	ANO
web based interface	ANO
thin client	ANO
thick client	ANO
prepared scheduled reports	ANO
self-serve / ad-hoc reports	ANO
report customization	ANO
conditional formatting	ANO
form input controls positionable anywhere in a report	ANO
mobile device access (responsive design)	ANO
Importable templates and graphics	
templates in HTML + CSS	ANO
templates in DOC(X)	NE



templates and graphics in SVG	NE
templates and graphics in PDF or EPS	NE
other templates and graphics	NE
Internationalization	
multiple languages - i18n (translation)	ANO
multiple locales - l10n (format date, etc)	ANO

Security

Vlastnosti	Ano/Ne
facts permissions	ANO
facts permissions	ANO

Celkem v oblasti Business Intelligence (BI) = 58 bodů

2.6.1.4 Knowledge Data Discovery (KDD)

KDD components

Vlastnosti	Ano/Ne
Server components	
KDD server	ANO
Editing components	
KDD console / script editor for expert users, e.g. SPSS Statistics (sKDD)	ANO
KDD visual tool for business users, e.g. SAS Visual Analytics (vKDD)	ANO
KDD visual tool with dataflow programming, e.g. SPSS Modeller (mKDD)	ANO
Viewing components	
analysis viewer	ANO

KDD features

Vlastnosti	Ano/Ne
KDD techniques	
descriptive statistics	ANO
associative rules	ANO
clustering (w/ custom metrics)	ANO



dimension reduction	ANO
classification trees	ANO
fuzzy classification	ANO
neural networks (details)	ANO
regression types	ANO
survival analysis	ANO
forecasting	ANO
neural networks (details)	ANO
threat / fraud detection	ANO
time series	ANO
text analytics	ANO
others - write your own	ANO
KDD standards support	
CRISP-DM principles support	ANO

Results visualization

Vlastnosti	Ano/Ne
customizable multi-page tables	ANO
customizable pivot tables	ANO
grouped and stacked charts (bar, column, pie, donut, area, line)	ANO
XY / scatter plot	ANO
timeseries	ANO
box plots	ANO
bubble chart	ANO
calendar heatmap	ANO
dendrogram	ANO
parallel coordinates	ANO
net chart	ANO
stock chart	ANO
gauge	ANO
Voronoi diagram	ANO
Sankey diagram	ANO



Gantt chart	ANO
map, geo and custom SVG ROI charts	ANO
external GIS support	ANO
customizable JavaScript library	ANO
D3js library support	ANO
other custom actions support	ANO

Celkem v oblasti Knowledge Data Discovery = 40 bodů.

CELKOVÉ HODNOCENÍ

Oblast	Body
Data Integration (DI)	31
Data Storage (DS)	28
Business Intelligence (BI)	58
Knowledge Data Discovery (KDD)	40
CELKEM bodů	157

Řešení MICROSOFT získalo v podrobném hodnocení 157 bodů.



2.7 Informatica

Architektura **PowerCenter** je založena na konceptu Service Oriented Architecture (SOA). Architektura orientovaná na služby (SOA) může být definována jako skupina služeb, které spolu komunikují.

PowerCenter je platforma pro integraci dat. Umožňuje práci v reálném čase a vytváření integračních kompetenčních center. Skládá se ze dvou komponent, klientské – Power Center Client a serverové – Power Center Server.

Power Center Client obsahuje aplikace Designer, Workflow Manager, Workflow Monitor.

Power Center Designer slouží k návrhu, jak přesouvat data mezi různými zdroji a cíli. Designer se používá k vytváření definic zdrojů, definic cílů a transformací, které lze dále využít pro vývoj mapování.

Power Center Workflow Manager Jedná se o uspořádanou sadu jedné nebo více relací a dalších úkolů navržených k dosažení celkového provozního účelu. Představuje řadu mapování (jako relace) a dalších úkolů.

Power Center Monitor je nástroj PowerCenter, se používá ke sledování provádění pracovních postupů a úkolů. Využívá se k zobrazení podrobností o pracovním postupu nebo spuštění úlohy v zobrazení Ganttova diagramu nebo zobrazení úlohy, k řízení pracovních postupů a úloh (spuštění, zastavení, přerušení).

Eviduje historii úloh, jež byly alespoň jednou spuštěny.

Server Informatica PowerCenter obsahuje komponenty s následujícími službami:

Úložiště - načítá, vkládá a aktualizuje metadata do databázových tabulek úložiště, včetně jeho správy.

Integrační služba spouští relace a pracovní postupy.

Služba SAP BW vyhledává požadavky RFC od SAP BW a zahajuje pracovní postupy pro extrahování dat nebo načítání dat do SAP BW.

Centrum webových služeb přijímá požadavky od klientů webových služeb a vystavuje pracovní toky PowerCenter jako služby.

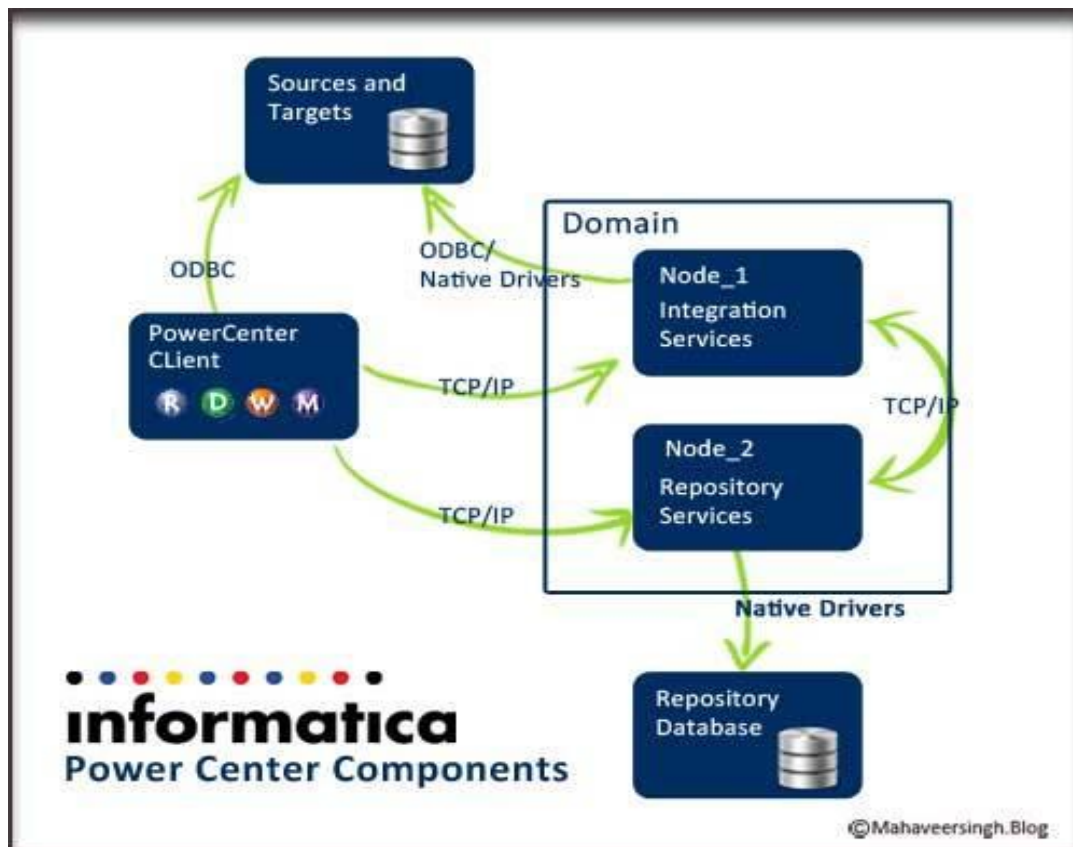
Portfolio nástrojů pro integraci dat společnosti Informatica zahrnuje nasazení v cloudu i pro řadu případů použití on-premise. Informatica kombinuje pokročilé hybridní integrační a řídicí funkce se samoobslužným uživatelským přístupem pro různé analytické operace. Rozšířená integrace je možná pomocí **CLAIRE Engine**, disponující umělou inteligencí (AI) engine řízeného metadata, který používá strojové učení. Společnost Informatica nabízí silnou interoperabilitu mezi rostoucím seznamem softwarových produktů pro správu dat.

CLAIRE zlepšuje schopnosti samoobslužných dat koncových uživatelů inteligentním doporučováním datových sad, které by mohly být využitelné. CLAIRE může automaticky odvodit podobná data z uživatelských značek, objevovat vztahy mezi daty a pomocí přiřazování schémat založených na AI propojit cílová schémata se zdrojovými daty a pomáhat automatizovat integraci dat. Schopnosti



přiřazování entit s vylepšenou inteligencí umožňují přesnější porovnávání nestructurovaných dat a dat strukturovaná částečně.

Obrázek 7: Schéma Informatica Power Center Components



2.7.1 Hodnocení

Z pohledu kompletnosti systémů Informatica, vyhovuje řešení všem obecným požadavkům.

Oblasti	Ano/Ne	Produkt
Data Integration (DI)	ANO	Informatica PowerCenter Advanced Edition
Data Storage (DS)	NE	
Business Intelligence (BI)	NE	
Knowledge Data Discovery (KDD)	NE	
Ostatní: Metadata management, Data Governance & Data Quality tool (DG)	ANO	Informatica PowerCenter Advanced Edition, Informatica Data Quality

Vzhledem k tomu, že nástroje Informatica nezahrnují veškeré oblasti zpracování dat, nebude řešení Informatica dále hodnoceno.



2.8 SAS

SAS Visual Analytics je webové prostředí, které podporuje několik aplikací. Umožňuje vytvářet interaktivní panely nebo reporty bezprostředně publikované na webu. Tento nástroj slouží k predikcím a analýzám. Existuje v distribuované a nedistribuované verzi.

SAS Analytic Data Store je datovým skladem společnosti SAS, **SAS Visual Statistics (VA)**, **SAS Enterprise Guide** jsou nástroje pro BI.

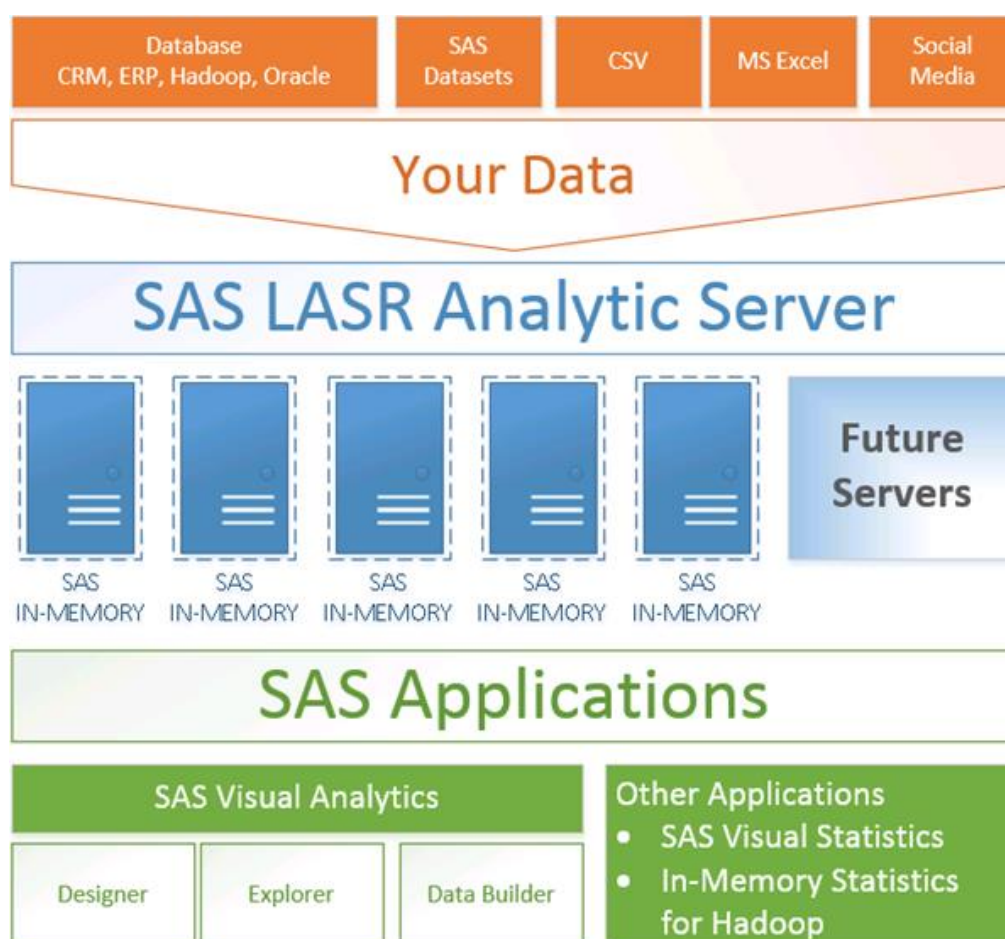
SAS Enterprise Guide (EG) je vývojové prostředí pro jazyk SAS, který zajišťuje import, transformaci, analýzu, vizualizaci a reporting dat.

SAS Visual Statistics (VA) a **SAS Enterprise Guide (EG)** jsou obdobné nástroje. SAS EG slouží k přípravě dat, jejich analýze a prezentaci. SAS VA také řeší tuto problematiku a vzhledem k webovému rozhraní je pro uživatele výrazně komfortnější.

SAS Enterprise Miner je KDD Knowledge Data Discovery nástroj pro data mining, který umožňuje vývoj popisného a prediktivního modelování a vyhodnocování v databázi.

Portfolio SAS Big Data Analytics obsahuje komponenty pro výkonné dolování dat SAS, SAS Enterprise Miner, SAS Scoring Accelerator, SAS Text Miner, SAS Model Manager a SAS Visual Statistics.

Obrázek 8: Schéma SAS Architektury





2.8.1 Hodnocení

Z pohledu kompletnosti systémů SAS, vyhovuje řešení všem obecným požadavkům.

Oblasti	Ano/Ne	Produkt
Data Integration (DI)	ANO	SAS Data Inegration Studio
Data Storage (DS)	ANO	SAS Analytic Data Store
Business Intelligence (BI)	ANO	SAS Visual Analytics, SAS Visual Statiscs, SAS Enterprise Guide
Knowledge Data Discovery (KDD)	ANO	SAS Enterprise Miner
Ostatní: Metadata management, Data Governance & Data Quality tool (DG)	ANO	SAS Data Quality Standard

Dále bylo provedeno hodnocení v jednotlivých oblastech:

2.8.1.1 Data Integration (DI)

Job Development

Vlastnosti	Ano/Ne
dataflow programming (https://en.wikipedia.org/wiki/Dataflow_programming)	ANO
separate data transformation flow and process (OK/error) flow	ANO
error (or exceptions) handling as a global trigger (callback)	ANO
data and error flow debugging - data preview at any point of a flow (+ breakpoints)	ANO
multilevel node grouping / container	ANO
include external jobs (subjobs)	ANO
iterations / loops	ANO

Engine type

Vlastnosti	Ano/Ne
ETL	ANO
manual in-database built-in function execution	ANO
automatic in-database built-in function execution	ANO
in-database UDF execution	ANO
ELT	ANO



Map&Reduce	NE
------------	-----------

Build-in features

Vlastnosti	Ano/Ne
visual data mapping (src to dest columns)	ANO
SQL commit / rollback	ANO
data normalization	ANO
data denormalization	ANO
data pivoting	ANO
data (tables) comparison	ANO
complex data (rows) generator	ANO
SCD level 1 generation	ANO
SCD level 2 generation	ANO
SCD level 3 generation	ANO
regular expressions	ANO
GPG signing, en/de-crypting, validating	NE
X509 signing, en/de-crypting, validating	ANO
symmetrical data encryption	NE
sensitive data processing support; sensitive data masking during job execution	NE
multiline EBNF or other grammar (parse input data using given grammar)	ANO
conditional data transformation operators/expressions (if, switch, etc.) inside nodes	ANO
business rules engine / rule based decisions	ANO

Celkem v oblasti Data Integration (DI) = 27 bodů

2.8.1.2 Data Storage (DS)

Storage type

Vlastnosti	Ano/Ne
row-oriented	ANO
column-oriented	NE
document-oriented	NE
multidimensional structure (OLAP)	NE



HDFS (Hadoop)	NE
in-memory column-oriented RDBMS	NE

Data types

Vlastnosti	Ano/Ne
SQL standard (int, varchar, date, etc)	ANO
binary	NE
JSON	NE
XML	NE

Backup, restore & replication

Vlastnosti	Ano/Ne
fast / bulk load	ANO
dump to SQL insert syntax	NE
hot-copy snapshots	ANO
hot-copy snapshots w/o query/update lock	NE
cold-copy snapshots	ANO
CLI backup tool	NE
client GUI backup tool	ANO
WebGUI backup tool	NE
journal	NE
master/slave replication	NE
multi-master replication	NE

Security

Vlastnosti	Ano/Ne
User permissions	
on actions - Standard SQL (select, update, ...)	ANO
to schema	ANO
to table / document	ANO
to row	ANO



to column / attribute	ANO
Encryption	
encryption support; please fill used methods too (e.g. AES256-CBC)	ANO
table / document encryption	ANO
column / attribute encryption	NE

Celkem v oblasti Data Storage (DS) = 13 bodů

2.8.1.3 Business Intelligence (BI)

BI components

Vlastnosti	Ano/Ne
Server components	
OLAP server	ANO
operational reports	ANO
Editing components	
OLAP cubes editor	ANO
visual OLAP cubes editor for business users	ANO
visual report editor for business users	ANO
Viewing components	
OLAP viewer	ANO
report viewer	ANO

OLAP engine

Vlastnosti	Ano/Ne
OLAP type	
MOLAP	ANO
ROLAP	ANO
HOLAP	ANO
OLAP actions	
customizable dimensions and attributes hierarchy	ANO
data drill up/down	ANO
data drill through	ANO



data drill across	ANO
time-series auto generation	ANO

Results visualization

Vlastnosti	Ano/Ne
customizable multi-page tables	ANO
customizable pivot tables	ANO
grouped and stacked charts (bar, column, pie, donut, area, line)	ANO
XY / scatter plot	ANO
timeseries	ANO
box plots	ANO
bubble chart	ANO
calendar heatmap	ANO
dendrogram	ANO
parallel coordinates	ANO
net chart	ANO
stock chart	ANO
gauge	ANO
Voronoi diagram	ANO
Sankey diagram	ANO
Gantt chart	NE
map, geo and custom SVG ROI charts	ANO
external GIS support	ANO
customizable JavaScript library	NE
D3js library support	NE
other custom actions support	NE

Reporting engine

Vlastnosti	Ano/Ne
General	
multi-page templates	ANO



web based interface	ANO
thin client	ANO
thick client	ANO
prepared scheduled reports	ANO
self-serve / ad-hoc reports	ANO
report customization	ANO
conditional formatting	ANO
form input controls positionable anywhere in a report	ANO
mobile device access (responsive design)	ANO
Importable templates and graphics	
templates in HTML + CSS	ANO
templates in DOC(X)	ANO
templates and graphics in SVG	ANO
templates and graphics in PDF or EPS	ANO
other templates and graphics	NE
Internationalization	
multiple languages - i18n (translation)	ANO
multiple locales - l10n (format date, etc)	ANO

Security

Vlastnosti	Ano/Ne
facts permissions	ANO
facts permissions	ANO

Celkem v oblasti Business Intelligence (BI) = 50 bodů

2.8.1.4 Knowledge Data Discovery (KDD)

KDD components

Vlastnosti	Ano/Ne
Server components	
KDD server	ANO
Editing components	



KDD console / script editor for expert users, e.g. SPSS Statistics (sKDD)	ANO
KDD visual tool for business users, e.g. SAS Visual Analytics (vKDD)	ANO
KDD visual tool with dataflow programming, e.g. SPSS Modeller (mKDD)	ANO
Viewing components	
analysis viewer	NE

KDD features

Vlastnosti	Ano/Ne
KDD techniques	
descriptive statistics	ANO
associative rules	NE
clustering (w/ custom metrics)	ANO
dimension reduction	ANO
classification trees	ANO
fuzzy classification	ANO
neural networks (details)	NE
regression types	ANO
survival analysis	ANO
forecasting	NE
neural networks (details)	NE
threat / fraud detection	ANO
time series	NE
text analytics	NE
others - write your own	NE
KDD standards support	
CRISP-DM principles support	ANO

Results visualization

Vlastnosti	Ano/Ne
customizable multi-page tables	ANO
customizable pivot tables	ANO
grouped and stacked charts (bar, column, pie, donut, area, line)	ANO



XY / scatter plot	ANO
timeseries	NE
box plots	NE
bubble chart	ANO
calendar heatmap	ANO
dendrogram	NE
parallel coordinates	ANO
net chart	NE
stock chart	NE
gauge	ANO
Voronoi diagram	NE
Sankey diagram	NE
Gantt chart	ANO
map, geo and custom SVG ROI charts	NE
external GIS support	NE
customizable JavaScript library	ANO
D3js library support	ANO
other custom actions support	NE

Celkem v oblasti Knowledge Data Discovery = 24 bodů.

CELKOVÉ HODNOCENÍ

Oblast	Body
Data Integration (DI)	27
Data Storage (DS)	13
Business Intelligence (BI)	50
Knowledge Data Discovery (KDD)	24
CELKEM bodů	114

Řešení SAS získalo v podrobném hodnocení 114 bodů.



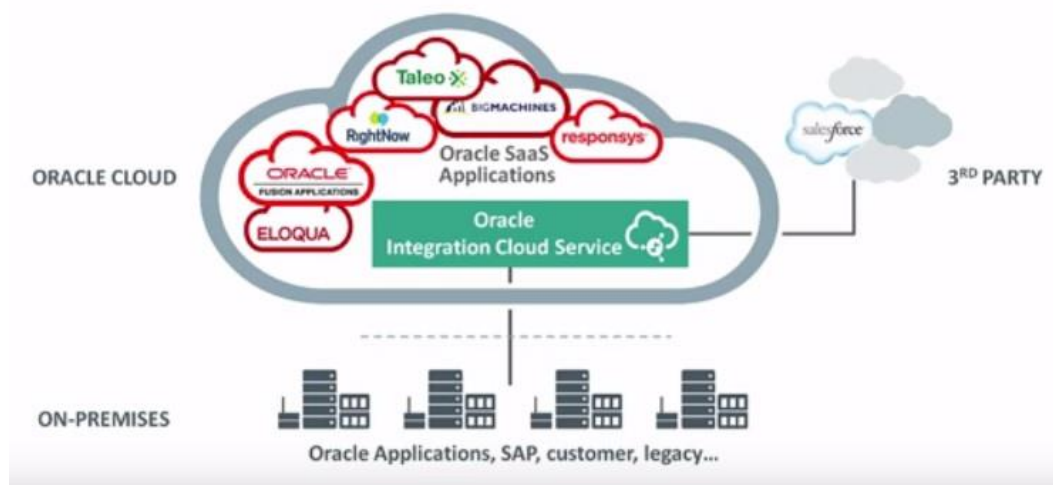
2.9 ORACLE

Platforma **Oracle Integration Cloud Service** poskytuje komponenty Oracle GoldenGate, Oracle Data Integrator, Oracle Big Data SQL, Oracle Service Bus, Cloud Data Integration Cloud Service.

Oracle nabízí celou škálu nástrojů pro integraci dat jak pro tradiční, tak i pro moderní případy, a to jak v implementacích, v cloudu, tak v on-premise. On-premise řešení je Oracle Dedicated Region Cloud @ Customer, které je ve vlastní infrastruktuře v podobě privátního cloudu ve správě společnosti Oracle.

Produktové portfolio společnosti zahrnuje technologie a služby, které organizacím umožňují plný pohyb a obohacení dat v životním cyklu. Integrace dat Oracle poskytuje všudypřítomný a nepřetržitý přístup k datům v heterogenních systémech prostřednictvím pohybu hromadných dat, transformace, obousměrné replikace, správy metadat, datových služeb a kvality dat pro zákazníky a produktové domény.

Obrázek 9: Oracle platforma



2.9.1 Hodnocení

Z pohledu kompletnosti systémů Microsoft, vyhovuje řešení všem obecným požadavkům.

Oblasti	Ano/Ne	Produkt
Data Integration (DI)	ANO	Oracle Data Integrator
Data Storage (DS)	ANO	Oracle Database
Business Intelligence (BI)	ANO	Oracle Business Intelligence
Knowledge Data Discovery (KDD)	ANO	Oracle Advanced Analytics, Oracle Business Intelligence, Oracle Data Miner
Ostatní: Metadata management, Data Governance & Data Quality tool (DG)	ANO	Oracle Data Integrator (ODI), Oracle Enterprise Metadata Management (OEMM), Oracle Data Quality (EDQ)



Dále bylo provedeno hodnocení v jednotlivých oblastech:

2.9.1.1 Data Integration (DI)

Job Development

Vlastnosti	Ano/Ne
dataflow programming (https://en.wikipedia.org/wiki/Dataflow_programming)	ANO
separate data transformation flow and process (OK/error) flow	ANO
error (or exceptions) handling as a global trigger (callback)	ANO
data and error flow debugging - data preview at any point of a flow (+ breakpoints)	ANO
multilevel node grouping / container	ANO
include external jobs (subjobs)	ANO
iterations / loops	ANO

Engine type

Vlastnosti	Ano/Ne
ETL	ANO
manual in-database built-in function execution	ANO
automatic in-database built-in function execution	ANO
in-database UDF execution	ANO
ELT	ANO
Map&Reduce	ANO

Build-in features

Vlastnosti	Ano/Ne
visual data mapping (src to dest columns)	ANO
SQL commit / rollback	NE
data normalization	ANO
data denormalization	ANO
data pivoting	ANO
data (tables) comparison	ANO
complex data (rows) generator	ANO



SCD level 1 generation	ANO
SCD level 2 generation	NE
SCD level 3 generation	NE
regular expressions	ANO
GPG signing, en/de-crypting, validating	NE
X509 signing, en/de-crypting, validating	NE
symmetrical data encryption	NE
sensitive data processing support; sensitive data masking during job execution	NE
multiline EBNF or other grammar (parse input data using given grammar)	NE
conditional data transformation operators/expressions (if, switch, etc.) inside nodes	ANO
business rules engine / rule based decisions	ANO

Celkem v oblasti Data Integration (DI) = 23 bodů

2.9.1.2 Data Storage (DS)

Storage type

Vlastnosti	Ano/Ne
row-oriented	ANO
column-oriented	ANO
document-oriented	ANO
multidimensional structure (OLAP)	ANO
HDFS (Hadoop)	ANO
in-memory column-oriented RDBMS	ANO

Data types

Vlastnosti	Ano/Ne
SQL standard (int, varchar, date, etc)	ANO
binary	ANO
JSON	ANO
XML	ANO



Backup, restore & replication

Vlastnosti	Ano/Ne
fast / bulk load	ANO
dump to SQL insert syntax	ANO
hot-copy snapshots	ANO
hot-copy snapshots w/o query/update lock	ANO
cold-copy snapshots	ANO
CLI backup tool	ANO
client GUI backup tool	ANO
WebGUI backup tool	ANO
journal	ANO
master/slave replication	ANO
multi-master replication	ANO

Security

Vlastnosti	Ano/Ne
User permissions	
on actions - Standard SQL (select, update, ...)	ANO
to schema	ANO
to table / document	ANO
to row	ANO
to column / attribute	ANO
Encryption	
encryption support; please fill used methods too (e.g. AES256-CBC)	ANO
table / document encryption	ANO
column / attribute encryption	ANO

Celkem v oblasti Data Storage (DS) = 29 bodů

2.9.1.3 Business Intelligence (BI)

BI components

Vlastnosti	Ano/Ne
------------	--------



Server components	
OLAP server	ANO
operational reports	ANO
Editing components	
OLAP cubes editor	ANO
visual OLAP cubes editor for business users	ANO
visual report editor for business users	ANO
Viewing components	
OLAP viewer	ANO
report viewer	ANO

OLAP engine

Vlastnosti	Ano/Ne
OLAP type	
MOLAP	ANO
ROLAP	ANO
HOLAP	ANO
OLAP actions	
customizable dimensions and attributes hierarchy	ANO
data drill up/down	ANO
data drill through	ANO
data drill across	ANO
time-series auto generation	NE

Results visualization

Vlastnosti	Ano/Ne
customizable multi-page tables	ANO
customizable pivot tables	ANO
grouped and stacked charts (bar, column, pie, donut, area, line)	ANO
XY / scatter plot	ANO
timeseries	ANO
box plots	ANO



bubble chart	ANO
calendar heatmap	ANO
dendrogram	ČÁSTEČNĚ
parallel coordinates	ČÁSTEČNĚ
net chart	ČÁSTEČNĚ
stock chart	ANO
gauge	ANO
Voronoi diagram	ČÁSTEČNĚ
Sankey diagram	ČÁSTEČNĚ
Gantt chart	ČÁSTEČNĚ
map, geo and custom SVG ROI charts	ČÁSTEČNĚ
external GIS support	ANO
customizable JavaScript library	ANO
D3js library support	ČÁSTEČNĚ
other custom actions support	ANO

Reporting engine

Vlastnosti	Ano/Ne
General	
multi-page templates	ANO
web based interface	ANO
thin client	ANO
thick client	NE
prepared scheduled reports	ANO
self-serve / ad-hoc reports	ANO
report customization	ANO
conditional formatting	ANO
form input controls positionable anywhere in a report	ANO
mobile device access (responsive design)	ANO
Importable templates and graphics	
templates in HTML + CSS	ANO
templates in DOC(X)	ANO



templates and graphics in SVG	NE
templates and graphics in PDF or EPS	ANO
other templates and graphics	ANO
Internationalization	
multiple languages - i18n (translation)	ANO
multiple locales - l10n (format date, etc)	ANO

Security

Vlastnosti	Ano/Ne
facts permissions	ANO
facts permissions	ANO

Celkem v oblasti Business Intelligence (BI) = 48 bodů

2.9.1.4 Knowledge Data Discovery (KDD)

KDD components

Vlastnosti	Ano/Ne
Server components	
KDD server	ANO
Editing components	
KDD console / script editor for expert users, e.g. SPSS Statistics (sKDD)	ANO
KDD visual tool for business users, e.g. SAS Visual Analytics (vKDD)	ANO
KDD visual tool with dataflow programming, e.g. SPSS Modeller (mKDD)	ANO
Viewing components	
analysis viewer	NE

KDD features

Vlastnosti	Ano/Ne
KDD techniques	
descriptive statistics	ANO
associative rules	ANO
clustering (w/ custom metrics)	ANO



dimension reduction	NE
classification trees	ANO
fuzzy classification	NE
neural networks (details)	ANO
regression types	ANO
survival analysis	ANO
forecasting	ANO
neural networks (details)	ANO
threat / fraud detection	ANO
time series	ANO
text analytics	ANO
KDD standards support	
CRISP-DM principles support	ANO

Results visualization

Vlastnosti	Ano/Ne
customizable multi-page tables	NE
customizable pivot tables	NE
grouped and stacked charts (bar, column, pie, donut, area, line)	NE
XY / scatter plot	NE
timeseries	NE
box plots	NE
bubble chart	NE
calendar heatmap	NE
dendrogram	NE
parallel coordinates	NE
net chart	NE
stock chart	NE
gauge	NE
Voronoi diagram	NE
Sankey diagram	NE
Gantt chart	NE



map, geo and custom SVG ROI charts	NE
external GIS support	NE
customizable JavaScript library	NE
D3js library support	NE
other custom actions support	NE

Celkem v oblasti Knowledge Data Discovery = 17 bodů.

CELKOVÉ HODNOCENÍ

Oblast	Body
Data Integration (DI)	23
Data Storage (DS)	29
Business Intelligence (BI)	48
Knowledge Data Discovery (KDD)	17
CELKEM bodů	117

Řešení ORACLE získalo v podrobném hodnocení 117 bodů.



2.10 IBM

IBM InfoSphere Information Server je platforma pro integraci dat, která zahrnuje řadu produktů, k vyčištění, transformování, sledování, a doručování dat. Informační server InfoSphere poskytuje možnosti masivního paralelního zpracování (MPP) a poskytuje vysoce škálovatelnou a flexibilní integrační platformu, která zpracovává všechny objemy dat.

Informační server InfoSphere je flexibilní k požadavkům na integraci informací - od integrace dat po kvalitu dat a správu dat - poskytovat důvěryhodné informace, které mají zásadní význam (jako jsou velká data a analytika, modernizace datového skladu, správa kmenových dat a analytika point-of-impact).

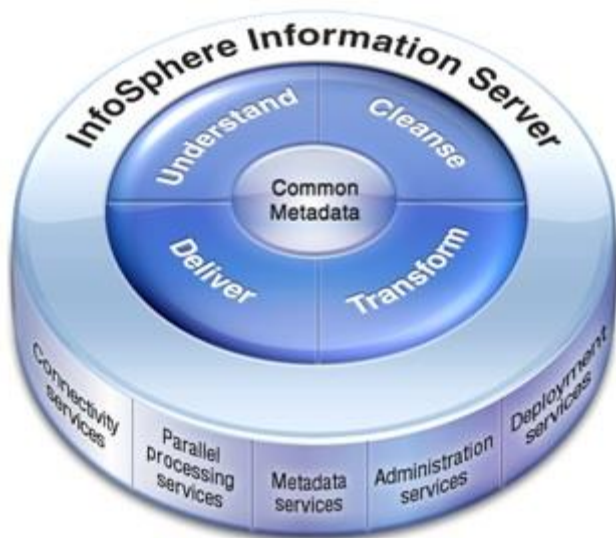
Tato výkonná, škálovatelná platforma pro extrakci, transformaci a načítání (ETL) zajistí flexibilní a integraci téměř všech typů dat v reálném čase, která lze nainstalovat jak on-premise tak v cloudu.

Server IBM Netezza Performance Server pro IBM Cloud Pak for Data je pokročilá datová skladová a analytická platforma dostupná jak on-premise, tak v cloudu. Netezza umožňuje strojové učení s objemy dat škálovanými až do petabytů.

Cognos Analytics je platforma pro business intelligence podporovaná umělou inteligencí, která podporuje celý analytický cyklus včetně vizualizace. Řešení je možné opět v cloudu, nebo on-premise.

Knowledge Data Discovery (KDD) - SPSS Modeler je přední řešení v oblasti vizuálních dat a strojového učení. Disponuje prediktivní analýzou s již předpřipravenými modely.

Obrázek 10: IBM platforma



2.10.1 Hodnocení

Z pohledu kompletnosti systémů IBM, nevyhovuje řešení všem obecným požadavkům.

Oblasti	Ano/Ne	Produkt
---------	--------	---------



Data Integration (DI)	ANO	IBM InfoSphere
Data Storage (DS)	ANO	IBM Netezza
Business Intelligence (BI)	ANO	IBM Cognos BI v10.2.2 (Express edition)
Knowledge Data Discovery (KDD)	ANO	IBM SPSS Modeler Professional v17.1
Ostatní: Metadata management, Data Governance & Data Quality tool (DG)	NE	

Vzhledem k tomu, že nástroje IBM nezahrnují veškeré oblasti zpracování dat, nebude řešení IBM dále hodnoceno.



3 Vyhodnocení průzkumu trhu

Požadavky uvedené v kapitole 1.2 Základní požadavky na navržené řešení byly rozčleněny na základní oblasti funkcionalit, které musí vybrané řešení obsahovat. Jedná se o:

- Data Integration (DI)
- Data Storage (DS)
- Business Intelligence (BI)
- Knowledge Data Discovery (KDD)
- Ostatní: Metadata management, Data Governance & Data Quality tool (DG)

Pro průzkum trhu byli vybráni klíčoví dodavatelé produktů v oblastech poskytujících komponenty pro datovou integraci, datový sklad, business intelligence a KDD (knowledge data Discovery).

Základním předpokladem pro hodnocení je skutečnost, že řešení zahrnuje veškeré výše zmíněné oblasti. **Dodavatelé, kteří tento předpoklad nenaplní, byli z podrobného hodnocení vyřazeni.**

Konkrétně se jednalo o řešení:

Tabulka 4: Seznam vyřazených dodavatelů z důvodu nenaplnění veškerých požadavků

Dodavatel	Nabízené řešení
CloverDX	CloverDX
Informatica	Informatica Data Integration Platform
IBM	IBM Analytics Platform

Dále byly výše uvedené oblasti rozděleny do podrobnějších požadavků/požadovaných funkcionalit (viz dále v textu u každého z dodavatelů), a bylo hodnoceno, zda řešení těmto požadavkům vyhovuje (ANO/NE, popř. ČÁSTEČNĚ). Na základě toho byly jednotlivým oblastem přiřazeny body (ANO – 1 bod, ČÁSTEČNĚ – 0,5 bodu, NE – 0 bodů).

Tabulka 5: Seznam hodnocených dodavatelů dle celkového počtu bodů

Dodavatel	Celkový počet bodů	Počet bodů DI	Počet bodů DS	Počet bodů BI	Počet bodů KDD
Microsoft	157	31	28	58	40
Teradata Big Data Analytics	138	27	21	51	39
Oracle	117	23	29	48	17
SAS	114	27	13	50	24
InterSystems	98	22	30	42	4

V podrobném hodnocení definovaných požadavků získalo nejvyšší počet bodů řešení MICROSOFT (na úrovni Microsoft SQL Server Enterprise).



3.1 Odůvodnění řešení MICROSOFT (Microsoft SQL Server Enterprise):

Microsoft SQL Server Enterprise zajišťuje pokročilé funkce s mimořádnými nároky na dostupnost, kapacitu a výkon databáze. Má neomezené množství výpočetních vláken – jader procesoru, neomezenou kapacitu operační paměti, neomezenou velikost databáze, více-nodové clusteru pro zajištění vysoké dostupnosti více databází. Poskytuje vyšší výkon díky čtení dat z více nodů clusteru. Obsahuje službu Software Assurance – neomezené množství databázových serverů.

Ve stávajících systémech ÚZIS je využívána jak edice MS SQL Server Standard, tak pro analytické účely MS SQL Server Enterprise.

Microsoft SQL Server Standard disponuje vysokým výkonem spolehlivostí. Je limitován 24 výpočetními vlákny – jádry procesoru. Maximálně 128 GB operační paměti. K zajištění vysoké dostupnosti jsou zde dvou-nodové clusteru pro jednu databázi. Čtení je jen z jednoho (primárního) clusteru. Pro každý server je potřeba individuální licence.

V rámci řešeného projektu „Optimalizace postupů při integraci prvků elektronického zdravotnictví do procesů ÚZIS a tvorba dokumentace k informačnímu systému vzácných onemocněních v ČR“, číslo CZ.03.4.74/0.0/0.0/15_025/0015811, a aktuálního enormního provozu registru infekčních nemocí (ISIN) a úložiště referenčních nemocnic (URN) dochází k nárůstu požadavků na analytické využití dat z výše uvedené databáze, proto při posuzování možných platforem bylo **zvoleno řešení Microsoft SQL Server Enterprise - jako řešení, které při výše uvedeném hodnocení získalo nejvíce bodů.**

Nebude se navíc jednat o pořízení kompletního nového řešení, ale pouze o upgrade Microsoft SQL Server Standard na Microsoft SQL Server Enterprise.

Národní zdravotnický informační systém využívá ve všech vrstvách procesu práce s jeho daty (sběr, ukládání, integrace, analýza, reporting) databázový nástroj Microsoft SQL Server (MS SQL) v edici Enterprise. Posledním prvkem kaskády serverů je analytický datový sklad NZIS registrů, který aktuálně již výkonově nevyhovuje (Microsoft SQL Server Standard – tedy nižší edice, nedostatečný počet licencí, absence šifrování uložených dat, absence škálování a řízení výkonu atd.).

Bude tedy nezbytné provést upgrade na MS SQL Server Enterprise, který by měl komplikované problémy vyřešit.

Nejen díky vlastnostem výše uvedeným bylo tedy zvoleno MS SQL, ale i s ohledem na činnosti, které již byly nad touto databází ale i nad veškerými databázemi NZIS provedeny. V případě nové databázové technologie by bylo nutné veškeré komponenty přepsat či upravit, v rozsahu od této analytické databáze až po všechny databáze NZIS. Dá se tedy předpokládat, že objemy prací by mohly jít až do několika tisíců MD.

Z důvodu efektivity a doposud vynaložené práce a nastavených procesů ve stávajících produktech bylo zvoleno řešení setrvání při již využívaném produktu Microsoft SQL Server, úprava spočívá pouze v jeho upgradu.

MS SQL je licencován podle edice a podle počtu využitých jader. V rámci v rámci projektu dojde k rozšíření licence o 32 výpočetních jader edice Enterprise (32 licencí per core).

Důvody k rozhodnutí na základě dosavadních zkušeností jsou pro MS SQL následující:



- Využití stávajících databází, procesů a nástrojů z rodiny MS SQL;
- Provoz databáze na běžném HW, nebo řešení virtualizací;
- Snadná škálovatelnost výkonu mezi jednotlivé databáze;
- Jednoduchý návrh DB bez nutnosti znát implementační detaily a metody optimalizace;
- Velmi rychlá odezva na analytické ad-hoc dotazy;
- Monitoring provozu, škálovatelnost, profilování, audit přístupu;
- Součástí balíku MS SQL již jsou nástroje pro správu a práci s MS SQL databází;
- Součástí jsou nástroje Business intelligence (SSIS, SSRS, SSAS);
- Dostatečná erudice personálu ÚZIS, dostupná dokumentace a vzdělávací materiály, kurzy;

3.2 Popis řešení / komponent

SQL Server Integration Services (SSIS) je součást databázového softwaru MS SQL Serveru, kterou lze použít k provádění široké škály úloh migrace dat.

SSIS je platforma pro integraci dat a aplikace a aplikace pracovního toku integraci dat a aplikace pracovního toku. Obsahuje nástroje pro datové sklady používané k extrakci, transformaci a načítání dat (ETL). Tento nástroj lze také použít k automatizaci údržby databází serveru SQL Server a aktualizací dat vícerozměrné OLAP krychle.

SSRS

SQL Server Reporting Services (SSRS) je serverový report generující softwarový systém. Je součástí sady služeb MS SQL Server, včetně SSAS (SQL Server Analysis Services) a SSIS (SQL Server Integration Services).

Je spravován prostřednictvím webového rozhraní a lze jej použít k přípravě a doručení různých interaktivních a tištěných zpráv. Služba SSRS poskytuje rozhraní do MS Visual Studio, k připojení k databázím SQL a používat nástroje SSRS k formátování sestav SQL mnoha složitými způsoby. Poskytuje také nástroj „Report Builder“ pro méně technické uživatele k formátování zpráv SQL ve zjednodušené formě.

SSAS

Microsoft SQL Server Analysis Services (SAAS), je nástroj pro online analytické zpracování OLAP kostek a dolování dat na MS SQL Serveru. SSAS je používán jako nástroj k analýze a pochopení dat, které mohou být rozloženy do více databází nebo v různých tabulkách nebo souborech.



4 Dodatečný nástroj

Krom nástrojů MS SQL Server Národní zdravotnický informační systém (NZIS) využívá pro analýzy v systému NRHZS **databázový nástroj HP Vertica**, který je naplněn daty v objemu přes 1 TB a je kontinuálně doplňován z jednotlivých registrů.

V rámci řešeného projektu „Optimalizace postupů při integraci prvků elektronického zdravotnictví do procesů ÚZIS a tvorba dokumentace k informačnímu systému vzácných onemocnění v ČR“, číslo CZ.03.4.74/0.0/0.0/15_025/0015811 budou částečně využívána rovněž data z výše uvedené databáze.

HP Vertica sice nespĺňuje veškeré požadavky uvedené a hodnocené v předchozích kapitolách, nicméně s ohledem na výsledek celkového hodnocení průzkumu a výběr již stávajícího používaného nástroje **MS SQL Server** (pouze v upgradované verzi), bylo z důvodu efektivity a doposud vynaložené práce ve stávajících produktech zvoleno řešení **připojení využití již používaného produktu HP Vertica ke zvolenému řešení MS SQL Server**. HP Vertica je licencována podle objemu zpracovávaných dat (po 1 TB) a v rámci projektu dojde k rozšíření licence.

Důvody pro volbu HP Vertica (jejího ponechání) - na základě dosavadních zkušeností - jsou následující:

- Provoz databáze je na běžném HW;
- Snadná škálovatelnost výkonu HW podle zátěže bez nutnosti dokupování licence;
- Jednoduchý návrh DB bez nutnosti znát implementační detaily a metody optimalizace;
- Velmi rychlá odezva na uživatelské ad-hoc dotazy;
- Monitoring provozu, profilování;
- Dostatečná erudice personálu ÚZIS, dostupná dokumentace a vzdělávací materiály, kurzy;

Krom výše uvedených vlastností, byl nástroj HP Vertica zvolen i s ohledem na náklady, které již byly v souvislosti s činnostmi v databázi a nad databází v NRHZS provedeny (viz níže tabulka).

Z toto pohledu se jeví volba ponechání HP Vertica jako efektivní řešení. V případě nové databáze by bylo nutné veškeré komponenty přepsat či upravit. Dá se předpokládat, že objemy prací by mohly jít až do několika stovek MD.

Tabulka 6: Činnosti již provedené v DB HP Vertica v rámci NRHZS

Komponenta	Metrika	Počet	Jednotky
db schéma	procesní a datové tabulky	1 300	ks
	procesní view	200	ks
	analytické tabulky	400	ks
	analytická view	690	ks
integrační proces	komponenty využívající DB	1 000	ks
	další sql skripty	1 000	kB
proces datové kvality	komponenty pro testy dat v DB	1 800	ks
	další sql skripty	750	kB
analytické výstupy		3 200	kB



4.1 Popis – HP Vertica

Sloupcově orientovaná platforma **Vertica Analytics** je navržena tak, aby spravovala velké, rychle rostoucí objemy dat a poskytovala velmi rychlý výkon dotazů při použití v datových skladech a dalších aplikacích náročných na dotazy. Výrazně vylepšuje výkon dotazů oproti tradičním relačním databázím a poskytuje vysokou dostupnost a škálovatelnost.

Je nezávislá na infrastruktuře a podporuje nasazení jak v cloudu, tak on-premise tak na cloudových platformách.

Klíčové vlastnosti HP Vertica:

- Sloupcově orientované paměťové uspořádání, které zvyšuje výkon sekvenčního přístupu k záznamům na úkor běžných transakčních operací orientovaných řádkově, jako je například vyhledávání jednoho záznamu, aktualizace a odstraňování.
- Standardní SQL rozhraní s mnoha vestavěnými analytickými schopnostmi, jako jsou vyplňování časově sériových řad/interpolací.
- Externí aktualizace a organizace hybridního úložiště, která zvyšují schopnost dotazování, vkládání a načítání, ale na úkor aktualizace a odstraňování.
- Komprese, která snižuje náklady na skladování dat a jejich propustnost (vstupní/výstupní). Vysoká komprese až 4:1, je možná, protože sloupce homogenního datového typu jsou uloženy společně, aktualizace na hlavním skladu jsou zpracovávány po dávkách.
- Nezávislá struktura (shared nothing architecture), která snižuje systémové neshody pro sdílené prostředky a umožňuje postupné snížení výkonu navzdory možnému selhání hardware.
- Snadné použití a údržba prostřednictvím automatizované replikace dat, serveru, optimalizace dotazů, a optimalizace skladování.
- Podpora pro standardní programovací rozhraní ODBC, JDBC a ADO.NET.

Obrázek 11: HP Vertica analytická platforma

The Analytics Platform





5 Závěr

Tento dokument popisuje proces „Výběru SW nástrojů“ v rámci projektu „Optimalizace postupů při integraci prvků elektronického zdravotnictví do procesů ÚZIS a tvorba dokumentace k informačnímu systému vzácných onemocnění v ČR“, registrační číslo CZ.03.4.74/0.0/0.0/15_025/0015811, a jeho **klíčové aktivity č. 3** *Pořízení, implementace a pilotní ověření využití softwarových nástrojů pro analýzy dat státní správy.*

Cílem aktivity je, vzhledem ke vzrůstající poptávce po rozsáhlých analytických hodnoceních dat z nových zdravotních registrů NZIS (zejména NRHZS), pořízení softwarového nástroje, sloužícího k efektivnímu zpracování a vytěžování datových zdrojů ÚZIS ČR, k jejich propojování a provádění transformací nad daty tak, aby byla zajištěna požadovaná validita a kvalita dat. Klíčovou vlastností je také rychlost zpracování (půjde o big data v rozsahu jednotek TB).

Zdůvodnění výběru SW nástrojů

Pro naplnění výše uvedené klíčové aktivity byl proveden průzkum trhu na základě top produktů v oblastech:

- Datová integraci (DI)
- Datový sklad (DS)
- Business Intelligence (BI)
- Knowledge Data Discovery (KDD)
- Další/ostatní funkcionality (Metadata management, Data Governance & Data Quality tool (DG))

5.1 Průzkum trhu

Pro průzkum trhu byli vybráni klíčoví dodavatelé produktů v oblastech poskytujících komponenty pro datovou integraci, datový sklad, business intelligence a KDD (knowledge data Discovery).

Základním předpokladem pro hodnocení je skutečnost, že řešení zahrnuje veškeré výše zmíněné oblasti. **Dodavatelé, kteří tento předpoklad nenaplnují, byli z podrobného hodnocení vyřazeni.**

Dále byly výše uvedené oblasti rozděleny do podrobnějších požadavků/požadovaných funkcionalit (viz dále v textu u každého z dodavatelů), a bylo hodnoceno, zda řešení těmto požadavkům vyhovuje.

Tabulka 7: Seznam hodnocených dodavatelů dle celkového počtu bodů

Dodavatel	Celkový počet bodů	Počet bodů DI	Počet bodů DS	Počet bodů BI	Počet bodů KDD
Microsoft	157	31	28	58	40
Teradata Big Data Analytics	138	27	21	51	39
Oracle	117	23	29	48	17
SAS	114	27	13	50	24
InterSystems	98	22	30	42	4



V podrobném hodnocení definovaných požadavků získalo nejvyšší počet bodů řešení MICROSOFT (na úrovni Microsoft SQL Server Enterprise).

Microsoft SQL Server Enterprise zajišťuje pokročilé funkce s mimořádnými nároky na dostupnost, kapacitu a výkon databáze. Má neomezené množství výpočetních vláken – jader procesoru, neomezenou kapacitu operační paměti, neomezenou velikost databáze, více-nodové clusteru pro zajištění vysoké dostupnosti více databází. Poskytuje vyšší výkon díky čtení dat z více nodů clusteru. Obsahuje službu Software Assurance – neomezené množství databázových serverů.

Ve stávajících systémech ÚZIS je využívána jak edice MS SQL Server Standard, tak pro analytické účely MS SQL Server Enterprise.

V rámci řešeného projektu „Optimalizace postupů při integraci prvků elektronického zdravotnictví do procesů ÚZIS a tvorba dokumentace k informačnímu systému vzácných onemocnění v ČR“, číslo CZ.03.4.74/0.0/0.0/15_025/0015811, a aktuálního enormního provozu registru infekčních nemocí (ISIN) a úložiště referenčních nemocnic (URN) dochází k nárůstu požadavků na analytické využití dat z výše uvedené databáze, proto při posuzování možných platforem bylo **zvoleno řešení Microsoft SQL Server Enterprise - jako řešení, které při výše uvedeném hodnocení získalo nejvíce bodů.**

Nebude se navíc jednat o pořízení kompletního nového řešení, ale pouze o upgrade Microsoft SQL Server Standard na Microsoft SQL Server Enterprise.

5.2 Dodatečný nástroj

Krom nástrojů MS SQL Server Národní zdravotnický informační systém (NZIS) využívá pro analýzy v systému NRHZS **databázový nástroj HP Vertica**, který je naplněn daty v objemu přes 1 TB a je kontinuálně doplňován z jednotlivých registrů.

V rámci řešeného projektu „Optimalizace postupů při integraci prvků elektronického zdravotnictví do procesů ÚZIS a tvorba dokumentace k informačnímu systému vzácných onemocnění v ČR“, číslo CZ.03.4.74/0.0/0.0/15_025/0015811 budou částečně využívána rovněž data z výše uvedené databáze.

HP Vertica sice nespĺňuje veškeré požadavky uvedené a hodnocené v předchozích kapitolách, nicméně s ohledem na výsledek celkového hodnocení průzkumu a výběr již stávajícího používaného nástroje **MS SQL Server** (pouze v upgradované verzi), bylo z důvodu efektivity a doposud vynaložené práce ve stávajících produktech zvoleno řešení **připojení využití již používaného produktu HP Vertica ke zvolenému řešení MS SQL Server**. HP Vertica je licencována podle objemu zpracovávaných dat (po 1 TB) a v rámci projektu dojde k rozšíření licence. Nástroj HP Vertica byl zvolen i s ohledem na náklady, které již byly v souvislosti s činnostmi v databázi a nad databází v NRHZS provedeny.

Z toto pohledu se jeví volba ponechání HP Vertica jako efektivní řešení. V případě nové databáze by bylo nutné veškeré komponenty přepsat či upravit. Dá se předpokládat, že objemy prací by mohly jít až do několika stovek MD.

Licence HP Vertica **nebude pořízena** z finančních prostředků předmětného projektu „Optimalizace postupů při integraci prvků elektronického zdravotnictví do procesů ÚZIS a tvorba dokumentace



Evropská unie
Evropský sociální fond
Operační program Zaměstnanost



k informačnímu systému vzácných onemocnění v ČR“, číslo CZ.03.4.74/0.0/0.0/15_025/0015811,
ale z jiných prostředků.