



Jedna verze pravdy pro všechny

Data Quality at a Glance

20.4.2010

vladimir.kyjonka@cze.sas.com

**THE
POWER
TO KNOW®**

Znáte toho muže?

žil:

- **427 – 347 př. Kr.**

vlastním jménem

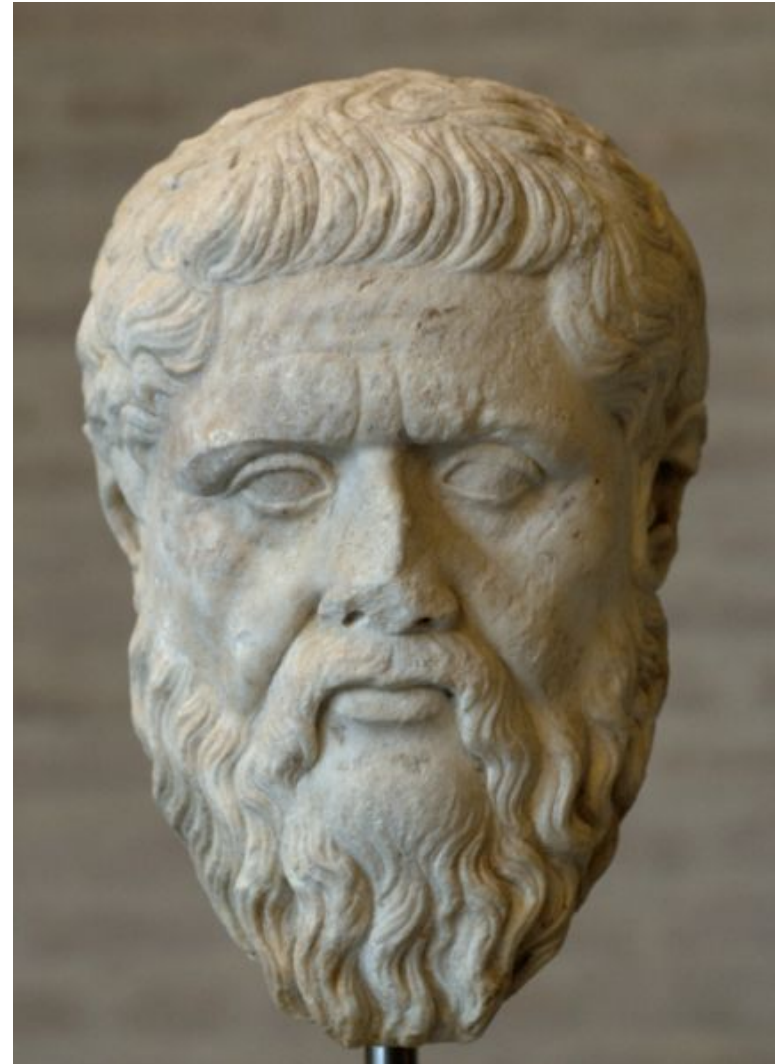
- **Aristoklés**

znám pod jménem:

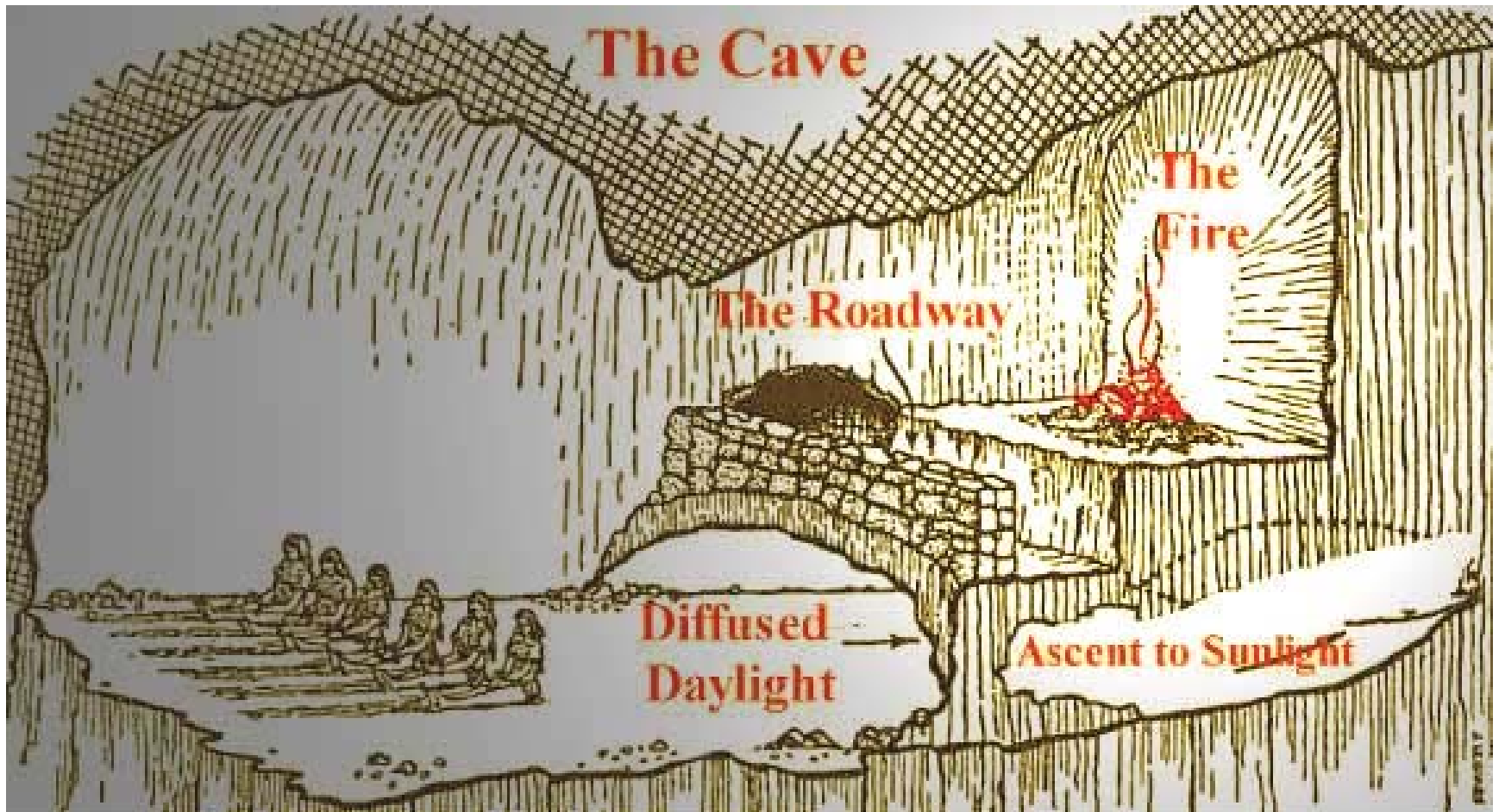
- **Πλάτων**

čili

- **Platón**



Podobenství o jeskyni



Jak je to s „jednou verzí pravdy“?

- Datové sklady
 - Cíl: Zjistit, jaký je pravý stav věcí
 - Cesta
 - Integrovat data o *něčem*
 - Forma:
 - Dimense
 - Konformované dimense
 - Centristický model DW
 - Bus Architecture
 - Realisace: Transformační vrstva DW
 - ETL, ELT, DQ...

Řešení v DW

- Vyzkoušené postupy pro konsolidaci dat
- Nejkritičtější, nejnáročnější, nejobtížnější a nejnákladnější část celého řešení DW
 - Táž informace z mnoha míst
 - Táž informace v různé podobě
 - Táž informace v různém čase
 - ...
- Transformace jedním směrem (downstream)
 - Neovlivňuje přímo provozní systémy
 - „Jenom“ vykazování, analýzy a reporting
 - Dodání informace není kritické ...

Historie: Ztráta jistot...

- 70.-80. léta
 - Centralismus, monolitické systémy, dávkové zpracování
 - „Jedna verze pravdy“ dána „seshora“:
 - Základní soubory
 - Kmenové soubory ...
- 80.-90. léta
 - Downsizing, rightsizing, distribuce...
 - => diversifikace, inkonsistence
- 90. léta až dnes:
 - Masivní zapojení „širokých mas“ do pořizování dat
 - Penetrace IT do stále více oblastí („Pokryto“)
- Dnes:
 - Dekáda integrace

Integrace

- Datová integrace, EAI, EII
- Datawarehousing
- Specialisované subjektově orientované systémy
 - CRM
 - (Centrální) zákaznické databáze a systémy
 - Zákaznické, produktové ODS
 - ...

Tři zdroje MDM

- Datové sklady
 - Dimense, konformované dimense ...
- Operativní systémy
 - Číselníky, kmenová data,...
- Speciální subjektově orientované systémy
 - Zákaznické systémy, CRM, ODS ...

Viditelný problém

- Důsledky diversifikace
 - Informace o subjektech se v datech vyskytují
 - mnohonásobně
 - na různých místech
 - v různých podobách

VS

- Požadavky integrace
 - Datové, funkční, informační propojení systémů
 - Integrovaný pohled na data (DW, portály, EII, BAM...)
 - Funkčně propojené systémy (EAI, Datová integrace, BPM...)

Kde se to projeví

- Technické problémy (příklady)
 - Nekompatibilita dat
 - Duplicity, balast => náklady
 - Špatná funkce systémů
 - ...
- Věcné (business) problémy (příklady)
 - Chybí jednotný pohled na klienta
 - Marketing, Risk management ...
 - Prokazatelnost reportovaných ukazatelů
 - Např. nesoulad analýz (DW) a kontroingu...
 - ...

Pustíme na to MDM

- Chtěli bychom:
 - aby podoby dat byly jednotné
 - aby jednotnost byla udržována:
- Údržba jednotnosti:
 - včas - t.j. když to business/aplikace vyžaduje
 - spolehlivě - t.j. správně a nezpochybnitelně
- Co je důležité:
 - Soustředit se na podstatné (oddělit zrna od plev) (nemrhat úsilím a penězi na marginality)
 - Kategorisace dle významu



MDM – co se od toho chce

Základní úlohy

- Rozeznat a najít všechna data o sybjektu (např. klient)
- Vytvořit, udržovat a poskytovat referenční data (master data)

Účel a podmínka

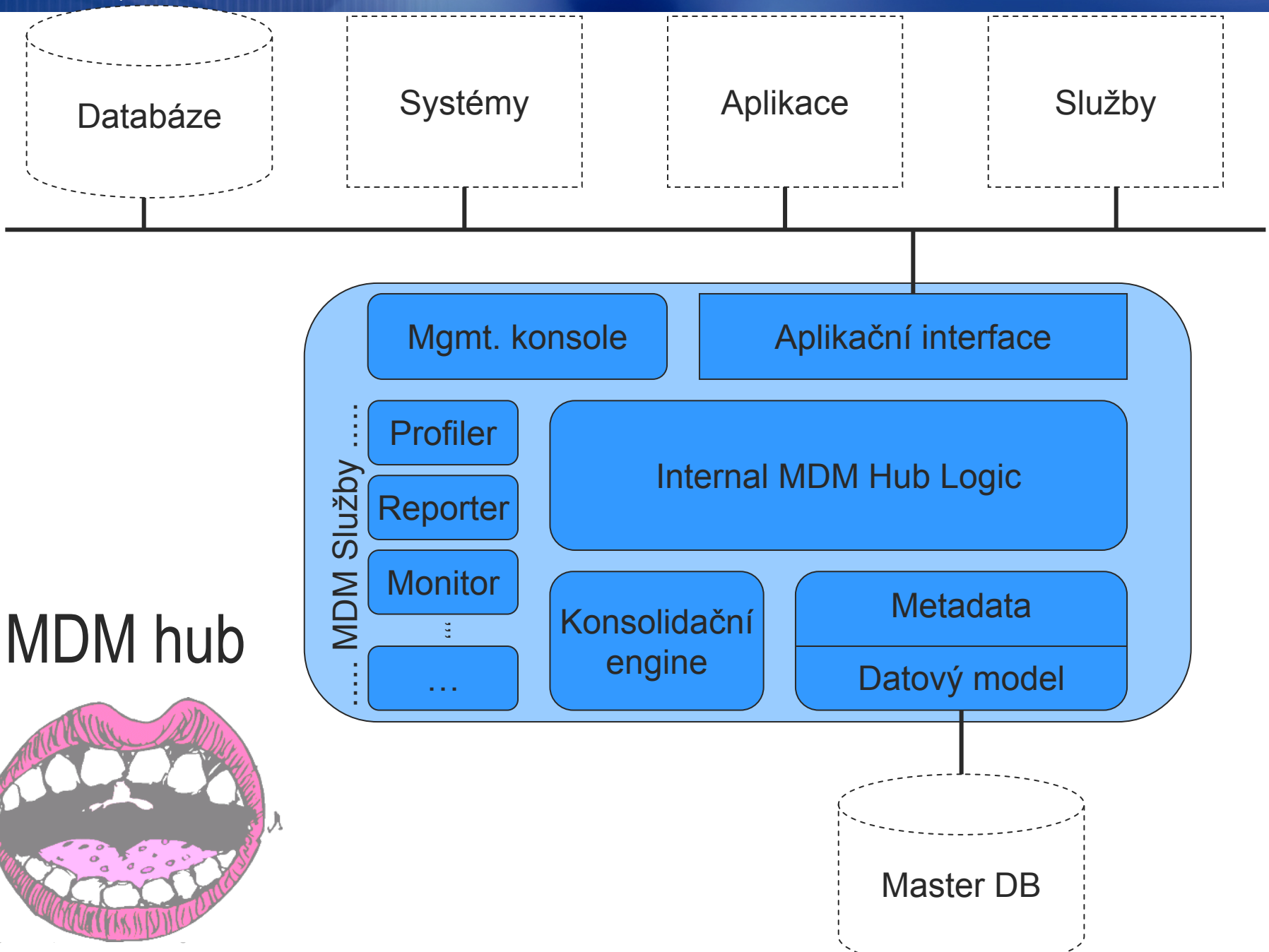
- Sdílení dat (informací)

3 součásti řešení MDM

- System of Record
 - Uložiště Master Dat
- MDM Hub
 - Služby přístupu k Master datům
 - Správa
- Integrovaná infrastruktura
 - Mechanismus sdílení dat

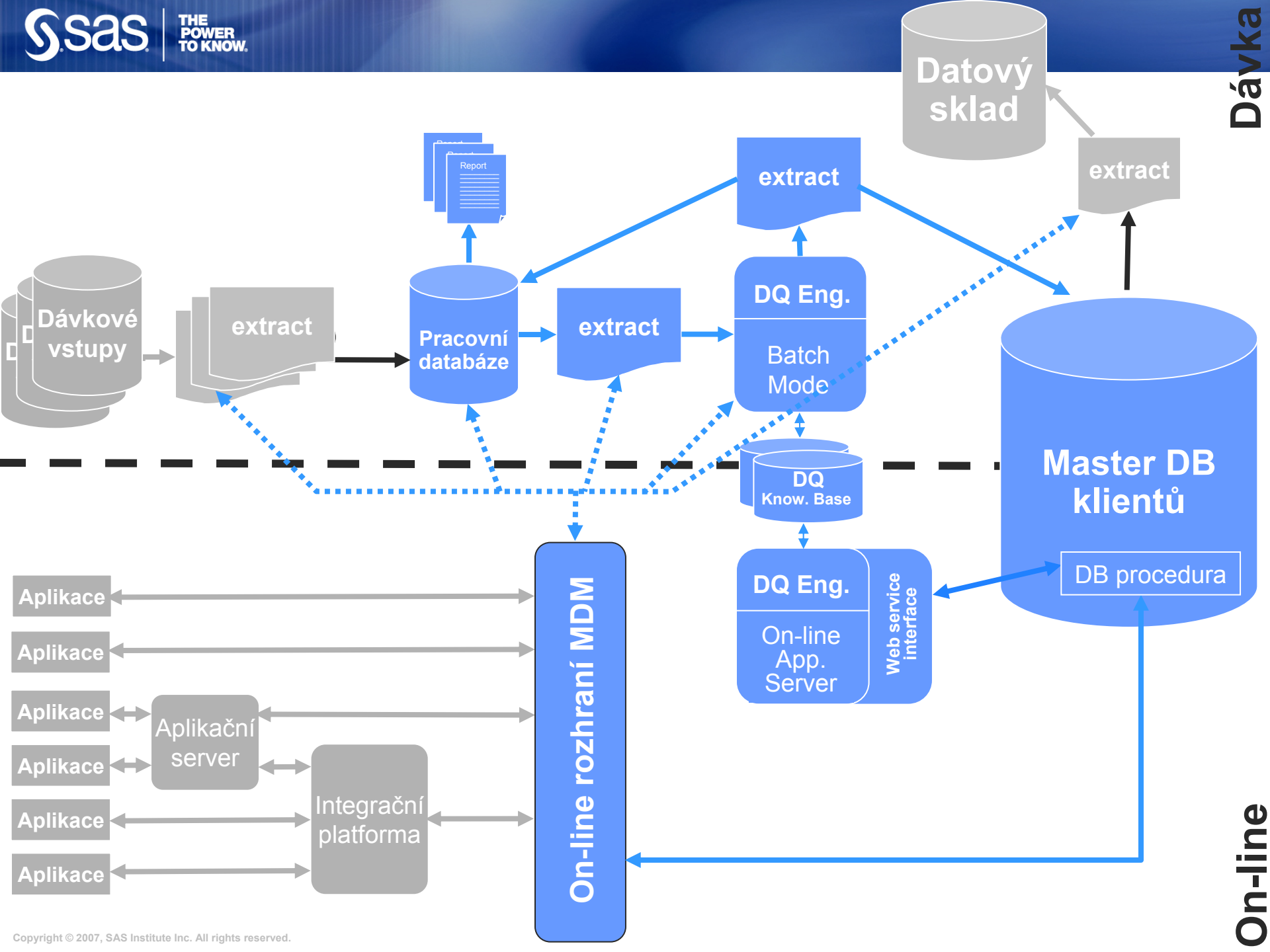
Typy MDM řešení

- Registry style
 - MDM Hub jenom eviduje záznamy ve zdrojových systémech
 - Systém of Record je jeden ze zdrojových systémů
 - Neposkytuje vyčištěná data
 - Snadno se implementuje
 - Vlastní informace je rozptýlená
- Transaction style (Repository style, Persisted Style)
 - Všechny atributy jsou ve speciální databázi MDM hubu (specialisovaná MDB)
 - Zdrojové systémy nemají vlastní data u sebe
 - Velmi náročná implementace, drahé, dlouhodobé...
- Hybrid (co-existence style)
 - Kombinace obého
 - Existuje speciální Master DB
 - Data zůstávají ve zdrojových systémech
 - Náročné na rekonciliaci (neustálá synchronisace)
- Consolidation style
 - „jednosměrné“ MDM
 - Pouze pro druhotné zpracování dat (např. publikace, analytické účely atd.)



MDM hub





Hlavní architektonické komponenty

- Podnikové aplikační systémy
 - On-line vstup dat
- Unifikovaný (jednotný) front-end pro *zákaznický* orientovaný přístup ke všem aplikacím
- Aplikační platformy (aplikační servery)
- Integrační prostředky (EAI, EII)
- Integrační platformy (ESB)
- Data Quality/Data Consolidation Engine
 - Včetně kontextu (Repository, číselníky atd.)
- (Centrální) databáze klientů (produktů...)
 - Včetně Master dat
- Front-end centrální databáze
 - Správa master dat
 - Primární aplikace pro některé agendy (např. Call center)
- DQ reporting a monitoring
- Rozhraní
 - primárních systémů
 - MDM služeb
 - Centrální DB
 - ...

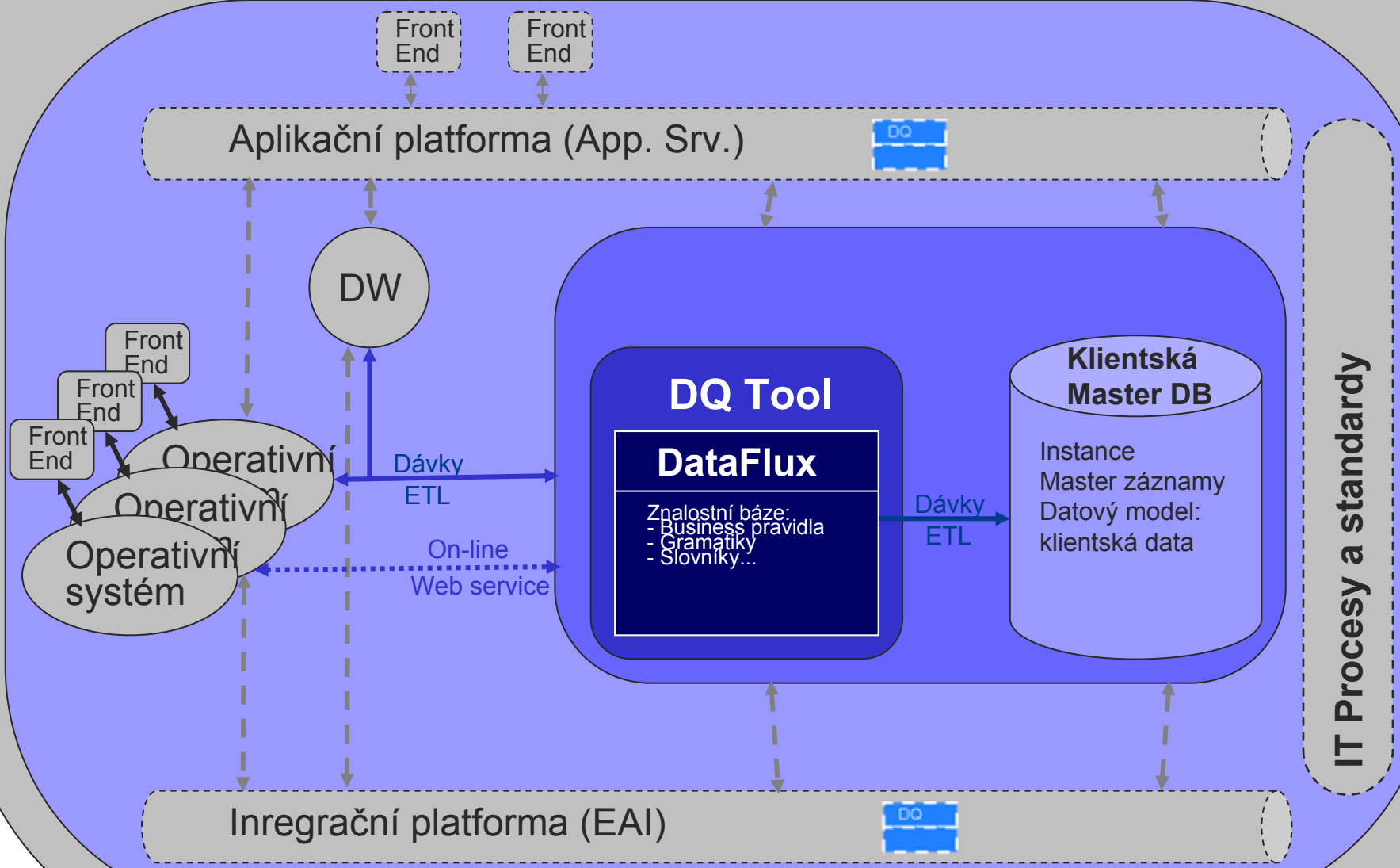
Požadované vlastnosti MDM produktu

Multientitní datový model
Podpora SOA
Hierarchy management
High-performance identity management
Podpora Data Governance (DQ monitor & inspection)
Podpora různých stylů práce
Podpora různých typů dat (semi a nestrukturovaná)
Podpora business procesů/policies

MDM v praxi

Podnikové procesy
Organisační struktura

Politika datové kvality
Data Governance ...



MDM v praxi

Podmínky fungování

- Technická infrastruktura
 - Architektura
 - Řešení
- Podniková iniciativa v oblasti péče o data a informace
 - např. Data Governance program, (ale i jiný – BI, SOA...)
 - Řízení na podnikobé úrovni (DG Body)
- Možnost a vůle sdílet informace (data)
 - Pravidla, „donucovací mechanismus“





Jedna verze pravdy pro všechny

Data Quality at a Glance

20.4.2010

vladimir.kyjonka@cze.sas.com

**THE
POWER
TO KNOW®**