



globtech

2. DUBNA 2013

ETL A MDM

ZKUŠENOSTI Z PRAXE

VÁŠ PARTNER NA CESTĚ OD DAT K INFORMACÍM

BUSINESS INTELLIGENCE DATA INTEGRATION SUCCESS
..... COMPLETE SOLUTIONS
WAREHOUSING

- 2003
ZALOŽEN POD ZNAČKOU **OPENTECH**
- 2005
VZNIKÁ SAMOSTATNÝ **GLOBTECH**,
SPECIALIZUJE SE NA DATOVÉ SKLADY
- DNES
30 PŘEVÁŽNĚ SENIORNÍCH KONZULTANTŮ,
GLOBTECH JE UZNÁVANÝM DODAVATELEM ANALYTICKÝCH A
OPERATIVNÍCH DATOVÝCH SKLADŮ A BI ŘEŠENÍ

■ DATOVÉ SKLADY

■ ANALYTICKÉ DATOVÉ SKLADY

- MENŠÍ POŽADAVKY NA AKTUÁLNOST
- MENŠÍ POŽADAVKY NA DOSTUPNOST – ČASTO DVA REŽIMY PRÁCE (DENNÍ / NOČNÍ)

■ OPERATIVNÍ DATOVÉ SKLADY (ODS)

- VYSOKÁ POTŘEBA AKTUÁLNOSTI DAT - „CLOSE TO REAL TIME“
- VYSOKÁ DOSTUPNOST

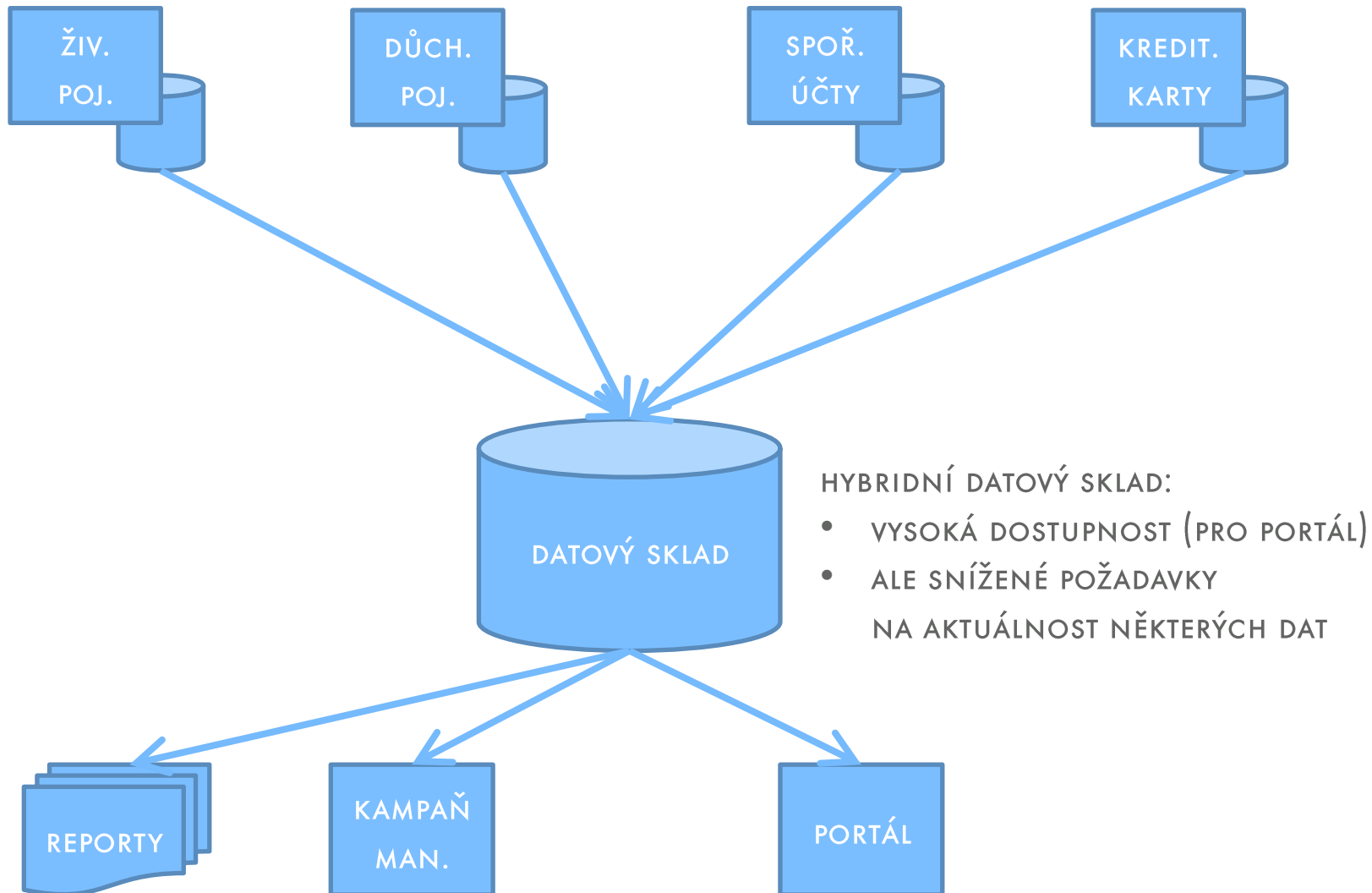
■ FINANČNÍ INSTITUCE

- NABÍZÍ NĚKOLIK FINANČNÍCH PRODUKTŮ: ŽIVOTNÍ, DŮCHODOVÉ, CESTOVNÍ POJIŠTĚNÍ, SPOŘÍCÍ ÚČTY, KREDITNÍ KARTY
- KAŽDÝ PRODUKT VZNIKL V JINÉ DOBĚ (V RŮZNÝCH FIRMÁCH)
- KAŽDÝ PRODUKT MÁ SVŮJ PRIMÁRNÍ / TRANSAKČNÍ SYSTÉM
- KAŽDÝ SYSTÉM BĚŽÍ NA JINÉ PLATFORMĚ (UNIX, AS400, WIN)
- KAŽDÝ SYSTÉM VYUŽÍVÁ JINÉ ÚLOŽIŠTĚ (DB2, ORACLE, ...)
- NĚKTERÁ DATA JSOU UKLÁDÁNA V NESTRUKTUROVANÉ FORMĚ (OPISY SMLUV A DOPISY V ELEKTRONICKÉM ARCHIVU)

■ AKTUÁLNÍ POŽADAVKY JSOU:

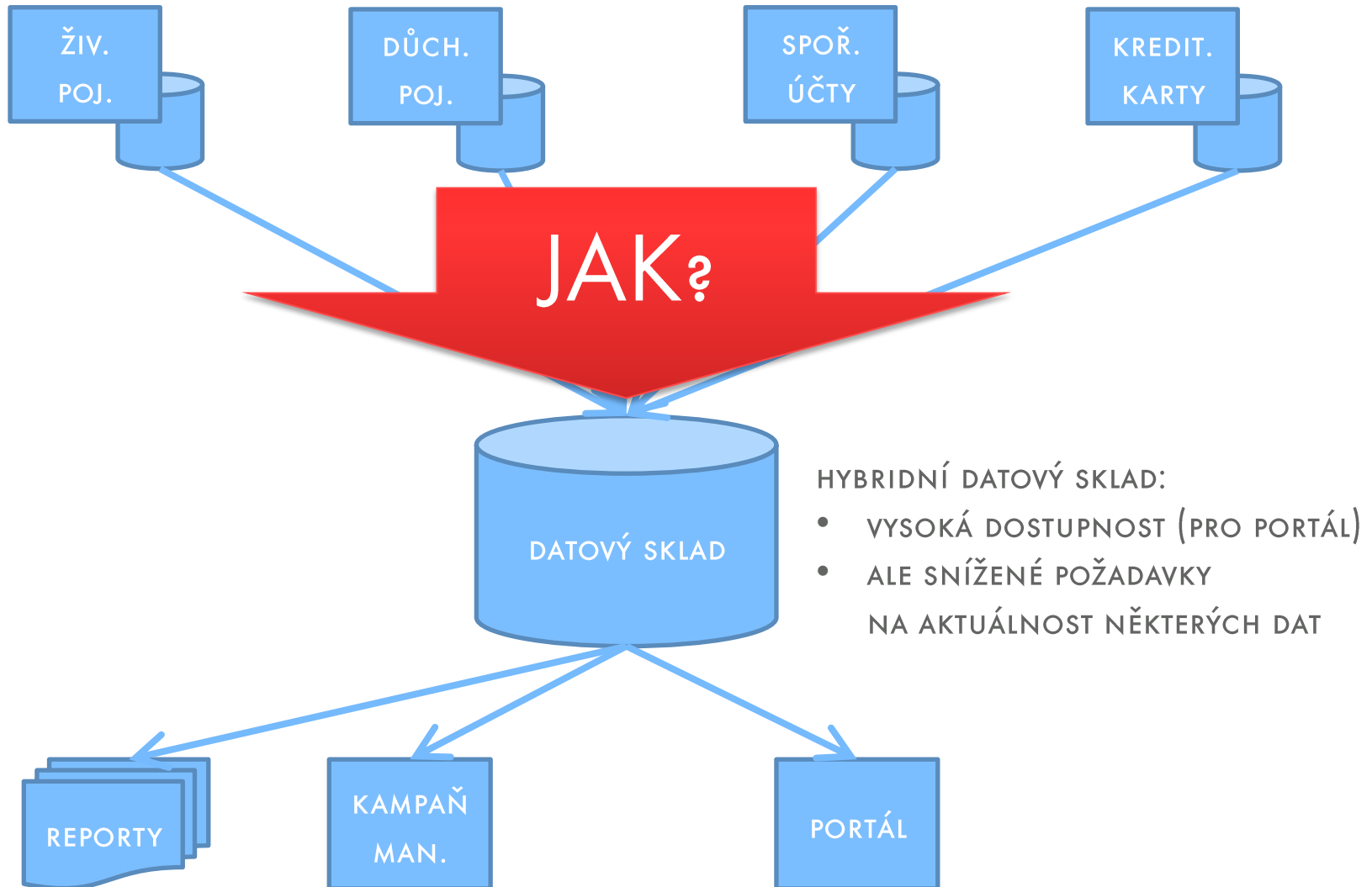
- VYBUDOVAT SYSTÉM NA ŘÍZENÍ KAMPANÍ (VÝBĚRY / CÍLENÍ ZÁKAZNÍKŮ PRO X-SELL/UP-SELL, PODPORA DIRECT MARKETINGU – TELE-MARKETINGU)
- VYBUDOVÁNÍ PORTÁLU PRO ZÁKAZNÍKY SE VŠEMI INFORMACEMI O JEJICH SMLOUVÁCH

CASE STUDY: NÁVRH ŘEŠENÍ



- PROČ DATOVÉ SKLADY VZNIKAJÍ
 - FRAGMENTOVANÉ IT PROSTŘEDÍ VE VĚTŠINĚ FIRMÁCH
 - NEKOMPATIBILNÍ TECHNOLOGIE
 - NESOURODÁ DATA
 - POTŘEBA ZNÁT CELOU HISTORII
(I KDYŽ PRIMÁRNÍ SYSTÉMY JI NEUCHOVÁVAJÍ)
 - POŽADAVKY NA ANONYMIZACI DAT
 - POTŘEBA „VIDĚT FIRMU JAKO CELEK“ A CHÁPAT SOUVISLOSTI

CASE STUDY: VÝZVY



- PŘI PŘENOSU JE TŘEBA VYŘEŠIT:
 - Z DŮVODŮ ZACHOVÁNÍ DOSTUPNOSTI PRIMÁRNÍCH SYSTÉMŮ, NENÍ MOŽNÉ DO NICH PŘISTUPOVAT PŘÍMO (DATABÁZOVÝM LINKEM)
 - VYSOKÉ POŽADAVKY NA ZABEZPEČENÍ VŠECH PŘENOSŮ DAT
 - SYSTÉMY JSOU VE FYZICKY ZNAČNĚ ODLIŠNÝCH LOKALITÁCH
 - DATA VE ZDROJOVÝCH SYSTÉMECH JSOU V RŮZNÝCH FORMÁTECH
 - (FORMÁTY DATUMŮ, KÓDOVÉ STRÁNKY TEXTŮ, FORMÁTY ČÍSEL)
 - SYSTÉMY OBSAHUJÍ „STEJNÁ DATA“
 - V KAŽDÉM SYSTÉMŮ JE „ZÁKAZNÍK“
 - V KAŽDÉM SYSTÉMU JSOU „PRODUKTY“
 - V KAŽDÉM SYSTÉMŮ JSOU „PLATBY“
 - ...
 - SYSTÉMY OBSAHUJÍ NESTEJNOU (NĚKDY ŽÁDNOU) HISTORII
 - DO CÍLOVÉHO ÚLOŽIŠTĚ MUSÍ BÝT DATA ULOŽENA S MAXIMÁLNÍ EFEKTIVITOU (V NORMALIZOVANÉ PODOBĚ A V KRÁTKÉM ČASE)

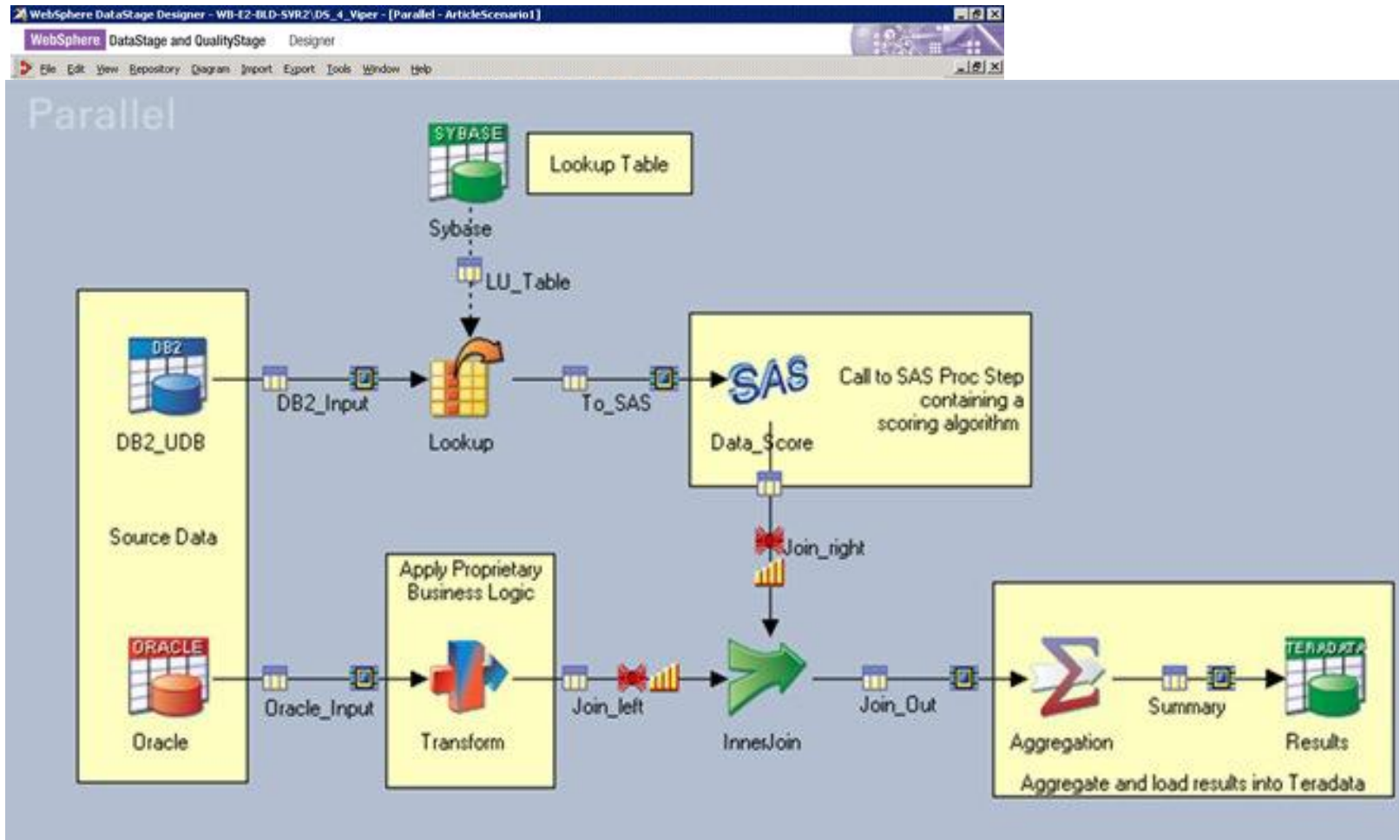
- PŘENOSY DAT Z PRIMÁRNÍCH SYSTÉMŮ DO DATOVÉHO SKLADU SE NAZÝVAJÍ ETL (ELT) PROCESY (JOBS)
- ETL (ELT) PROCES JE POSLOUPNOST KROKŮ (STEPS) ZAJIŠŤUJÍCÍCH:
 - EXTRACTION
 - PŘENOS DAT ZE ZDROJE (S/FTP, SCP, DB LINK / ODBC, SOA, ...)
 - TRANSFORMATION
 - KONVERZE DAT (CSV, XML, DB, ...)
 - NORMALIZACE DAT, GENEROVÁNÍ KLÍČŮ
 - UNIFIKACE (ELIMINACE REDUNDANTNÍCH DAT)
 - HISTORIZACE
 - AGREGACE (PŘI TRANSFORMACÍCH VE VYŠŠÍCH VRSTVÁCH DAT. SKLADŮ)
 - LOADING
 - ULOŽENÍ DO CÍLOVÉHO ÚLOŽIŠTĚ
- ÚKOLEM ETL NÁSTROJŮ JE MAXIMÁLNĚ ZJEDNODUŠIT A ZEFEKTIVNIT IMPLEMENTACI A SOUČASNĚ PROVOZ ETL PROCESŮ

- NEJZNÁMĚJŠÍ ETL NÁSTROJE:
 - IBM INFO SPHERE DATA STAGE A QUALITY STAGE
 - INFORMatica POWER CENTER
 - ORACLE DATA INTEGRATOR (ODI)
 - ORACLE WAREHOUSE BUILDER (OWB)
 - MICROSOFT SQL SERVER INTEGRATION SERVICES (SSIS)
 - ...
- CHARAKTERISTIKA
 - APLIKAČNÍ SERVERY
 - ZPRAVIDLA ODDĚLENÉ OD VLASTNÍHO ÚLOŽIŠTĚ DAT
 - VYSOKÉ NÁROKY NA PROCESSING (VÝKON A PAMĚŤ)

- GUI PRO NÁVRH ETL PROCESŮ
 - PŘEHLEDNÉ GRAFICKÉ PROSTŘEDÍ
 - KAŽDÁ OPERACE JE REPREZENTOVÁNA KOMPONENTOU; MEZI KOMPONENTAMI SE „TAHAJÍ“ DRÁTY; KOMPONENTY MAJÍ ŘADU NASTAVITELNÝCH PARAMETRŮ (JEDNA KOMPONENTA MŮŽE ZAJISTIT NAPŘ. SCD2 HISTORIZACI)
 - MOŽNOST VYTVÁŘET „CUSTOM“ KOMPONENTY
- REPOSITORY
 - OBSAHUJE DETAILNÍ POPIS PROCESŮ (VČ. MAPOVÁNÍ VŠECH ATRIBUTŮ) A STATISTIKY
 - NĚKDY NABÍZÍ API NEBO SAMOSTATNÝ NÁSTROJ NA ANALÝZU (IBM INFOSPHERE METADATA WORKBENCH)
- EXECUTION PLANNING AND MONITORING
 - UMOŽŇUJE DEFINOVAT ZÁVISLOSTI MEZI PROCESY A SPOUŠTĚT JE V DEFINOVANÉM ČASE (DEFINOVÁNÍ HRADEL A SEMAFORŮ)
 - GRAFICKÉ PROSTŘEDÍ UMOŽŇUJÍCÍ SLEDOVAT VÝSLEDKY JEDNOTLIVÝCH PROCESŮ; NOTIFIKACE / ALERTY; SLEDOVÁNÍ KVALITY A SLA
 - SLABÝ ČLÁNEK VĚTŠINY NÁSTROJŮ
 - PŘEVÁŽNĚ SEKVENČNÍ ZPRACOVÁNÍ, MUSÍ ALE UMOŽNIT PARALELNÍ EXEKUCI PROCESŮ; OMEZENÁ MOŽNOST DEFINICE POMOCÍ DATOVÝCH ZÁVISLOSTÍ; NEMOŽNOST SLEDOVAT ZÁVISLOSTI A PROCESY VE VÍCE PROSTŘEDÍCH

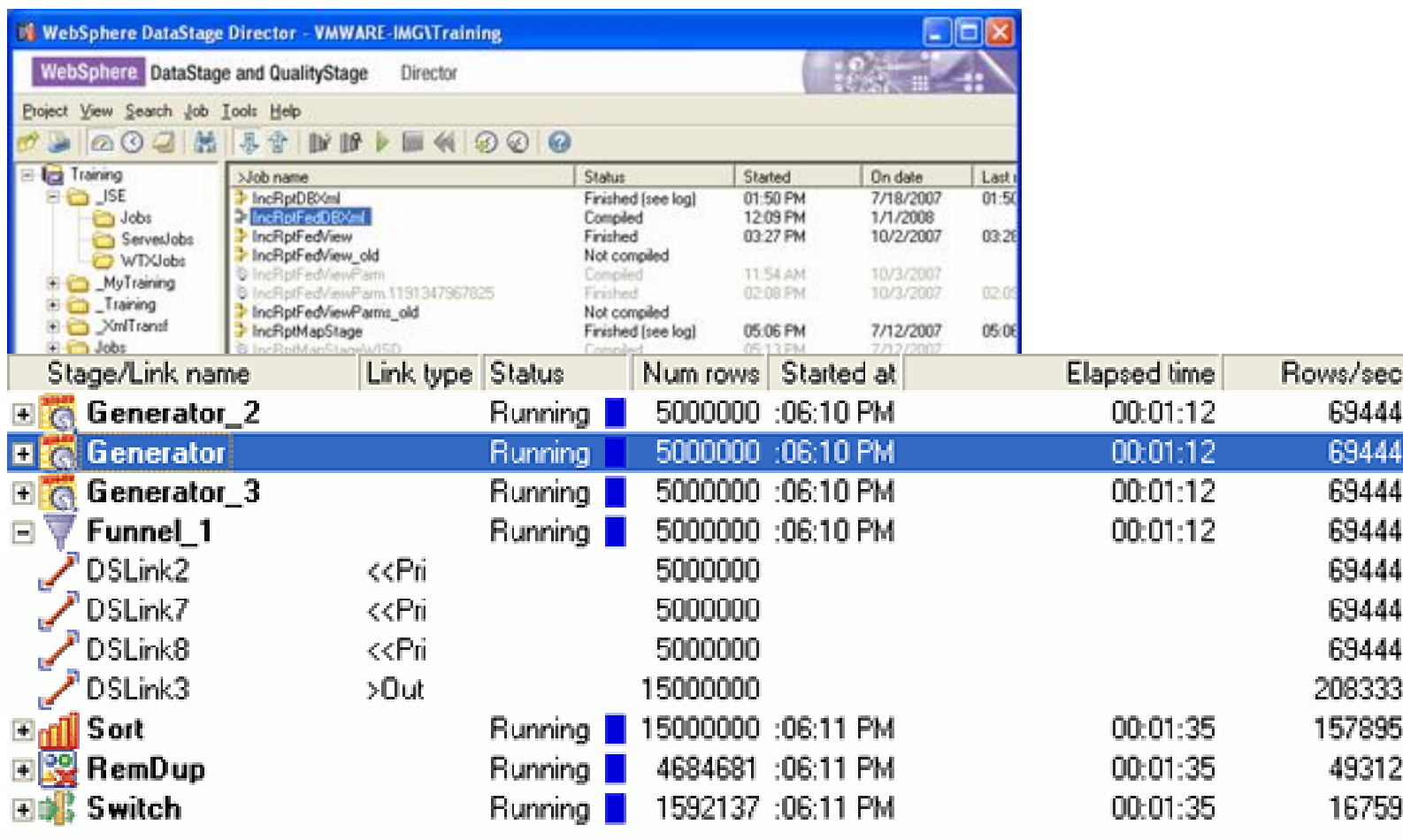
PŘÍKLAD PROCESU

■ IBM INFO SPHERE DATA STAGE



PŘÍKLAD MONITORINGU

■ IBM INFO SPHERE DIRECTOR



The screenshot shows the IBM InfoSphere Director interface. The top window title is "WebSphere DataStage Director - VMWARE-IMGITraining". The main area displays a tree view on the left and a job monitoring table on the right. The table lists various stages and links in a workflow, including Generators, a Funnel, DataStage Links, Sort, RemDup, and Switch stages. The table columns include Stage/Link name, Link type, Status, Num rows, Started at, Elapsed time, and Rows/sec.

Stage/Link name	Link type	Status	Num rows	Started at	Elapsed time	Rows/sec
Generator_2		Running	5000000	:06:10 PM	00:01:12	69444
Generator		Running	5000000	:06:10 PM	00:01:12	69444
Generator_3		Running	5000000	:06:10 PM	00:01:12	69444
Funnel_1		Running	5000000	:06:10 PM	00:01:12	69444
DSLink2	<<Pri		5000000			69444
DSLink7	<<Pri		5000000			69444
DSLink8	<<Pri		5000000			69444
DSLink3	>Out		15000000			208333
Sort		Running	15000000	:06:11 PM	00:01:35	157895
RemDup		Running	4684681	:06:11 PM	00:01:35	49312
Switch		Running	1592137	:06:11 PM	00:01:35	16759

■ PŘÍNOSY

- SNADNÝ VÝVOJ I PRO MÉNĚ ZKUŠENÉ VÝVOJÁŘE NA ZÁKLADĚ MAPOVÁNÍ
- OBSAHUJE PŘEDPŘIPRAVENÉ ČASTO SLOŽITÉ TRANSFORMACE
- MOŽNOST PROVÁDĚT SNADNO IMPACT ANALÝZY – ZAJIŠTĚNÍ LEPŠÍ AUDITOVATELNOSTI A KONTROLY (BEZPEČNOSTI)
- ŠKÁLOVATELNOST PROSTŘEDÍ – PARALELISMUS, PIPELINING
- „SAMODOKUMENTOVATELNOST“

■ NEVÝHODY

- NĚKDY MŮŽE BÝT VÝVOJ DELŠÍ (NUTNOST „KLIKAT“)

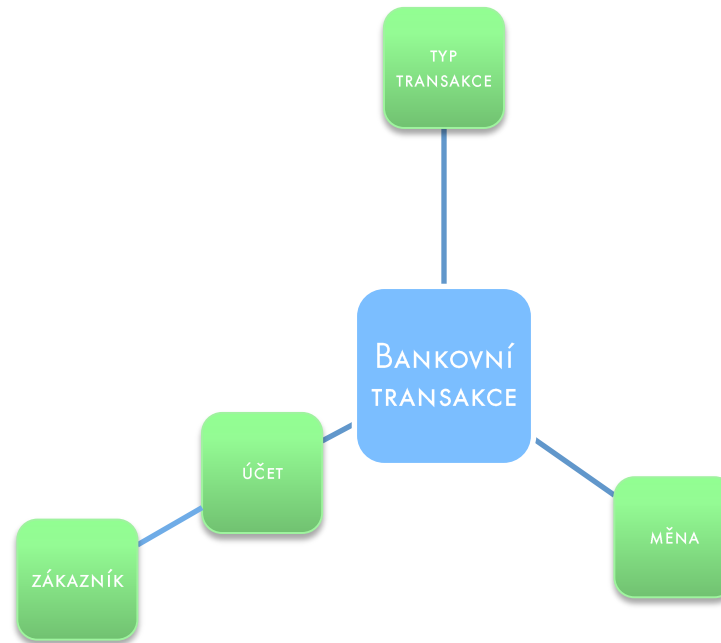
■ OD ETL K ELT: VÍCE-VRSTVÁ ARCHITEKTURA DATOVÝCH SKLADŮ => „IN-DATABASE TRANSFORMATIONS“

- - PLNĚ ZÁVISLÉ NA MOŽNOSTECH A VÝKONU ÚLOŽIŠTĚ (DATABÁZE)
- + MINIMALIZACE PŘENOSŮ DAT

- CUSTOM KOMPONENTA: UNIFIKACE ZÁKAZNÍKA
 - VELICE SPECIFICKÉ POŽADAVKY NA UNIFIKACI (AKTUALIZACE DAT NA ZÁKLADĚ PRIORIT NA ÚROVNI ATRIBUTŮ)
 - PROCES: „DE-UNIFIKACE“
 - POTŘEBA ZACHOVAT JAK PŮVODNÍ DATA, TAK UNIFIKOVANÁ DATA
- PDF PARSING AND GENERATION:
 - NĚKTERÁ DATA PŘICHÁZEJÍ V PODOBĚ VYGENEROVANÝCH PDF
 - METADATA Z TĚCHTO PDF JE TŘEBA PARSOVAT A POUŽÍVAT PŘI ULOŽENÍ DO ELEKTRONICKÉHO ARCHIVU A PRO PROPOJENÍ S OSTATNÍMI DATY
 - Z DAT JE PAK OBČAS TŘEBA ZNOVU VYTVÁŘET DOKUMENTY NA ZÁKLADĚ ŠABLON
- GLOBTECH SCHEDULER
 - ZPRACOVÁNÍ DĚLÁ INFORMATICA, ORACLE PL/SQL SKRIPTY, SSIS

MDM: MASTER DATA MANAGEMENT

- MASTER DATA JSOU KLÍČOVÉ INFORMAČNÍ ENTITY, KTERÉ BY MĚLY BÝT NAPŘÍČ FIRMOU (A SYSTÉMY) INTERPRETOVÁNY JEDNOTNĚ. KAŽDÁ POLOŽKA, KTERÁ JE LOOK-UP (NA KTEROU JE REFERENCE) JE TEORETICKY MASTER DATY.



- V PRAXI JE DEFINICE MASTER DAT VELICE OBTÍŽNÁ.

MDM: MASTER DATA MANAGEMENT

- MASTER DATA TVOŘÍ JEDEN Z PILÍŘŮ DATOVÉ KVALITY.
- MDM JE SOUČÁST DATA GOVERNANCE.
- CÍLEM MDM (JAKO I CELÉ DG) JE ZAJISTIT, KONTROLOVANÉ A KONZISTENTNÍ VYTVÁŘENÍ MASTER DAT A JEJICH KVALITU
- SKUTEČNÉ MDM SE ZPRAVIDLA IMPLEMENTUJE PRO RŮZNÉ ČÍSELNÍKY A PRO KATALOG PRODUKTŮ

DĚKUJEME ZA POZORNOST!



globtech

VÁŠ PARTNER NA CESTĚ OD DAT K INFORMACÍM

BUSINESS INTELLIGENCE DATA INTEGRATION SUCCESS
COMPLETE SOLUTIONS
WAREHOUSING