

NOTATKA WSTĘPNE KONSULTACJE RYNKOWE

„System Hurtownia Danych wraz z narzędziem raportowym jako zadanie w ramach Projektu Transformacji Cyfrowej i Procesowej w obszarze DSM”, znak sprawy: CUI-ZZ.322.1.2021

Data:	22.03.2021r.
Miejsce/forma:	Konsultacje prowadzone za zgodą Stron na łączach video konferencji
Uwagi:	Wstępne konsultacje rynkowe pozwolą na uzyskanie informacji w zakresie najlepszych i najkorzystniejszych: technicznie, technologicznie, organizacyjnie oraz ekonomicznie rozwiązań w zakresie przedmiotu konsultacji. Informacje o minimalnych środkach, które zamawiający podejmie w celu zapobieżenia zakłócenia konkurencji na etapie postępowania: - upublicznienie informacji uzyskanych w trakcie konsultacji - upublicznienie informacji o podmiotach biorących udział w konsultacjach

PORUSZANE ZAGADNIENIA:

L.p.	Poruszane kwestie	Ustalenia
1.	Prezentacja firmy, wdrożenia referencyjne	O firmie: - Firma - zajmują się rozwiązaniami na platformach <i>Big Data</i>, <i>Cloud</i> i <i>Microservices</i>. Modelowanie i zaawansowana analityka danych. Dane trzymają albo na sprzęcie klienta albo w chmurze. - Potencjalny zespół dla pojedynczego projektu składa się z: PM, BA, architekt, lider od <i>Cloud</i>, administrator, zespół dev (ilość zależna od projektu, zazwyczaj 3-4). Wdrożenia referencyjne: - Pierwsza prezentacja HD dla Ubezpieczeń, hurtownia pracująca lokalnie i w chmurze (Azure). Raportowanie: Power BI (Cloud). - Druga prezentacja CLOUD DATA WAREHOUSE (Retail/Gastro) – HD w Cloud z repo w technologii Google BigQuery. Procesy ETL zrealizowane w SnapLogic. Efektem raporty dostępne dla kilku tys. Restauracji / użytkowników. - Trzecia HD (Retail/Motoryzacja) – repozytorium danych w technologii <i>data lake</i> – oparte na OpenSource (Apache Hive) - analizowanie danych z wielu źródeł. - Data Lake dla danych roamingowych – system zarządzania i monitorowania połączeń telefonicznych w roamingu. Wolumen: 300 TB. Okres tworzenia: 6 miesięcy, Zespół 5 osób merytorycznych. - Data Lake (techn. Apache Hadoop) dla prognozowania (energetyka) – efektem było skrócenie czasu przygotowania raportów bilingowych, skrócenie czasu przygotowania predykcji zużycia energii z kilku h na kilka min.

		Wolumen: 100 TB. Zespół – 4 osoby merytoryczne. Bazy danych i analityka danych dla obszaru finansów publicznych. Bazy danych MySQL, Postgres; Zapewnienie poprawności działania baz danych, modele analityczne z wizualizacją danych
2.	Prezentacja systemu	- prezentacja modeli Hurtowni Danych, grafiką <i>Data sources</i> -> <i>ETL</i> -> <i>DW</i>: - model Billa Inmon'a (klasyczny z repozytorium danych detalicznych) - model Ralpa Kimball'a (do szybkiego raportowania, pozbawiony repozytorium) - Big Data: prezentacja architektury <i>Data Lake</i> - prezentacja architektury <i>Lambda</i> (model hybrydowy Hurtowni Danych – Batch Layer z dostępem do danych „on-line” – Speed Layer) - prezentacja schematu <i>Lakehouse</i> (połączenie HD, Speed Layer i Analiz Ad-hoc) - Główne czynniki wpływające na wydajność: - Latencja (opóźnienie) – czas od rozpoczęcia ładowania do udostępnienia danych - wpływ czasu ładowania HD na wymaganą wydajność procesów ETL - odseparowanie infrastruktury systemów źródłowych i Systemu HD - MOLAP, ROLAP i HOLAP – metody przetwarzania analitycznego danych <i>Partycjonowanie i indeksowanie danych</i> – jako metody na skrócenie czasu dostępu do danych <i>ETL – główne funkcjonalności</i>: - Obsługa różnych źródeł danych - Realizowanie złożonych transformacji - Elastyczne harmonogramowanie przetwarzania <i>ETL a ELT – różne konfiguracje procesów</i> - ETL – LIDERZY: - Informatica – najlepszy, ale bardzo drogi, - Oracle - Microsoft Integration Services – popularny, dobry dla średnich hurtowni danych. - Talend – dobra relacja korzyści biznesowych do ceny - SAS - Jakość danych (monitorowanie, zapewnianie jakości) - Wydajność raportowania - Tryby pracy analitycznej - Raporty statyczne – już wygenerowane, - Raporty dynamiczne / odświeżane na moment wywołania – muszą mieć połączenie z bazą danych - Raporty Self Service (np. Tableau, Power BI) – budowane samodzielnie przez użytkownika – także Alteryx, MS Objects, Qlik View. - Analizy wizualne – bazujące na światach obiektów biznesowych. - Technologie Notebook – służą do samodzielnego, swobodnego tworzenia raportów z wykorzystaniem języków programowania. - Metadane - Procesowe Służą do określenia rodzajów, jakości, prawidłowości procesów itp. - Techniczne Służą do Data Lineage (skąd są, z jakiej tabeli, z

		jakiego itp.) - Biznesowe Mówią o tym, z jakimi danymi mamy do czynienia (co mamy w bazach danych) Rola Właściciela Biznesowego: Data Governance – zarządzanie zakresem danych, definicjami zakresów raportowania, udzielanie uprawnień dostępu do danych / raportów, kluczowe pojęcia biznesowe – ustrukturyzowane na potrzeby HD - Platformy analityczne – LIDERZY: - Microsoft (Power BI) - Tableau - Oracle - SAP Business Objects - Qlik - SAS Visual Analytics
3.	Integracja z systemami	- Sposoby integracji danych: - integracja na poziomie infrastruktury - integracja na poziomie danych - integracja wirtualna - model semantyczny - lakehouse
4.	Wymagania techniczne	- Bezpieczeństwo danych – - bezpieczeństwo użytkowania (role użytkowników, obszary raportowe) - bezpieczeństwo techniczne danych (backupy, macierze dyskowe), technika backupu - klastrowanie / redundancja (2 macierze, klastrowanie serwerów itp.) – HA - Disaster Recovery (środowisko DR rozproszone – centra przetwarzania przynajmniej w 2 miejscach) – realizacja projektu dla Deutsche Telekom - SKALOWALNOŚĆ - w klasycznych bazach danych – zarządzanie ilością serwerów w klastrach, macierzami dyskowymi - Liderzy w zarządzaniu Bazami Danych: - Microsoft - Oracle - Amazon Web Services - Google - SAP - LIDERZY - Zarządzanie danymi dla Analityki - Oracle - Microsoft - Amazon Web Services - SAP
5.	Sposób licencjonowania i szacowany czas i koszt wdrożenia	- LICENCJONOWANIE: - Najczęściej Licencje na HD, ETL – per rdzeń procesora, niektóre ETL – per ilość transakcji - Narzędzia raportowe (BI) – per rdzeń lub model subskrypcyjny (ilość użytkowników)
6.	Inne	Harmonogram ramowy: - Analiza - MVP - Agile vs waterfall - Ryzyka - Utrzymanie - Czas wdrożenia. MVP – Minimal Valuable Product Krótkoterminowa realizacja minimalnego produktu –

		jak najszybsze dostarczenie pierwszego „uproszczonego” produktu, aby klient mógł zobaczyć pierwsze wyniki. Dopiero później realizowany jest pełny proces wytwórczy z zachowaniem wszystkich wymaganych elementów (pełne przyrosty, struktury danych, procesy ETL itd.) Cykliczne dostarczanie wartości biznesowej – w poszczególnych przyrostach.
--	--	---

SPORZĄDZILI:

Jacek Tyburczy/Jerzy Sawczyszyn

.....
[imię i nazwisko]