

# Wprowadzenie do Querony



**YOU**NEED **IT**

#Microsoft (Gold, ISV CEE)  
#Hortonworks #Cdata #Microfocus



Kapitał polski, na rynku od 2011 roku

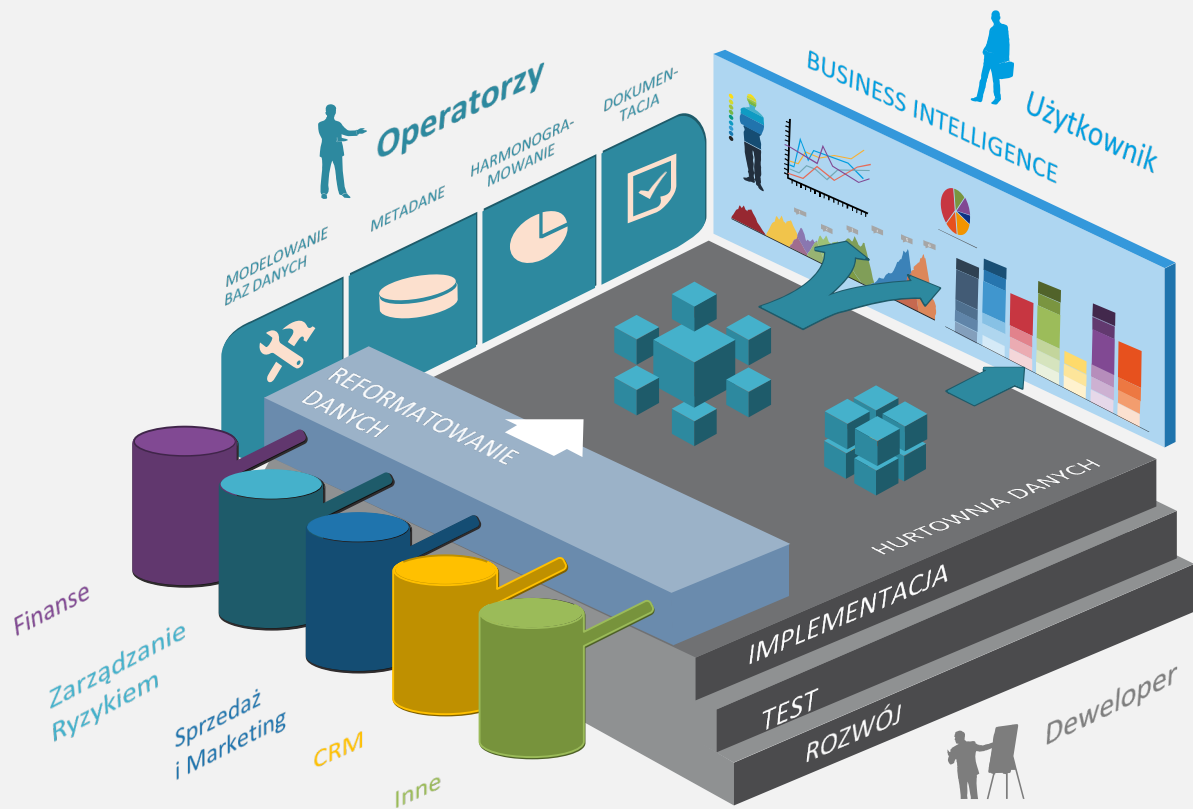
50+ specjalistów tworzących rozwiązania oparte o dane

Klasyczne hurtownie danych, wirtualne hurtownie danych, bigdata, analityka, wizualizacja danych, oprogramowanie na zamówienie

Partnerzy:

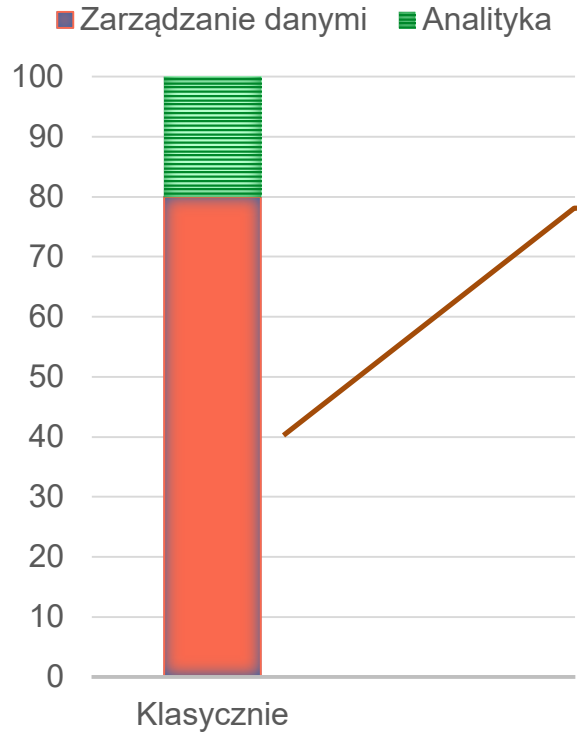


# Przepływ danych w przedsiębiorstwie



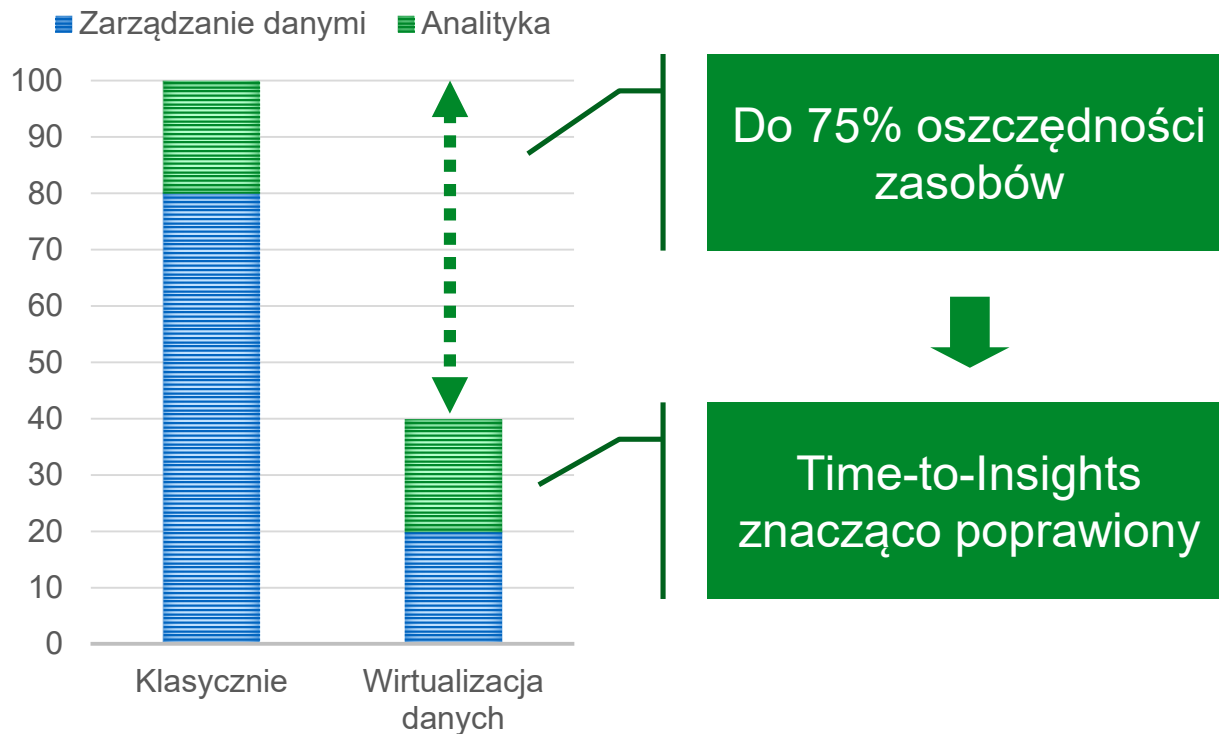
- **Wysokie koszty Hurtowni Danych**  
W klasycznej architekturze wysokie koszty ekstrakcji i konsolidacji danych.
- **Niska Elastyczność**  
Każdorazowe dodanie nowego źródła wymaga pełnego procesu integracji z HD.
- **Wysoka zależność od IT**  
Analitycy BI muszą czekać na zmiany IT zanim zaczną pracę analityczną. BI powinno być zwinnym procesem odkrywczym, który nie powinien dodatkowo obciążać IT.
- **Rosnące wymagania bezpieczeństwa**  
Organizacje muszą na bieżąco sprostać zmieniającym się wymaganiom odnośnie ochrony danych (RODO, UODO).

# Dane są potrzebne wszędzie, ale...



Lwia część zasobów  
zużywana jest na  
integrację i konsolidację  
danych!

# Rozwiązanie: Wirtualizacja danych z Queroną



# Potrzeba podejścia innego niż klasyczne

**Wzrost produktywności** – Wirtualizacja Danych (WD) jest łatwa w użyciu i koncepcyjnie prosta do zrozumienia, co sumarycznie powoduje skrócenie czasu potrzebnego na wdrożenie użytkowników, w szczególności użytkowników biznesowych, oraz odciążenie IT.

**Skrócenie Time-to-Solution i Time-to-Insights** – Projekty oparte o WD trwają krócej, zatem rezultaty biznesowe są dostarczone szybciej. Niższe koszty projektów są dodatkową korzyścią.

**Optymalizacje technologii** – WD pomaga zoptymalizować użycie istniejącej infrastruktury. Zmniejszając zapotrzebowanie na składowanie danych, oszczędności na sprzęcie, utrzymaniu i nadzorze mogą być znaczące.

# Cele wirtualizacji danych

- Zwiększyć **elastyczność i wydajność** tradycyjnych rozwiązań
- Udostępnić **dane szybko**, bez **konieczności ruszania ich z miejsca**
  - Za dużo danych by opłacało się je ruszać z miejsca
  - Okno czasowe jest zbyt wąskie by zdążyć z przeniesieniem danych
  - Dane nie mogą być legalnie skopiowane poza obszar jurysdykcji
- Umożliwić użytkownikom dostęp do **różnych platform przetwarzania danych**
- Zmniejszyć opóźnienia; umożliwić **analitikę „near real-time”**
- **Zmniejszyć redundancję** i skrócić czas przetwarzania danych
- Ułatwić **wdrożenie strategii „polyglot persistence”** (M. Fowler)
  - Użyj najlepszego sposobu magazynowania danych
  - ... i przetwarzania

# Korzyści dla organizacji i użytkowników

**Jeden hub danych**, spójny technicznie i przeszukiwalny

- Meta i meta-meta dane umożliwiają odnalezienie danych w źródłach
- Dostęp do danych poprzez standardowy interfejs SQL

**Demokratyzacja danych i upodmiotowienie konsumentów** danych

- Organizacja może szybciej wejść w trend działań sterowanych danymi

**Integracja i modernizacja** bez rewolucji

- Użytkownik używa swojego ulubionego narzędzia analitycznego, korzystając z danych które normalnie nie byłyby w nim dostępne

Znacząco **polepszone Time-to-Solution i Time-to-Insights**

- podejście zwinne i prototypowanie, ograniczony lub uproszczony ETL
- skalowalność, maksymalne opóźnienie decyzji o materializacji danych

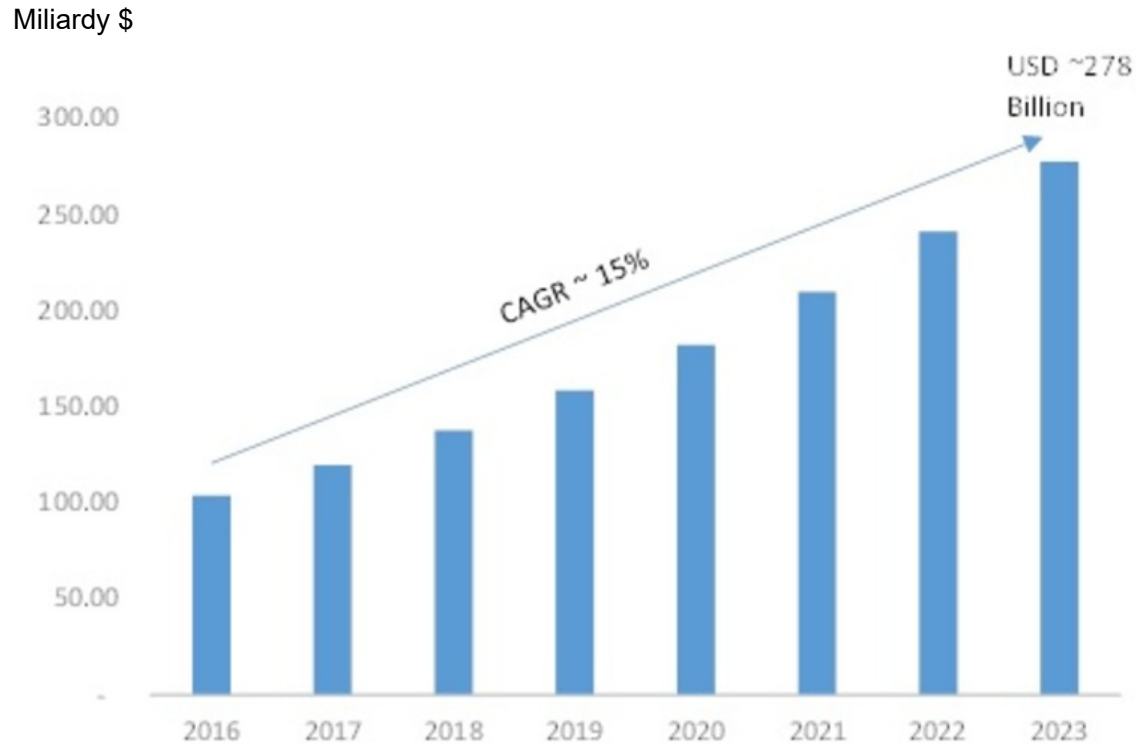
**Krokowa modernizacja** istniejących rozwiązań

- Zwiększenie wydajności raportowania (kolumnowość, MPP)



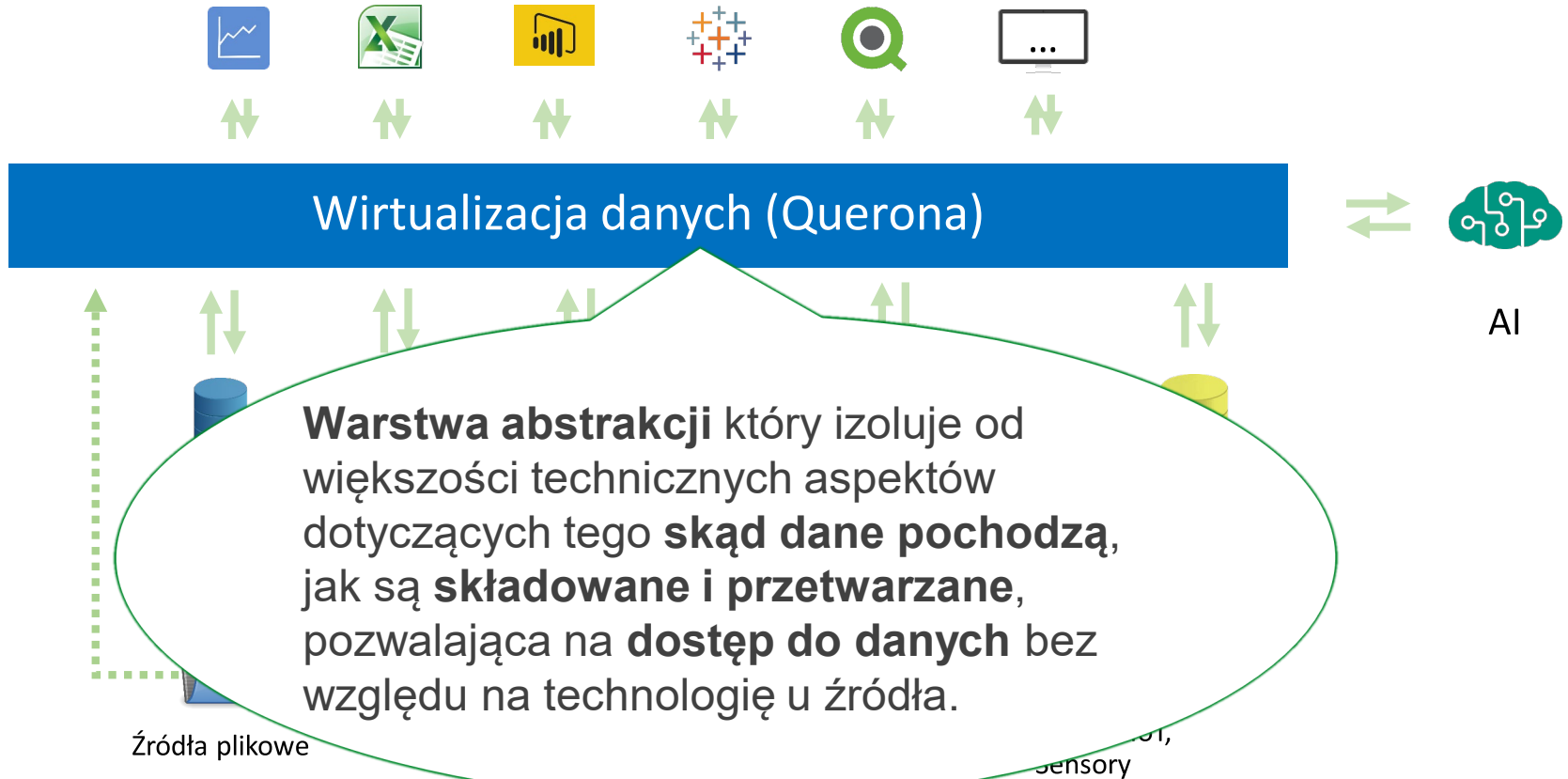


# Rynek wirtualizacji danych

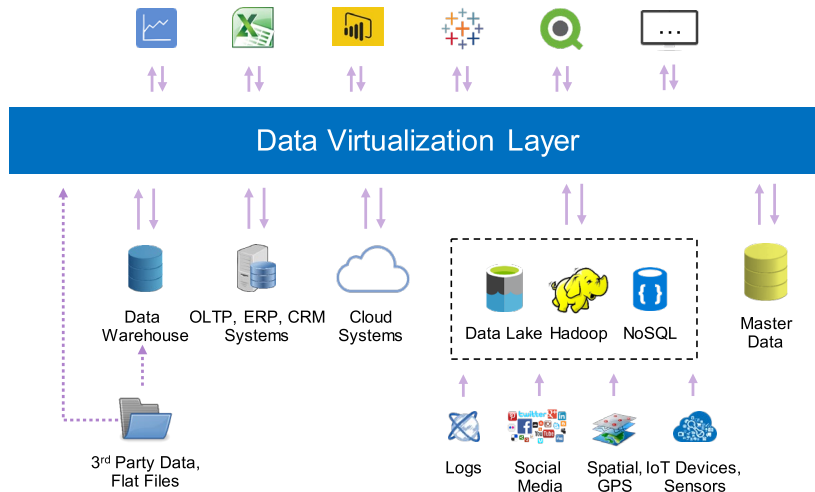


(June, 2018) <https://www.marketresearchfuture.com/reports/data-virtualization-market-4447>

# Architektura z wykorzystaniem wirtualizacji danych



# Kluczowe wartości



## Dane dostępne bez kopiowania

Uniknij zbędnych projektów (ETL) kopiujących dane i skróć projekty analityczne o **75%**



## Połącz ze sobą wszystkie dane

Podłącz **~128** typów źródeł danych i popatrz na dane z pełniejszej perspektywy



## Zarządzanie danymi w trybie samoobsługowych

Analitycy biznesowi mogą osiągnąć **więcej** bez każdorazowej potrzeby angażowania IT

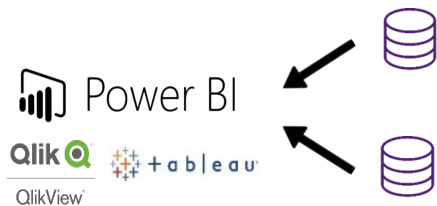
# Kluczowe funkcjonalności i jak je realizujemy

Każde źródło danych	128 typów dedykowanych źródeł danych + wsparcie źródeł generycznych (JDBC, ODBC)
Każdy silnik przetwarzania danych	14 silników przetwarzania danych, które mogą być użyte równocześnie
Każda platforma	Windows (wkrótce Unix/Linux)
Każdy model wdrożeniowy	On-premise, Chmura, Hybryda
Każdy konsument danych	Emulacja Microsoft SQL Server
Każda skala	MPP, Multicloud Silnik BigData Apache Spark wbudowany

# Wirtualizacja Danych polepsza narzędzia Business Intelligence

## FAKTY

Platformy Business Intelligence potrafią łączyć się z różnymi źródłami danych



## PROBLEMY

- Ograniczone możliwości oczyszczania i poprawiania jakości danych
- Joiny pomiędzy wieloma bazami danych są mocno ograniczone lub niemożliwe w trybie rzeczywistym.
- Cała logika zarządzania danymi jest zaszyta w warstwie raportów (brak re-używalności)

## ARCHITEKTURA WIRTUALIZACJI DANYCH

Użycie Querony jako centralnego hub-a danych



## KORZYŚCI

- Tworzenie współdzielonego modelu danych
- Akceleracja zapytań z pomocą kombinacji różnych silników i mechanizmu query rewrite
- Centralizacja ochrony danych
- Uniwersalny interfejs, który jest natywnie wspierany przez narzędzia analityczne i raportowe
- Tryb direct-query zawsze dostępny

# WD pomaga uniknąć budowania klasycznej Hurtowni Danych

## MIT

Tradycyjna, wielowarstwowa hurtownia danych jest konieczna do raportowania.

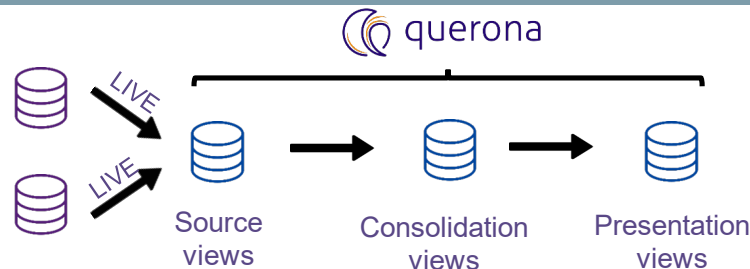


## PROBLEMY

- Tworzenie Raportów jest wstrzymywane tworzeniem hurtowni (6-9 tyg.)
- Konieczna jest implementacja procesu ETL oraz tabel na każdej warstwie.
- Zmiany są kosztowne i zazwyczaj wymagają ponownego ładowania danych

## ARCHITEKTURA QUERONA

Użycie Querony jako Logicznej Hurtowni Danych



## KORZYŚCI

- Materializujemy tylko wybrane widoki w celu optymalizacji wydajności
- Wirtualną hurtownię danych możemy stworzyć szybciej i jest ona prostsza do utrzymania.
- Brak ETL

# Use case: Transparentna modernizacja Hurtowni Danych

Wyzwania:

1. Znacząco **poprawić czasy odpowiedzi zapytań analitycznych**, w szczególności korzystających z zabezpieczenia danych na poziomie wierszy.
2. Zoptymalizować wdrożenia Data Martów **unikając tworzenia kopii i odtwarzania baz danych**



Data  
Marty  
(> 1 TB)



Raportowanie  
Tableau

Ograniczenia:

- **Żadnych zmian we front-end**. Istniejące raporty muszą działać bez zmian w ich implementacji.
- **Enterprise Data Warehouse** musi być nieświadomy jakichkolwiek zmian w martach.
- Dopuszczalne tylko minimalne zmiany w implementacji istniejących widoków na dane.

# Use case: Transparentna modernizacja | Podejście

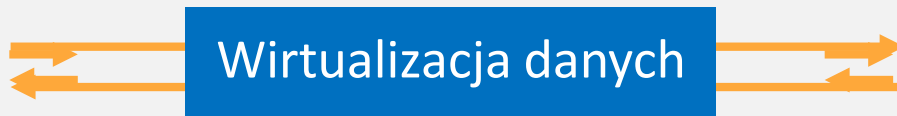
Bez zmian w  
DM lub EDW,  
**bez ETL!**

Querona jako pośrednik  
pomiędzy Data Martami i Raportowaniem  
(transparentne przejście na technologię  
kolumnową i in-memory)

Bez zmian  
w warstwie  
raportowej



Data  
Marty  
(>1 TB)



Raportowanie  
Tableau

alternatywnie

alternatywnie

Chmura

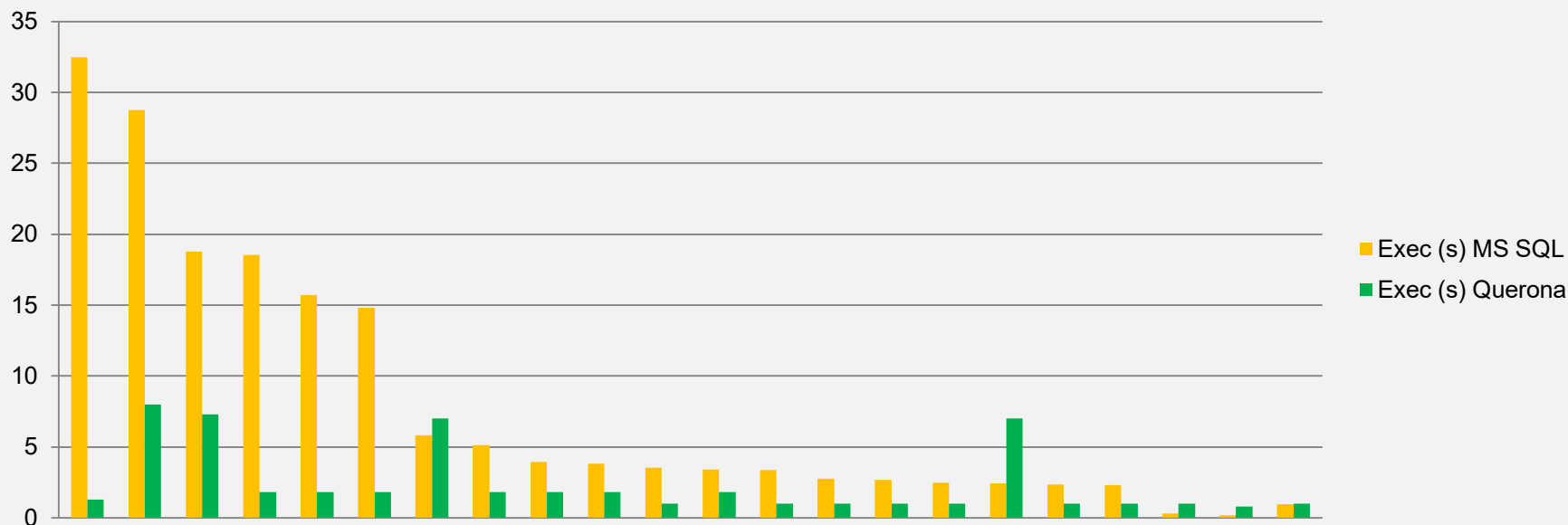


Mix kilku silników  
działających jak jeden  
(spośród 14 dostępnych)



# Use case: Transparentna modernizacja | Rezultaty

Średni przyrost wydajności **+407%** (maks. +2500%); 3 tygodnie wdrożenia



Przetestowane na: 1 x Windows Server, 8 v-rdzeni (2.6 GHz), 32 GB RAM, 400 GB HDD (<< 1TB! :)

Użyte optymalizacje: spychanie predykatów, buforowanie danych bez ETL, preagregaty / surogaty

# Use case | Finanse i bankowość



**Dostęp On-line do zbiorów informacji publicznych** jak: KRS, CEIDG, itp.

Korzyści:

- Oszczędność na rocznych kosztach zakupu informacji, która jest publicznie dostępna
- Skrócenie czasu stworzenia rozwiązania o 70%
- Organizacja zyskała dostęp do aktualnych danych z publicznych rejestrów
- Oszczędności poprzez brak potrzeby kupowania dedykowanej platformy bazy danych (dzięki wbudowanemu Apache Spark)
- Querona pozwala integrować o wiele więcej typów źródeł: (Google Analytics, Facebook, REST, JSON). We can talk to SAS over ODBC.



Google Analytics

{JSON} JSON



REST



Facebook



OData



## Use case | Sektor publiczny – MF



### **Transfer publicznych informacji pomiędzy systemami różnych typów.**

- Przesłanie kilka miliardów rzędów z bazy DB2, Spark on Hadoop, Oracle do bazy docelowej, z użyciem pseudonimizacji danych w locie

#### Korzyści:

- Byliśmy w stanie utrzymać KPI transferu danych na poziomie natywnego systemu DB2 przy jednocześnie dołożonej pseudonimizacji w locie.
- Prosty, webowy interfejs użytkownika pozwolił klientowi skonfigurować rozwiązanie we własnym zakresie.

# Use case | System zarządzania zdarzeniami marketingowymi w sektorze bankowym



## Marketing oparty o zdarzenia.

### Korzyści:

- Oferta wysyłana do klienta w czasie rzeczywistym w oparciu o konkretne zdarzenia (np. nieudana próba wypłaty środków – przekroczony limit).
- Kampanie marketingowe w czasie rzeczywistym
- Prosta integracja z bazami danych wewnątrz banku – analiza ryzyka klienta, czarna lista, baza zgód marketingowych.
- Integracja z systemem zarządzającym zdarzeniami (Magistrala IBM Enterprise, Kafka)

# Use case | Branża Farmaceutyczna



**Modernizacja** istniejącej **Hurtowni Danych** klienta, zawierającej terabajty danych, opartej o Microsoft SQL Server i Tableau Server.

Korzyści:

- **Brak potrzeby programowania** po stronie warstwy backend i frontend
- Poprawa wydajności rozwiązania o średnio **+407%** na 21 raportach analitycznych
- Brak dodatkowych kosztów związanych z inwestycjami w infrastrukturę

**Techniki optymalizacyjne:**

- Push-down predykatów
- Cache danych – konfigurowany poprzez harmonogram zadań
- Join na pojedynczym źródle
- Pre-agregaty

# Use case | Edukacja



Uniwersytet Michigan wybrał narzędzie Querona jako **preferowany system** przy nauczaniu programu **Data Science and Analytics**

Korzyści:

- **Prostota integracji** wielu źródeł danych używanych na uczelni z systemem chmury Google
- **Niższe koszty** rozwiązania w porównaniu z konkurencją
- **Krótki czas** potrzebny na migrację obecnych rozwiązań do systemu Querona
- Uniwersytet Michigan jest w czołówce rankingów, działa od 1926, a od 1996 stał się jedną z pierwszych szkół z programami online w Stanach Zjednoczonych.
- Pracujemy nad wewnętrznymi artykułami uniwersytetu wypierając Denodo przez Queronę w programie dydaktycznym.

## Use case | Retail oraz e-commerce



**Dostęp On-line do stanów magazynowych** we wszystkich sklepach sieci, w tym e-sklepie

Korzyści:

- Rozwiązanie zbudowane o **~70%** szybciej niż w przypadku klasycznego podejścia
- Zespół Analityczny ma dostęp do bieżącej informacji ze sklepów sieci
- **Oszczędność poprzez zmniejszenie kosztów** w porównaniu do dedykowanej platformy bazodanowej, dzięki wbudowanemu Apache Spark

## Use case | Retail oraz e-commerce



Wirtualna Hurtownia Danch stworzona w **2 miesiące**:

- Własne rozwiązanie e-sklepu,
- Google Analytics,
- ERP,
- Finanse

Korzyści:

- **Skrócony** o **~60%** czas stworzenia pełnego rozwiązania
- Zespół analityczny zarządza nowymi źródłami danych w trybie **self-service**
- **Oszczędność poprzez zmniejszenie kosztów** w porównaniu do dedykowanej platformy bazodanowej, dzięki wbudowanemu Apache Spark



# Demo



**YOU**NEED **IT**



## REJESTRACJA ŹRÓDEŁ

~128 typów źródeł danych



## KLASYFIKACJA METADANYCH

Zdefiniowanie tagów, przypisanie do obiektów



## ZABEZPIECZ. DANYCH

Zintegrowane uwierzytelnienie, maskowanie danych, ...



## KONSUMPCJA DANYCH

Standardowe połączenie jak do MS SQL Server

Source object

- NorthwindOracleTested
  - "NorthwindOracleTested"."Categories"
  - "NorthwindOracleTested"."Customers"
  - "NorthwindOracleTested"."Employees"
  - "NorthwindOracleTested"."IntegrationTestsDatabaseHash"
  - "NorthwindOracleTested"."Order Details"
  - "NorthwindOracleTested"."Orders"
  - "NorthwindOracleTested"."Products"
  - "NorthwindOracleTested"."Region"
  - "NorthwindOracleTested"."Shippers"
  - "NorthwindOracleTested"."Suppliers"
- SCOTT
  - "SCOTT"."BONUS"
  - "SCOTT"."DEPT"
  - "SCOTT"."EMP"
  - "SCOTT"."SALGRADE"

Results: 3

First name x

[AdventureWorksDW2012].[dbo].[DimCustomer].[FirstName]

[No comment]

[ContosoRetailDW].[dbo].[DimCustomer].[FirstName]

[No comment]

Privilege

	Allow	Grant	Deny
Alter	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Alter view	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Control	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Control view	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Delete	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Insert	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Select	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Update	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
View definition	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

	A	B	C
1	CustomerKey	GeographyKey	CustomerAlternateKey
2	11000	26	AW00011000
3	11001	37	AW00011001
4	11002	31	AW00011002
5	11003	11	AW00011003
6	11004	19	AW00011004
7	11005	22	AW00011005
8	11006	8	AW00011006
9	11007	40	AW00011007
10	11008	32	AW00011008
11	11009	25	AW00011009
12	11010	22	AW00011010

Ekstrakcja wybranych metadanych z systemów źródłowych.

Tabele i kolumny wybrane do importu metadanych.

Zmiany, komentarze i etykietowanie metadanych do wyszukiwania.

Katalog etykiet jest zarządzany przez Klienta.

Wprowadzenie zarządzanego modelu bezpieczeństwa kompatybilnego z Microsoft SQL Server.

Opcjonalne wprowadzenie zabezpieczeń na poziomie wierszy i maskowania danych.

Konsumpcja danych przy użyciu standardowych sterowników do Microsoft SQL Server dostępnych we wszystkich narzędziach.

Zapytanie realizowane w dialekcie T-SQL.

# Kluczowe Korzyści

- Pełne rozwiązanie np. hurtowni danych można stworzyć do **75% szybciej**
- **Współtworzenie modelu danych** – demokratyzacja wiedzy biznesowej
- **Brak ograniczenia do jednego producenta**, łatwość łączenia różnych technologii pozwala znaleźć najwydajniejsze rozwiązanie
- Wydajność, bezpieczeństwo i elastyczność w oparciu o **heterogeniczne źródła danych**
- **Prosty proces wdrożenia**: zerowa potrzeba konfiguracji poprzez pełną zgodność standardu z Microsoft SQL
- **Platforma BigData jest wbudowana** oraz gotowa do użycia (Apache Spark)

# Potwierdzenie wartości biznesowej

**1** IDENTYFIKACJA WYZWAŃ

Identyfikujemy obszary które są wyzwaniem

Wybieram obszar z możliwie najlepszym potencjałem poprawy

**2** USE CASE + WARSZTAT

Querona stara się jak najlepiej poznać organizację i procesy

Definiujemy proces i jego wyzwania

Określamy zakres i wymagania techniczne

**3** ANALIZA

Określamy zakres i wymagania dla POV

Weryfikujemy dostępne dane

Określamy miary dla przypadku biznesowego

Określamy mapę realizacji POV

**4** POTWIERDZENIE WARTOŚCI (POV)

Pomagamy klientom przygotować środowisko

Wdrażamy Queronę

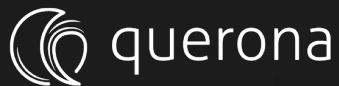
Warsztat celem szkolenia klienta jak używać Querona do budowania rozwiązań

Weryfikujemy rozwiązanie i porównujemy rezultaty



Data management easier  
Actionable results faster

[info@querona.io](mailto:info@querona.io) | [www.querona.io](http://www.querona.io)

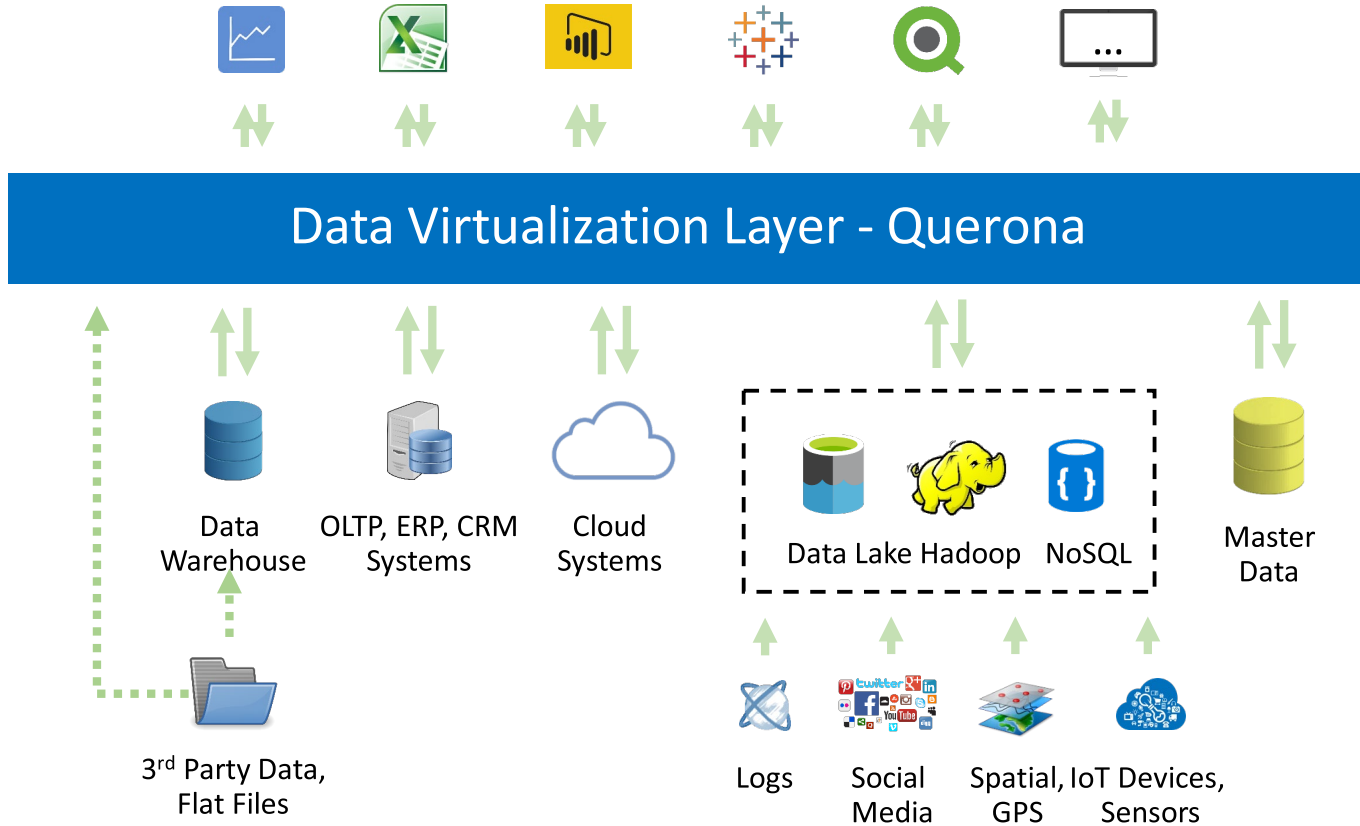


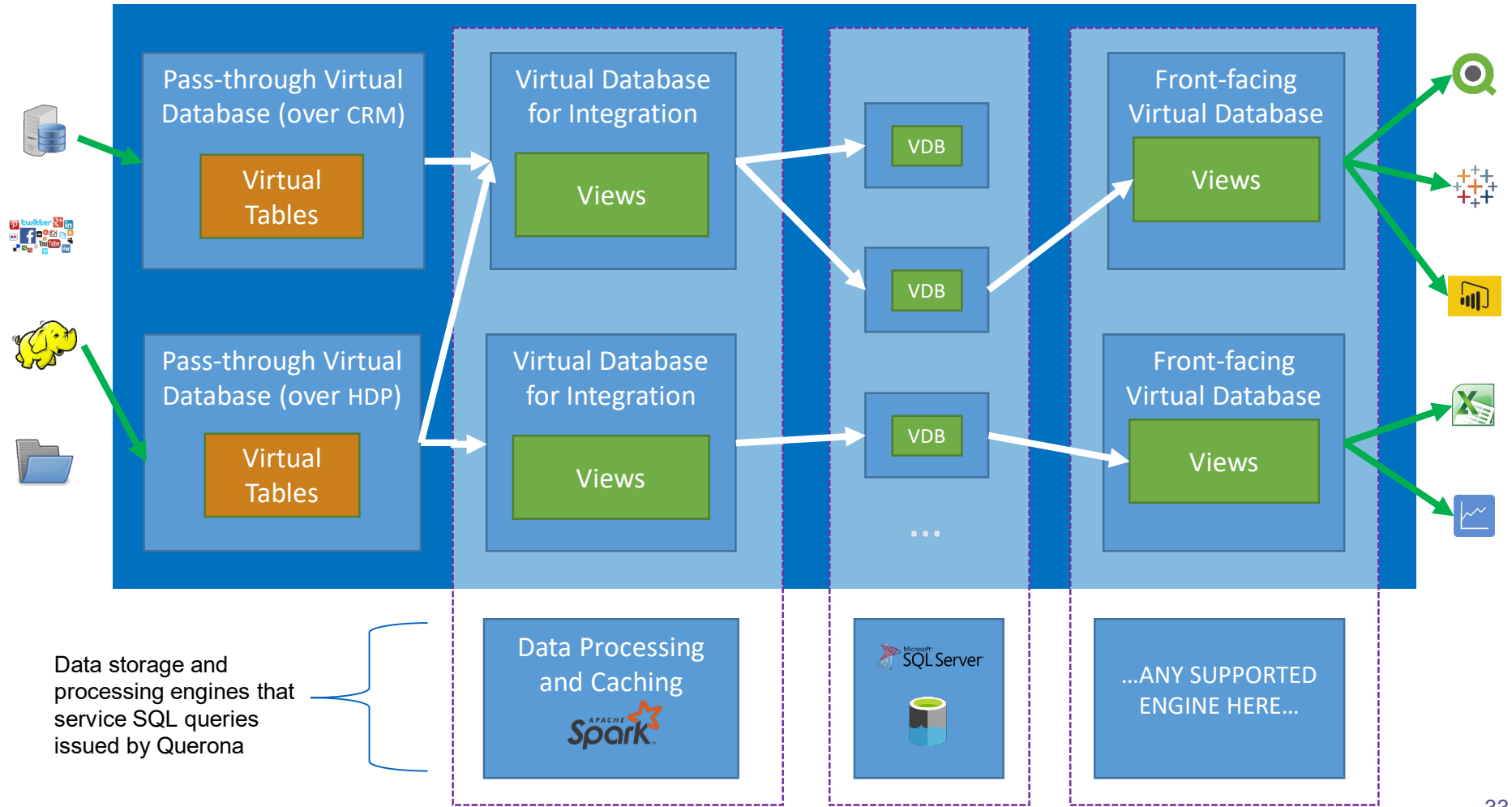
# Bonus slides | Technology



**YOU**NEED **IT**

# Architecture overview







## CRM & Marketing Automation



Salesforce.com



Microsoft Dynamics 365



Dynamics 365 Sales (CRM)



Microsoft Dynamics CRM (On-...



NetSuite CRM



SugarCRM



SuiteCRM



Highrise



Zoho CRM



Act CRM



Bullhorn CRM



Oracle Sales Cloud



Veeva



Salesforce Marketing Cloud



Google Analytics



DoubleClick



DoubleClick For Publishers



Google AdWords



Bing Ads



MailChimp



Oracle Eloqua



HubSpot



Marketo



SendGrid



YouTube





YouTube Analytics




Magento

## Collaboration & ERP

 Microsoft SharePoint (On-pre...

 Excel Online

 Excel Online

 SharePoint Excel Services

 Google Apps

 Office 365

 Basecamp

 Microsoft Project

 Smartsheet.com

 SurveyMonkey

 Magento

 Dynamics 365 Sales (CRM)

 Dynamics 365 Business Centra...

 Dynamics 365 Finance & Oper...

 Microsoft Dynamics GP

 Microsoft Dynamics NAV


 Microsoft Dynamics AX

 NetSuite ERP

 SAP NetWeaver

 Acumatica ERP

 odoo ERP

 Google Directory

 ServiceNow

 Splunk

 Basecamp

 GitHub

 Jira

## NoSQL & Big Data



DynamoDB



MongoDB



Cosmos DB



Google BigQuery



Amazon Athena



Azure Table



IBM Cloudant



Cassandra



Couchbase



HBase



Elasticsearch



Apache Hive



HPCC Systems



Apache Spark



DynamoDB



Google BigQuery



Redis



Amazon Redshift



Amazon Athena



Cassandra



1010data



HPCC Systems



SQL Analysis Services

## Accounting Systems



Microsoft Dynamics GP



QuickBooks



QuickBooks Online



QuickBooks Point-Of-Sale



Sage 50



Sage 50 UK



Xero Accounting



Exact Online



Intacct



FreshBooks



SAP Concur



Reckon

## REST Services



OData

{JSON} JSON



REST

## Document & File Formats



Excel Files



Excel Online



SharePoint Excel Services



Google Spreadsheets



CSV/TSV



XML

## Social Networking



Salesforce Chatter



Twitter



Facebook



LinkedIn



Slack

## Networking & Auth



OData

{JSON} JSON

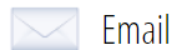


REST



LDAP

OFX Transactions



Email

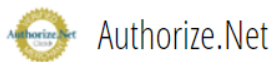


RSS



FTP

## Financial & E-Commerce



Authorize.Net



Square

eBay



PayPal



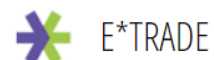
Shopify



Stripe



OFX Transactions



E\*TRADE



OpenExchangeRate



Quandl

## Uncategorized Sources



Twilio



Google Calendar



Google Contacts



Google Contacts



Google Gmail



Azure Management



Bing Search



Salesforce Einstein



Microsoft Exchange



Google Search