

Potencjał Big Data w elektroenergetyce

Anna Gorczyca-Goraj – Uniwersytet Śląski, Instytut Fizyki, PSE Innowacje

Streszczenie

System elektroenergetyczny podlega obecnie fundamentalnym zmianom. Funkcjonujące do tej pory paradygmaty jego działania wymagają redefinicji, pociągając za sobą konieczność dynamicznego realizowania przez operatorów sieci przesyłowych i dystrybucyjnych szeregu zmian dotyczących wielu aspektów związanych z pracą i sterowaniem w systemie elektroenergetycznym. Zmiany te związane są z postępującą decentralizacją, gdzie udział źródeł rozproszonych rośnie coraz silniej. Dodatkowo, świadomość odbiorców energii elektrycznej wzmacnia się i ich udział jako prosumentów również stanowi impuls do zachodzących w elektroenergetyce zmian. Sterowanie pracą systemu bazujące na odczytach z układów automatycznego opomiarowania, wraz z instalacją inteligentnych liczników, powodują, że zarówno napływ rejestrowanych danych, jak i towarzyszące mu operacje przetwarzania tych danych, mają charakter masowy. Dlatego też termin Big Data jest jednym ze słów kluczowych towarzyszących transformacji systemu elektroenergetycznego między jego czysto fizyczną postacią, a docelową, cyber-fizyczną.

1. Wstęp

Ilość wytwarzanych w różnych sektorach danych rośnie w ogromnym tempie, z czego ostatnie lata przynoszą prawdziwą lawinę napływających zewsząd rejestrowanych danych. Należy mieć jednak świadomość, że sam termin, jaki używany jest obecnie powszechnie do określania dużych zbiorów danych, czyli Big Data jest już znany od 2005 roku, kiedy został pierwszy raz użyty przez Rogera Magoulasa z O'Reilly Media. Termin został wprowadzony do określenia szerokiej klasy zbiorów danych, tak dużych, że niemożliwe stało się zarządzanie zgromadzonymi danymi czy ich przetwarzanie, z wykorzystaniem dostępnych „tradycyjnie” narzędzi. Ta niemożliwość związana jest nie tylko z ilością zgromadzonych danych, ale także z ich złożonością. Dlatego w uproszczeniu można stwierdzić, że Big Data są to dane zbyt duże, lub zmieniające się zbyt szybko, żeby możliwe było ich dopasowanie do zwykłych struktur bazodanowych.

Termin Big Data już dawno przestał być modnym i chwytliwym powiedzonkiem, czymś co po angielsku określa się jako „*buzzword*”. Takim był w sektorze IT w 2012 roku, ale od tego czasu, wyewoluował w stronę kompleksowego podejścia do pracy z danymi o ogromnym rozmiarze, dużej szybkości napływu i wielkiej różnorodności. Podejście to ma na celu wyszukanie istniejących w morzu danych wzorców i informacji, które wcześniej ukryte były ze względu na ilości czasu koniecznego do przetworzenia danych. Obecnie dzięki dostępności odpowiedniej wydajności sprzętu, a także przeniesieniu części infrastruktury do rozwiązań chmurowych i darmowemu oprogramowaniu *open-source*, które dedykowane jest przetwarzaniu Big Data, technologia jest już dostępna nawet dla niewielkich projektów start up-owych, w których możliwe jest niedrogo wykonywanie obliczeń w chmurze.

Wartość Big Data dla biznesu dzieli się na dwie kategorie. Jedna wynika z analitycznego wykorzystania danych, wyciągnięcia wniosków z pewnej historii zarejestrowanej w nich, a druga daje firmom możliwość proponowania nowych produktów i usług. Pomysły na nie wywodzić się mogą ze znalezionych, często zaskakujących, wzorców ukrytych w Big Data. Wykorzystanie Big Data, które ma zakończyć się dla firmy sukcesem i rozwojem, wymaga od zespołu Data Center ciekawości, jak również przedsiębiorczego spojrzenia na dane i wyniki ich analiz.

Należy sobie teraz zadać definicyjne pytanie o to, czym są dane charakteryzowane terminem Big Data. Sam termin może wydawać się odrobinę mglisty, w podobny sposób jak obecnie powszechnie dostępne rozwiązania chmurowe, obejmujące szereg różnych technologii. Danymi wejściowymi dla systemów informatycznych dedykowanych Big Data mogą być czaty z mediów społecznościowych, logi z serwerów, odczyty z czujników ruchu ulicznego, obrazy satelitarne, transakcje bankowe, nagrania radiowe, pliki muzyczne MP3, skany dokumentów rządowych, dane GPS, dane pochodzące z giełdy itd. Taka lista może być nadal rozwijana. Jak łatwo zauważyć, wymienione źródła Big Data są bardzo zróżnicowane, ale z pewnością są wspólne cechy, które pozwalają je skategoryzować w ramach tego terminu. Początkowo definicja Big Data obejmowała 3 V od angielskich słów *volume*, *velocity*, *variety*. Obecnie używa się jeszcze jednego określenia, mianowicie *veracity*.

Znaczenie poszczególnych V jest następujące:

- **Volume** – odnosi się do ilości danych określanych jako Big Data.
Modele matematyczne konstruowane podczas przetwarzania danych na potrzeby analityki, wnioskowania czy też prognozowania mogą przynieść doskonałe wyniki, jeżeli działać będą na ogromnych zbiorach treningowych. Należy w tym miejscu podkreślić jednak, że takie wolumeny nie mogą być przechowywane w relacyjnych bazach danych. Architektura Big Data wymagana jest w przypadku regularnego lub wręcz ciągłego przetwarzania tera- lub petabajtów danych. W przypadku danych obsługiwanych na komputerach osobistych, rzadko przekraczających 100 GB, wystarczy serwer skalowalny; elastyczny, taki, który działając w chmurze obliczeniowej pozwala na dynamiczne, automatyczne zwiększanie zasobów, ale również ich obniżenie w razie potrzeby. Jednakże w przypadku ilości danych tera- lub petabajtowych, wymagane jest stosowanie skalowalnych rozwiązań opartych na innych paradygmatach. Przykładowo, Apache Hadoop w języku Java jest platformą *open-source* bazującą na klastrach składających się z szeregu indywidualnych komputerów.
- **Velocity** – oznacza szybkość napływu danych do bazy, w której mają być przechowywane.
Przykładem mogą być firmy z sektora finansowego bazujące na analizach danych giełdowych. Ale nie tylko. Szybkość dostarczanych danych wynika obecnie przede wszystkim z szerokiego wachlarza usług internetowych, w których dane pochodzące od klienta wracają do dostawcy i podlegają analizie, żeby zaproponować nowe, lepsze usługi i produkty możliwie jak najszybciej, wyprzedzając konkurencję. Każde kliknięcie w aplikacji internetowej jest rejestrowane i przetwarzane. Dodatkowo, źródłem danych takich jak zdjęcia geolokalizacyjne czy nagrania audio są smartfony. Należy w tym miejscu podkreślić, że sam termin *velocity* w odniesieniu do Big Data również uległ ewolucji. Początkowo odnosił się do szybkości napływających danych, co szybko przestało być problematyczne. Możliwe było gromadzenie spływających do bazy ogromnych ilości danych w odpowiednim tempie, na potrzeby przyszłego przetwarzania. Waga terminu *velocity* wiąże się dziś również z szybkością otrzymania odpowiedzi o wynikach analiz danych, czyli szybkością między otrzymaniem danych wejściowych, a podjęciem decyzji w oparciu o nie, co wpływa wprost na konkurencyjność produktów czy usług.
- **Variety** – oznacza szeroko rozumianą różnorodność danych.
Najczęstszą sytuacją w przypadku zbiorów Big Data jest przetwarzanie danych nieustrukturyzowanych, w którym pierwszym etapem jest proces ich strukturyzacji na potrzeby pracy analityków, bądź działania odpowiednich algorytmów przetwarzania przygotowanych danych. Oznacza to, że dane zasilające bazę są różnorodne, gdyż pochodzą przykładowo z mediów społecznościowych, część stanowi zdjęcia, a dodatkowo część to surowe odczyty z liczników pomiarowych. Te dane zasilic mają jedno repozytorium, co od razu sugeruje, że relacyjne bazy danych nie znajdują tu zastosowania. Dla celów biznesowych ważne jest przechowywanie wszystkich typów danych. Pomysł na ich wykorzystanie może przyjść z czasem, a dane utracone lub nie pozyskane, mogą przynieść szkodę z punktu widzenia rozwoju przedsiębiorstwa. Przechowywanie tak zróżnicowanych danych wpływa na architekturę konkretnego rozwiązania bazodanowego i poświęcona zostanie temu zagadnieniu osobna sekcja niniejszego artykułu.
- **Veracity** – wiarygodność danych. Jest to cecha zbiorów Big Data przekładająca się wprost na ich jakość. Dane o wysokim stopniu prawdziwości są danymi wysokiej jakości, czyli takimi których analizy przyniosą wymierne korzyści biznesowe. Zbiory o niskiej prawdziwości zawierają dane zaszumione, spośród których wyekstrahowanie cennych informacji analitycznych może okazać się problematyczne. Przykładem danych o wysokiej prawdziwości są dane pochodzące z eksperymentów lub prób medycznych.

Jednakże jak wynika z literatury, nie ma jednoznacznej definicji terminu Big Data, ani też polskiego odpowiednika, który dobrze oddawałby znaczenie. O ile początki określenia Big Data sięgają tylko znaczenia zbiorów danych, które nie były przechowywane w relacyjnych bazach danych, o tyle obecnie, termin ten jest bardzo pojemny, ponieważ niektórzy autorzy zaliczają do niego również narzędzia analityczne, czyli techniki Data Mining. Dodatkowo, poza wymienionymi powyżej 4 V określającymi podstawowe cechy wielkich zbiorów danych, niektórzy autorzy dodają też kolejne. Mianowicie, jedna z czołowych firm światowych w dziedzinie nowoczesnych technologii, Gartner Group, wskazuje dwie dodatkowe charakterystyki: *variability*, czyli zmienność, oraz *complexity* – złożoność. Te dwie cechy dodaje do definicji Big Data również SAS (Analytics Software and Solutions). Wymiarom zbiorów Big Data dodanym przez Qlink Tech, coraz bardziej istotnym przy obserwowanym zalewie danymi, jest *vagueness*, czyli trudność w zidentyfikowaniu danych na potrzeby analizy konkretnego problemu, ich nieokreśloność czy nieostrość.

2. Architektura Big Data – krótka charakterystyka logiczna

Architektura Big Data stworzona została do obsługi, wprowadzania, przetwarzania i analizy danych, które są zbyt duże, lub zbyt złożone dla tradycyjnych systemów baz danych. Konkretnie rozwiązanie powinno zawierać co najmniej jeden z następujących komponentów:

- przetwarzanie wsadowe źródeł Big Data w stanie spoczynku,
- przetwarzanie w czasie rzeczywistym Big Data w ruchu,
- interakcyjna eksploracja danych,
- analiza predykcyjna i uczenie maszynowe.

Z punktu widzenia zastosowań biznesowych, zwłaszcza w elektroenergetyce, wśród sytuacji, kiedy wymagane jest zastosowanie architektury Big Data, najbardziej istotne są:

- gromadzenie danych (np. pomiarowych) z rozległej sieci,
- ilości gromadzonych danych przekraczają porcje 100 GB, a procesy przetwarzania tych danych trwają dłużej niż 8 godzin,
- dane gromadzone są w postaci nieustrukturyzowanej, ale ustrukturyzowania wymagają ze względu na ich dalszą analizę,
- analityka Big Data motywowana jest celami biznesowymi OSP i OSD, jak również zachowaniem bezpieczeństwa pracy sieci.

Wśród poziomów logicznych architektury Big Data, można wymienić następujące:

- 1) Poziom źródła danych.
Dane pochodzą z różnych źródeł; baza musi umożliwiać ich gromadzenie w trybie wsadowym, a także w czasie rzeczywistym;
- 2) Poziom komunikacji ze źródłami danych i decyzji o sposobie ich przechowywania.
Ta warstwa faktycznie otrzymuje dane ze źródła. W razie potrzeby musi mieć możliwość przetworzenia nieustrukturyzowanych danych do konkretnej postaci w celu ich dalszej analizy.
 - Dane ustrukturyzowane, typowe dla relacyjnych baz danych, *Relational Database Management Systems* (RDBMS);
 - Dane nieustrukturyzowane w specjalnych systemach plików takich jak *Hadoop Distributed File System* (HDFS), [1]
 - Dane nieustrukturyzowane w nierelacyjnych bazach danych typu NoSQL.
- 3) Poziom analizowania, zarówno danych ustrukturyzowanych, jak i nieustrukturyzowanych.
- 4) Prezentacja wyników analityki Big Data, dostępna w różnych formach, np. poprzez oglądanie raportów przez użytkowników, aplikacje, procesy biznesowe.

Oprócz takiej warstwowej struktury architektury systemów Big Data, istnieją też procesy działające między warstwami. Istotnymi w tej grupie są te, mające zapewnić bezpieczeństwo danych i użytkowników, a także związane z zarządzaniem infrastrukturą. Architektura Big Data najczęściej funkcjonuje na dużej ilości klastrów, o wysokiej skalowalności i wydajności. Jednostka IT przedsiębiorstwa musi na bieżąco monitorować oraz zapewniać działanie i bezpieczeństwo systemu Big Data. Podobne działania zapewniające bezpieczeństwo niezbędne są również w przypadku architektury w chmurze.

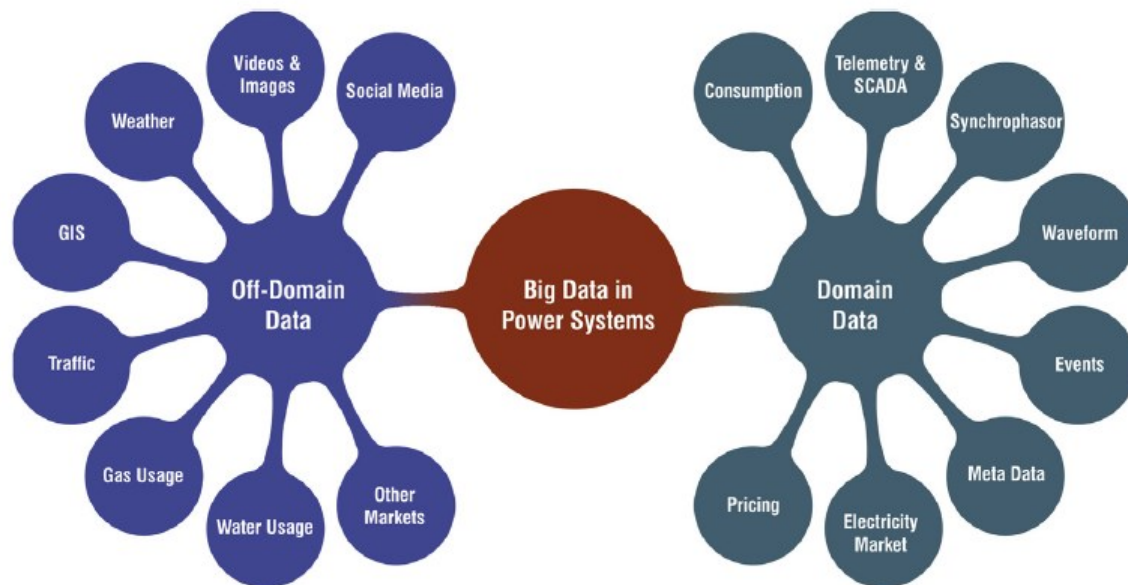
3. Źródła Big Data w elektroenergetyce

W sektorze elektroenergetycznym, czy szerzej energetycznym, analizy biznesowe prowadzone są w oparciu o dostępne dane od wielu lat, tak samo jako to ma miejsce w innych gałęziach gospodarki. O ile w prężnie rozwijającej się branży szeroko rozumianych usług internetowych, analityka danych jest poniekąd domyślnie zaimplementowana w oferowane serwisy, w branży elektroenergetycznej dane do statystyk trzeba pozyskiwać z instalowanych coraz powszechniej urządzeń pomiarowych, zwłaszcza typu smart. Należy pamiętać, że w elektroenergetyce dane pozyskiwane są w wielu obszarach działalności sektora i ich analizy służyć mają nie tylko celom biznesowym, związanym z oferowaniem klientom nowych usług i produktów, ale także mają wspierać operatorów w utrzymaniu bezpiecznej i stabilnej pracy systemu elektroenergetycznego, poprzez utrzymanie efektywnej pracy przedsiębiorstwa, wsparcie procesów optymalizacji kosztów utrzymania i modernizacji infrastruktury sieciowej.

Można jednakże zadać sobie obecnie pytanie czy już w systemie elektroenergetycznym mamy do czynienia z danymi określanymi pojemnym terminem Big Data. Oczywiście jeśli chodzi o *volume*, czyli ilość generowanych w systemie elektroenergetycznym danych, odpowiedź brzmi „tak”. Na przestrzeni ostatnich lat wskutek szybkiego wzrostu liczby zainstalowanych inteligentnych liczników u odbiorców końcowych oraz rozwoju systemów automatycznego opomiarowania infrastruktury, ilość danych przez nie generowanych zawrotnie wzrosła. Dodatkowo, dane w sektorze energetycznym podlegają kryterium *variety*. Pochodzą z szeregu rozmaitych źródeł, czujników zainstalowanych w różnych lokalizacjach,

mających różne zastosowanie w różnych miejscach w systemie. W większości sytuacji dane sieciowe generowane są z bardzo małymi interwałami, więc podlegają kryterium *velocity*. Podstawowe ograniczenie jakie pojawia się przy określeniu czy dane w sektorze elektroenergetycznym są już Big Data wynika z tego w jaki sposób są one agregowane. W przypadku większości nawet najnowocześniejszych czujników pomiarowych, dane są gromadzone w sposób krótkotrwały, pozwalający na detekcję zdarzenia systemowego i odpowiednią reakcję. Jednakże po takim zdarzeniu w krótkim czasie dane są nadpisywane. Dodatkowo, dotychczas w większości aplikacji czujników pomiarowych generujących automatycznie dane z małym krokiem czasowym, dane gromadzone były lokalnie w dedykowanych serwerowniach. Z różnych względów, z których najważniejszym jest cyberbezpieczeństwo infrastruktury krytycznej, dane nie były przesyłane do centralnych repozytoriów danych, gdzie mogłyby podlegać wielkoskalowej analizie, w przeciwieństwie do tego, jak ma to miejsce w przypadku sektora IT.

Należy podkreślić, że dane z sektora elektroenergetycznego w znaczeniu ilości, różnorodności, szybkości ich generacji, spełniają kryteria Big Data. Dodatkowo, warto rozróżnić je na dwie kategorie, związane ze źródłem ich pochodzenia. Klasyfikowane są wówczas jako dane domenowe i spoza domeny. Zgodnie z referencją [2] można przedstawić następujący diagram.



Rys. 1. Klasyfikacja i źródła Big Data w systemie elektroenergetycznym [2]

Dane domenowe pochodzące wprost z układów opomiarowania w systemie elektroenergetycznym pochodzą z następujących źródeł:

- dane telemetryczne i z systemów SCADA;
- synchrofazory;
- zapotrzebowanie, zużycie – dane pomiarowe;
- zdarzenia systemowe;
- dane finansowe, pomiarowo-rozliczeniowe.

Praca systemu elektroenergetycznego i jego sterowanie, oparte są także o inne dane, określane jako pozadomenowe, czyli dane nie związane wprost z sektorem elektroenergetycznym. Do tej grupy można zaliczyć dane meteorologiczne, dane dotyczące wyładowań atmosferycznych, dane geoprzestrzenne (tzw. dane GIS – *geographic information system*). Obecnie katalog źródeł danych, które mogą wspierać analitykę Big Data dla sektora elektroenergetycznego, pozostaje otwarty i właściwie nieograniczony. Mogą się w nim znaleźć dane dotyczące ruchu ulicznego, dane z serwisów społecznościowych, obrazy, strumienie video, dane giełdowe.

Podejście analityki Big Data stało się możliwe dzięki zaistnieniu źródeł danych, na przykład poprzez układy automatycznego opomiarowania, ale także dzięki powstaniu platform do nich, dostarczających rozwiązania chmurowe, które w skuteczny sposób podnoszą dostępność cenową akwizycji danych masowych, ich przesyłania i gromadzenia. Ponadto, najważniejszą cechą analityki Big Data jest to, że prowadzona jest ona w charakterze odkrywczym czy też wyjaśniającym pewne trendy, bez wymagania o konieczności posiadania wstępnego rozeznania, jakie trendy powinny być poszukiwane. Obecnie analityka Big Data wyposażona jest w szereg sprawdzonych metodologii, stanowiących hybrydowe połączenie technik analitycznych pochodzących z sektora IT, sterowanych danymi, z tradycyjnymi metodami analitycznymi

wypracowanymi w systemie elektroenergetycznym, bazującymi na odpowiednich modelach. Metodologie wypracowane dla analityki Big Data w systemie elektroenergetycznym są bardzo interesujące i powinny stanowić w całości treść odrębnego artykułu.

4. Potencjał analiz Big Data w systemie elektroenergetycznym

Mimo, iż zgodnie z podstawowymi własnościami Big Data, właściwie nie ma ograniczeń na to, ile i jakiego typu nowych informacji analityka Big Data może przynieść dla sektora elektroenergetycznego, niektóre z obszarów ich zastosowania już zostały nakreślone [3], [4], [5], [6], [7], [8], [9], [10], [11]. Najważniejszymi zidentyfikowanymi jak do tej pory, są następujące kierunki zastosowań analityki Big Data w sektorze elektroenergetycznym:

- a) poprawa dynamiki sterowania pracą systemu elektroenergetycznego;
Zagadnienie to rozumiane jest w kontekście analityki Big Data, jako nowatorskie podejście do problemu *Demand Response*. Do tej pory kreowanie odpowiedzi systemu na zapotrzebowanie stymulowane było wielkoskalowo, w znaczeniu zarówno infrastruktury, jak i dużych uczestników rynku energii. Analityka Big Data daje szansę na przeniesienia zagadnienia do rozwiązań chmurowych, gdzie obserwowane będą przez odpowiednie podmioty wzorce zachowań milionów odbiorców końcowych i na tej podstawie wyciągane wnioski, które spośród grup użytkowników mogą partycypować w *Demand Response* i wpływać poprzez efekt skali na pracę systemu.
- b) prognozowanie o wysokiej dokładności (ziarnistości);
Analityka Big Data, przeniesiona do chmury obliczeniowej, daje możliwość prognozowania zapotrzebowania oraz generacji, z rosnącą precyzją i granulacją prognoz, w miarę dostarczania kolejnych warstw danych do modeli obliczeniowych. Przykładem platform dostępnych do celów testowych za darmo, jest m.in. IBM PAIRS GEOSCOPE (*Physical Analytics Integrated Data Repository and Services*), gdzie wielowarstwowa analityka Big Data została już z powodzeniem zastosowana do prognozowania generacji farm wiatrowych, a także do wsparcia działań z rodziny *Predictive Maintenance*. W tym kontekście połączenie danych geoprzestrzennych, dostępnych w ramach platformy, z danymi dotyczącymi składu gleby oraz roślinności na wybranych obszarach Stanów Zjednoczonych, wsparło amerykańskiego operatora Oncor we wprowadzeniu inteligentnego systemu przycinania roślinności wzdłuż linii napowietrznych (*smart trimming*). W obszarze prognozowania generacji ze źródeł odnawialnych, ogromne nadzieje pokładane są w wykorzystaniu sieci neuronowych, czyli mechanizmów sztucznej inteligencji. Oczywiście kwestią jest tu problem walidacji tworzonych modeli oraz dostępności danych treningowych i warstw do tworzonych modeli (np. prognoz meteorologicznych), ale potencjalne korzyści płynące z takiej metodyki prognozowania są ogromne.
- c) wykorzystywanie danych pochodzących wprost z sektora elektroenergetycznego i pozadomenowych do wykrywania zwarc i wyłączeń;
W tym kontekście nawet dane pochodzące z mediów społecznościowych, wskazujące na pewne wzorce zachowań wśród odbiorców końcowych, wraz z trwającym procesem decentralizacji wytwarzania, mogą stanowić cenne źródło informacji dla inteligentnych systemów wykrywania zwarc. Projekty pilotażowe w tym obszarze realizowane były z wykorzystaniem technologii takich jak logika rozmyta czy sieci neuronowe. Pozytywne wyniki stanowią rekomendację dla rozwijania takich technik w domenie wykrywania zwarc.
- d) wsparcie dla sterowania pracą systemu poprzez wzrost obserwowalności stanu;
- e) monitorowanie sprzętu i wydłużanie jego życia poprzez predykcyjne zapobieganie zwarciom, oraz szerzej – *Predictive Maintenance*;
- f) zaawansowana wizualizacja danych oraz wsparcie w walidacji i weryfikacji, czyli wyborze danych wartościowych w analizach;

5. Potencjalne przeszkody w wykorzystaniu analityki Big Data w sektorze elektroenergetycznym

Jako impuls do intensywnego rozwoju obszarów wykorzystania analityki Big Data w systemie elektroenergetycznym należy traktować regulacje wprowadzone przez ENTSO-E, nakazujące zrzeszonym operatorom działania w stronę transformacji istniejącego systemu w postaci fizycznej do wynikowej, cyber-fizycznej. Działania takie obserwowane są na całym świecie. Hasła, bardzo istotne z punktu widzenia bezpieczeństwa energetycznego i ciągłości dostaw energii elektrycznej, jakiej oczekujemy jako odbiorcy, a które warunkowane są również dokonującą się transformacją, to poprawa niezawodności, elastyczności, dostępności mocy i energii.

Aby wykorzystanie ogromnego potencjału analityki Big Data dla pozytywnej rozbudowy systemu było możliwe, konieczny jest rozwój domen związanych ze zbieraniem, komunikacją i zarządzaniem danymi.

Wymienione kwestie związane są z zagadnieniami technologiczno-informatycznymi i nie rzadko wymagają innowacyjnego podejścia. Jednocześnie podczas projektowania infrastruktury informatycznej i oprogramowania wspierającego analitykę Big Data, konieczne jest odpowiednie rozpoznanie i zaadresowanie wyzwań, oczekiwań oraz funkcjonalności, jakie niesie ze sobą analityka Big Data w sektorze elektroenergetycznym. W szczególności zagadnienia takie, jak wymienione poniżej, mogą być traktowane jako ograniczenia czy też wręcz bariery na drodze do efektywnego wdrożenia wielkoskalowej analityki danych w systemie elektroenergetycznym:

- a) pozyskiwanie danych, do tej pory odrzucanych jako nie posiadające wartości dla analityki danych systemowych;
- b) pozyskiwanie "silosowanych" danych, ze źródeł o bardzo ograniczonym dostępie dla użytkowników z zewnątrz;
- c) konieczność wsparcia dla analiz przeprowadzanych w czasie rzeczywistym;
- d) współistnienie zarządzania danymi centralnego i rozproszonego;
- e) wypracowanie efektywnych mechanizmów pozwalających na gromadzenie, zarządzanie i analizę szybko zmiennych i szybko napływających danych.

Należy w tym miejscu podkreślić, że niezwykle istotnym aspektem przy projektowaniu centralnych repozytoriów danych systemowych (a także pozadomenowych, ale docelowo wspierających analitykę systemową), jest zachowanie względów cyberbezpieczeństwa. W wybranych państwach zrzeszonych w ENTSO-E takie scentralizowane bazy danych już działają. Zasilane są danymi pomiarowymi pochodzącymi wprost z infrastruktury sieciowej, ale nie tylko. Szereg podmiotów, np. administracja rządowa, wysyła dane o cechach statystycznych, które jak już wspomniano, łącznie z danymi domenowymi przy odpowiednim zastosowaniu analityki Big Data dają ogromny potencjał ulepszający pracę systemu, jak i wspierający usługi biznesowe OSP i OSD. Ważne jest zachowanie uprawnionego dostępu do danych pierwotnych, czy też do raportów, odpowiednio do pełnionej roli. Po stronie operatora takiej scentralizowanej bazy danych stoi zapewnienie infrastruktury przechowującej napływające dane, a także opracowanie formatu, w którym będą przekazywane. Prace takie w Polsce trwają i służyć mają wprowadzeniu instytucji Operatora Informacji Rynku Energii (PSE S.A.), który zarządzać będzie centralnym repozytorium danych, CSIRE (Centralny System Informacji Rynku Energii).

Ciekawym zagadnieniem w kontekście zapewnienia cyberbezpieczeństwa infrastruktury krytycznej jest rozpoznanie wpływu informatyki kwantowej na bezpieczeństwo pracy systemu elektroenergetycznego. Należy mieć świadomość, że liczący się uczestnicy rynku IT, tj. Google, IBM, Microsoft czy D-wave, ogłosili już opracowanie nie tylko koncepcji, ale również realizację fizyczną konkretnej "maszyny". Wobec prawdziwie kwantowych komputerów klasyczne algorytmy szyfrowania, zapewniające nasze cyfrowe bezpieczeństwo, są bezradne. Algorytmy stosowane obecnie w systemach szyfrujących nigdy nie były niezniszczalne, jednakże zapewniały bezpieczeństwo poprzez fakt, że złamanie klucza zajmuje mnóstwo czasu, przy wykorzystaniu komputera klasycznego. Gwarancji takiej nie ma przy zastosowaniu komputera kwantowego. Temat cyberbezpieczeństwa w obliczu rozwijających się, nie tylko teoretycznie, informatyki i kryptografii kwantowych powinien stać się treścią odrębnego artykułu.

6. Podsumowanie

System elektroenergetyczny podlega obecnie intensywnym zmianom. Jego transformacja jest procesem wielopłaszczyznowym, związanym ze zmianami w sektorze wytwórczym i intensywnym wzrostem udziału odnawialnych źródeł energii, jak również z procesami takimi jak instalacja inteligentnych liczników oraz systemów automatycznego opomiarowania wspierających np. monitorowanie i sterowanie pracą KSE, czy też wzrostem aktywnego udziału odbiorców w rynku energii. Zachodzące zmiany prowadzą do transformacji systemu elektroenergetycznego z jego fizycznej postaci do cyber-fizycznej, co pociąga za sobą intensywny wzrost ilości przetwarzanych danych związanych z jego pracą. Ilości te wymagać będą zastosowania technik gromadzenia, zarządzania i analityki dedykowanych danym typu Big Data. Aby analityka Big Data faktycznie zaistniała i przyniosła operatorom korzyści w postaci poprawy obserwowalności stanu systemu elektroenergetycznego, sterowania jego pracą, elastyczności, a także propozycję nowych usług biznesowych i marketingowych, konieczne jest odejście od tradycyjnego podejścia gromadzenia danych wewnątrz dedykowanych hurtowni o wysoce limitowanym dostępie. Sposób gromadzenia danych domenowych i pozadomenowych, które podlegać powinny analizie Big Data musi oczywiście spełniać wymogi cyberbezpieczeństwa, ale jednocześnie wymaga nowatorskiego podejścia umożliwiającego nowoczesną analitykę, odbywającą się w czasie rzeczywistym. Działania te docelowo służyć będą na poziomie OSP m. in. poprawie dynamiki reakcji na występujące w systemie zdarzenia, a na poziomie OSD dostosowaniem oferty kierowanej do odbiorców końcowych, których świadomość energetyczna obecnie silnie wzrasta.

Literatura

- [1] <http://hadoop.apache.org/>
- [2] Akhavan-Hejazi H., Mohsenian-Rad H., *Power systems big data analytics: An assessment of paradigm shift barriers and prospects*, Energy Reports 4, strony 91–100 (2018)
- [3] Bui N., Castellani A.P., Casari P., Zorzi M., *The internet of energy: A web enabled smart grid system*. IEEE Netw. 26 (2012)
- [4] Huang Z., Luo H., Skoda D., Zhu T., Gu Y., *E-sketch: Gathering large-scale energy onsumption data based on consumption patterns*, IEEE International Conference on Big Data, strony. 656–665 (2014)
- [5] Stimmel C.L., *Big Data Analytics Strategies for the Smart Grid*, Auerbach Publications (2014)
- [6] Kezunovic M., Xie L., Grijalva S., *The role of big data in improving power system operation and protection*. Bulk Power System Dynamics and Control, Optimization, Security and Control of the Emerging Power Grid (IREP), IREP Symposium. IEEE, strony. 1–9 (2013)
- [7] Moreno-Munoz A., Bellido-Outeirino F., Siano P., Gomez-Nieto M., *Mobile social media for smart grids customer engagement: Emerging trends and challenges*, Renew. Sustain. Energy Rev. 53, strony 1611–1616 (2016)
- [8] Hu J., Vasilakos A.V., *Energy big data analytics and security: Challenges and opportunities*. IEEE Trans. Smart Grid 7 (5), strony 2423–2436 (2016)
- [9] Yu N., Shah S., Johnson R., Sherick R., Hong M., Loparo K., *Big data analytics in power distribution systems*, Innovative Smart Grid Technologies Conference (ISGT), IEEE Power & Energy Society, strony 1–5 (2015)
- [10] Yin J., Sharma P., Gorton I., Akyoli B., *Large-scale data challenges in future power grids*, IEEE 7th International Symposium on Service Oriented System Engineering (SOSE). IEEE, strony 324–328 (2013)
- [11] Qing L., Boyu Z., Qinqian L., *Impact of Big Data on Electric-power Industry*, IEEE 2nd International Conference on Big Data Analysis (2017)

