Expanding Global Big Data Solutions with Innovative Analytics

Umeshwar Dayal, PhD   Masaharu Akatsu, PhD   Chetan Gupta, PhD
Ravigopal Vennelakanti   Massimiliano Lenardi, PhD

Through the proliferation of sensors, smart machines, and instrumentation, industrial operations are generating ever increasing volumes of data of many different types. At the same time, advances in computing, storage, communications, and big data technologies are making it possible to store, process, and analyze enormous volumes of data at scale and at speed. The convergence of OT and IT, powered by innovative analytics, holds the promise of creating new social innovation businesses. Hitachi's R&D Group has established the Global Big Data Innovation Lab (GBDIL) to coordinate world-wide analytics research activities in support of the global expansion of the social innovation solution businesses by providing innovative analytics to the recently launched Hitachi Global Center for Innovative Analytics (HGC-IA). This article introduces GBDIL and HGC-IA, and describes a common reference architecture for developing, deploying, and operating big data solutions that leverage Hitachi's innovative analytics technologies. Through the use of example use cases, the article explains the strategy to expand the global big data solution business.

1. MARKET AND TECHNOLOGY TRENDS

TODAY, we are at the dawn of transformative changes across industries, from agriculture to manufacturing, from mining to energy production, from healthcare to transportation. These transformations hold the promise of making our economic production more efficient, cost effective, and, most importantly, sustainable. These transformations are being driven by the convergence of the global industrial system [operations technology (OT)] with the power of integrating advanced computing, analytics, low-cost sensing, and new levels of connectivity [information technology (IT)].

Through the increasing use of interconnected sensors and smart machines and the proliferation of social media and other open data, industrial operations and physical systems produce a continuous stream of sensor data, event data, and contextual data. This unprecedented amount of rich data needs to be stored, managed, analyzed, and acted upon for sustainable operations of these systems. Big data technologies, driven by innovative analytics, are the key to creating novel solutions for these systems that achieve better outcomes at lower cost, substantial savings in fuel and energy, and better performing and longer-lived physical assets.

Opportunities to create big data solutions abound in many industries (power distribution, oil and gas exploration and production, transportation, telecommunications, healthcare, agriculture, and mining, to name a few) and in the public sector (homeland security, smart cities, population health management, etc.). To realize operational efficiencies and to create new revenue-generating lines of business from the deluge of data requires the convergence of IT × OT by leveraging an analytics framework to translate data-driven insights from a multitude of sources into actionable insights delivered at the speed of the business. To achieve this, innovations in analytics will be required: first, to deal with the vast volumes, variety, and velocity of data; and second, to create increasing value by moving from descriptive or historical analytics (what has happened and why?) to predictive analytics (what is likely to happen and when?), and finally to prescriptive analytics (what is best course of

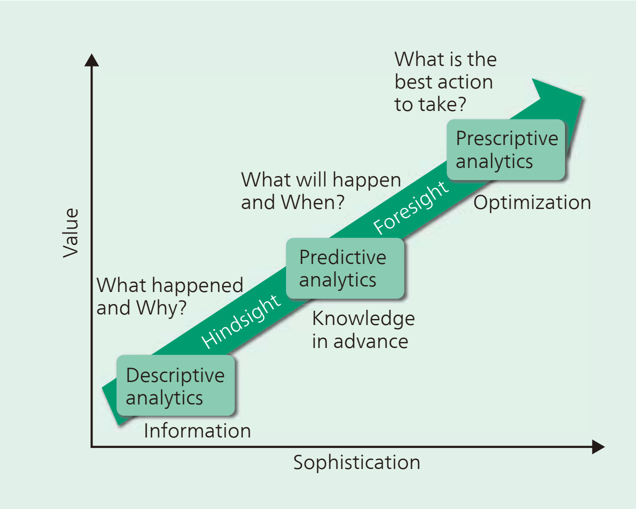


Fig. 1 | Evolution of Analytics Sophistication.

Analytics is evolving in both sophistication and value from “descriptive analytics” to “predictive analytics” to “prescriptive analytics.” Descriptive analytics helps describe the current state, predictive analytics attempts to predict what might happen next, and finally, prescriptive analytics aims to automatically suggest the best action to take.

action to take next?) (see **Fig. 1**).

2. HITACHI’S RESPONSE TO MARKET AND TECHNOLOGY TRENDS

Through its strengths in IT and OT, Hitachi is uniquely positioned to lead this transformation by developing Social Innovation Business in the global market.

Hitachi has taken a lead in addressing these opportunities by establishing the Hitachi Global Center for Innovative Analytics (HGC-IA). The HGC-IA was launched in June 2013 to provide customers globally with innovative solutions that leverage the power of big data by linking Hitachi operating units with researchers and engineers located in the U.S., Europe, Japan, and other regions around the globe. “The HGC-IA will provide opportunities for Hitachi to co-innovate with customers and partners to create value through advanced data analytics, developing solutions to meet customers’ challenges in areas such as sales, productivity, and cost. Vertical solutions are being developed across several industries, including: Health Care, Communications & Media, Energy, Transport, Mining, as well as enterprise fields such as logistics, manufacturing, and business intelligence for opportunities to deploy Hitachi technologies.”* The Global Big Data Innovation Lab, which serves as the R&D arm of the HGC-IA, harnesses the expertise and inventive power of over 300 researchers worldwide to deliver cutting-edge analytics technologies to the customer solutions being developed by the HGC-IA. Uniquely, in a single global structure, Hitachi has solution stacks, data scientists, architects, and consultants who build the right big data analytics solutions to reinvent the way global business operates.

Aided by these new organizations, a key element of Hitachi’s strategy is to create repeatable, high-value, industry-specific solutions that are developed over a common solution framework, and are delivered as appliances, packaged applications, or cloud services. By building on a common framework, Hitachi can develop repeatable industry solutions quickly and efficiently. This strategy leverages Hitachi’s assets and deep expertise in social infrastructure, IT, and solutions.

3. COMMON SOLUTION FRAMEWORK

Early adopters of industrial automated control systems deployed custom-built analytics applications implemented over a siloed architecture. However, this approach has several drawbacks: it is expensive and resource intensive; it is inefficient since every application needs to be built from scratch; it makes it challenging to scale and to deal with new functionality, new data types, and new quality-of-service requirements; and finally it is incomplete – most good application designers are not domain experts and are unfamiliar with the analytics needed over sensor and event data, and not that many domain analysts are familiar with application design.

Instead, the market demands a solution framework for repeatable, scalable solutions that cater particularly to the needs of analytics over operational data obtained from physical systems and other data sources such as social media. This framework must also abstract the underlying complexities of provisioning such a system for application architects while supporting different deployment models: centralized, core-edge, and in the cloud. Such a framework will make it easier to build prototypes for customers by reusing components, make it more efficient for system integrators to deploy the solutions in the customer environment, and enable the efficient development of “analytics-in-a-box” solutions for various verticals.

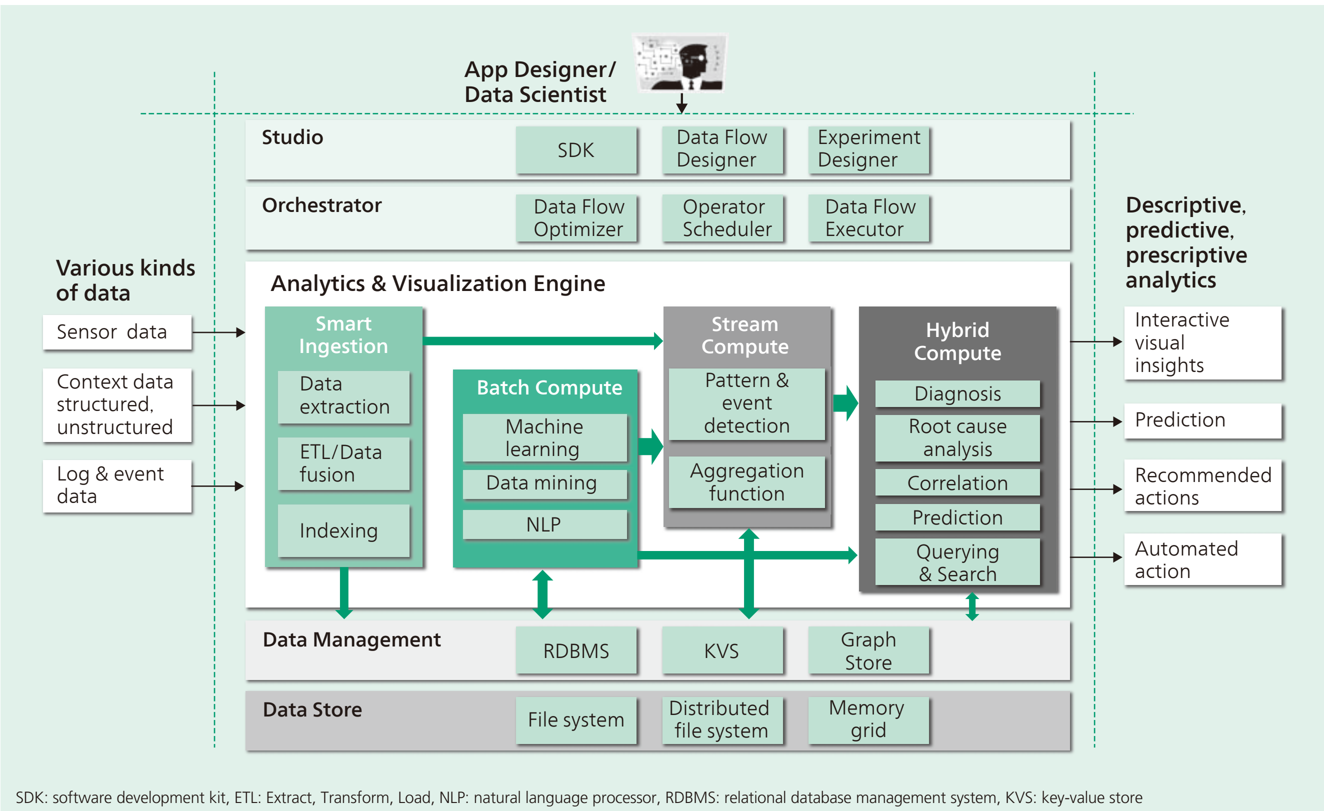
Our solution framework is based on years of experience building IT × OT solutions for a variety of vertical industries. In designing the framework, Hitachi has identified the following requirements: (i) ability to ingest, process, manage, and analyze both streaming data and large volumes of stored data; (ii) ability to process both time series and event data; (iii) ability to process both structured and unstructured data; (iv) ability to process data within the enterprise and outside the enterprise; and (v) ability to be deployed in the cloud and on-premise.

Because the framework needs to satisfy a variety of data processing and analytic needs, it will require a small number of data processing engines with different capabilities; plug-and-play capability to interchange engines; and the ability to create, optimize, and execute tasks that run over a combination of the engines. Appropriately defining the components and interfaces of this framework ensures that it can be used to develop and deploy solutions for a variety of industries.

The outputs of our data management and analytics framework over these various kinds of data should be interactive visual insights, predictions, and recommended manual or automated actions.

Fig. 2 shows the reference architecture of such a framework. The architecture consists of:

- Connectors to various data sources.
- An ingestion layer to prepare data for analytics processing and for storing the data.
- A data store layer, which provides different engines for storing data to be used for historical analysis.
- A data management layer, which provides different engines for managing and processing data of different types.
- A stream processing core for streaming analytics such as pattern and event detection and streaming on-line analytics processing.
- A batch processing layer for analytics over historical data, including machine learning, data mining, and natural language processing methods.
- A hybrid computation layer for analytics over combinations of streaming and stored data.
- A visualization, presentation, and automation layer for output in terms of interactive insights and automated actions.
- A development studio, which includes an Analytics Workbench for designing complex analytic data flows and experiments, and a repository of common analytics operations that can be reused in many applications.
- An orchestration layer, which includes an optimizer for analytics data flows, a scheduler, and an execution



SDK: software development kit, ETL: Extract, Transform, Load, NLP: natural language processor, RDBMS: relational database management system, KVS: key-value store

Fig. 2 | Common Reference Architecture.

Hitachi's common reference architecture aims to support the development and deployment of big data solutions across verticals. It consists of components for ingesting, storing, managing, analyzing, and visualizing big data over a single node or a cluster.

manager for managing the execution of analytics data flows on multiple execution engines. The framework allows for solutions that can run on a single node and on a large cluster.

4. CURRENT ACTIVITIES

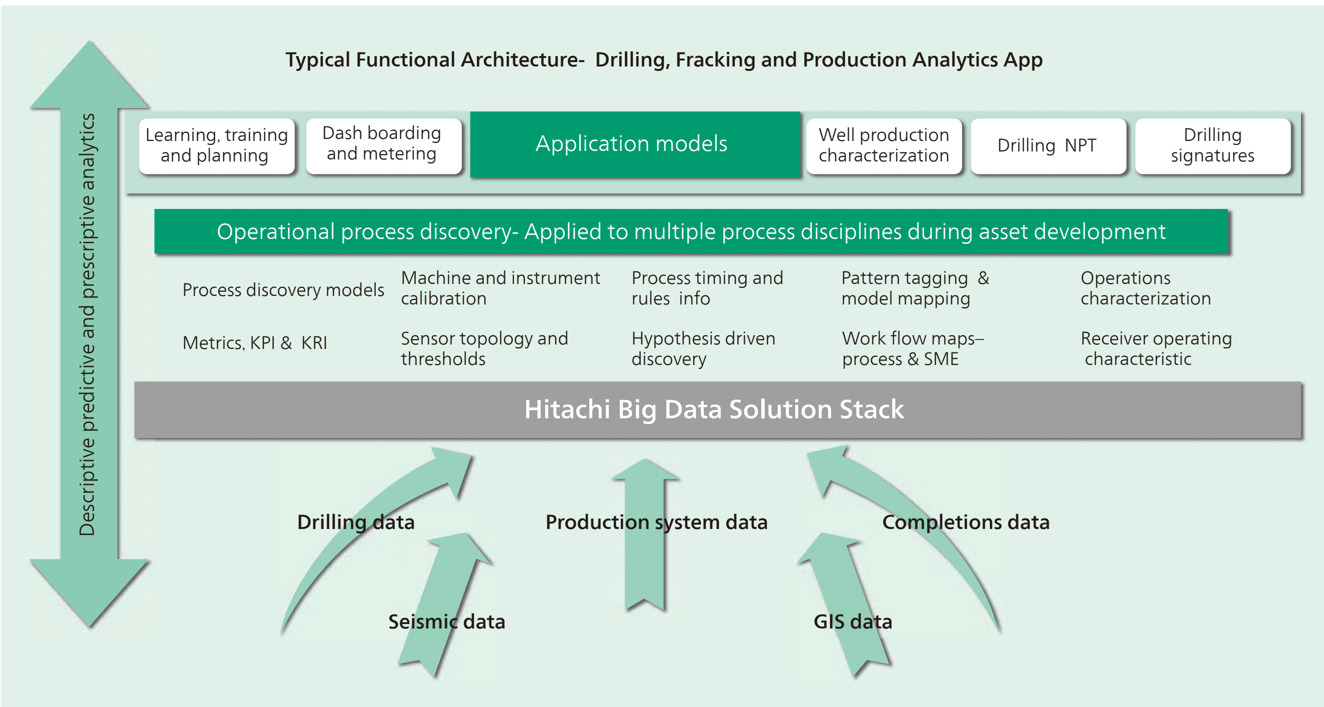
Besides architecting the common framework, HGC-IA and the Global Big Data Innovation Lab are involved in several proof of concept (PoC) activities. This includes being actively engaged in creating customer solutions for the communications, mining, shale oil and gas, healthcare, and other industries, with the goal of improving the efficiency and reliability of operations in these industries. The following sections describe two representative examples.

4.1 Oil & Gas Industry Solution

Industry experts recognize that technological advances in the development of energy sources such as shale gas, oil sands, and deep water resources have the potential to transform global energy markets. In particular, shale gas has a critical role to play in supplying future energy needs. Advances in horizontal directional drilling and hydraulic fracturing technologies have unlocked the potential for recovering enough natural gas from shale to power the US economy for a century. However, the problem of maximizing output from a shale gas reservoir is not well understood. This makes production decisions and sizing top-side facilities difficult. Furthermore, operators often struggle with realtime performance of support for down-hole gauges, semi-submersible pumps, and other equipment. Non-productive time may constitute over 30% of the cost of drilling operations. Descriptive, predictive, and prescriptive analytics applied to addressing these problems could have a profound impact on the production cost, efficiency, and environmental impact.

In keeping with an objective of making repeatable solutions, Hitachi is aiming to build a solution for this industry. **Fig. 3** presents a schematic of this solution.

To strengthen the vision and execution of Hitachi's interests in the oil and gas technology domain, the US Big Data Lab has established collaboration with the Energy & Environmental Research Center (EERC) at the University of North Dakota. The EERC is a contract research organization that focuses on energy and related environmental issues. Hitachi and EERC have prioritized the research focus to solve the hard problems operators face today in developing the Bakken Field for sustained



NPT: Non-productive time, KPI: Key performance indicator, KRI: Key risk indicator, SME: Structure-mapping engine, GIS: Geographic information system

Fig. 3 | Schematic of an Oil & Gas Solution.

The upstream oil and gas industry solution processes several kinds of data, including geology, drilling, production, and operations data. The data is managed and analyzed through Hitachi's big data solution stack and is combined with domain specific knowledge to build several oil and gas industry specific applications such as well production characterization and reduction of non-productive time.

operations across the processes.

The Bakken Production Optimization program has two primary goals:

(1) Optimizing well productivity by better understanding the underlying Bakken field geology and its response to different wellbore trajectories and completion techniques.

(2) Improving drilling efficiency and reducing the surface impact of drilling by potentially using water and gas generated at the well site, as well as other techniques.

To meet the above goals, our solution will derive insights from a vast amount of data collected and stored as a "system of record" today. These insights will help transform the orthodox approach of deploying systems of record to deploying "systems of engagement" to reduce the cost of operations, and to improve the productivity, reliability, and sustainability of operations (see **Fig. 4**).

Fig. 4 | Production Characterization Application.

The dashboard for the production characterization application displays the complex, non-intuitive interrelationships between geology, completions, production, and operations data.

4.2 Healthcare Industry Solution

Companies operating in the health care industry today are dealing with the significant challenge of providing services and systems that improve the quality of medical care for populations in the face of rising medical care costs, often associated with the additional dimension of an aging society. To tackle this problem, Hitachi established the European Big Data Laboratory (EBDL) within the University of Manchester Innovation Centre (UMIC) to develop and deliver Proof of Concept (PoC) projects for the conurbation of Greater Manchester, UK, working closely with the National Health Service (NHS) England partners in Salford, Central and South Manchester.

The PoC projects are being undertaken by EBDL together with NHS in Greater Manchester (NHS England in GM), North West e-Health (NWeH) and the Manchester Academic Health Science Centre (MAHSC), with the aim of achieving an increased quality of life while containing medical costs by making full use of IT and big data. Specifically, the projects involve the development and verification of a secure health care data integrated (federated) platform and a life-style improvement program for pre-diabetics.

Fig. 5 shows the overview of the secure healthcare data integrated platform and key technologies. This platform would enable a variety of healthcare services to securely and appropriately share patient data to better inform the delivery and development of services. The platform has

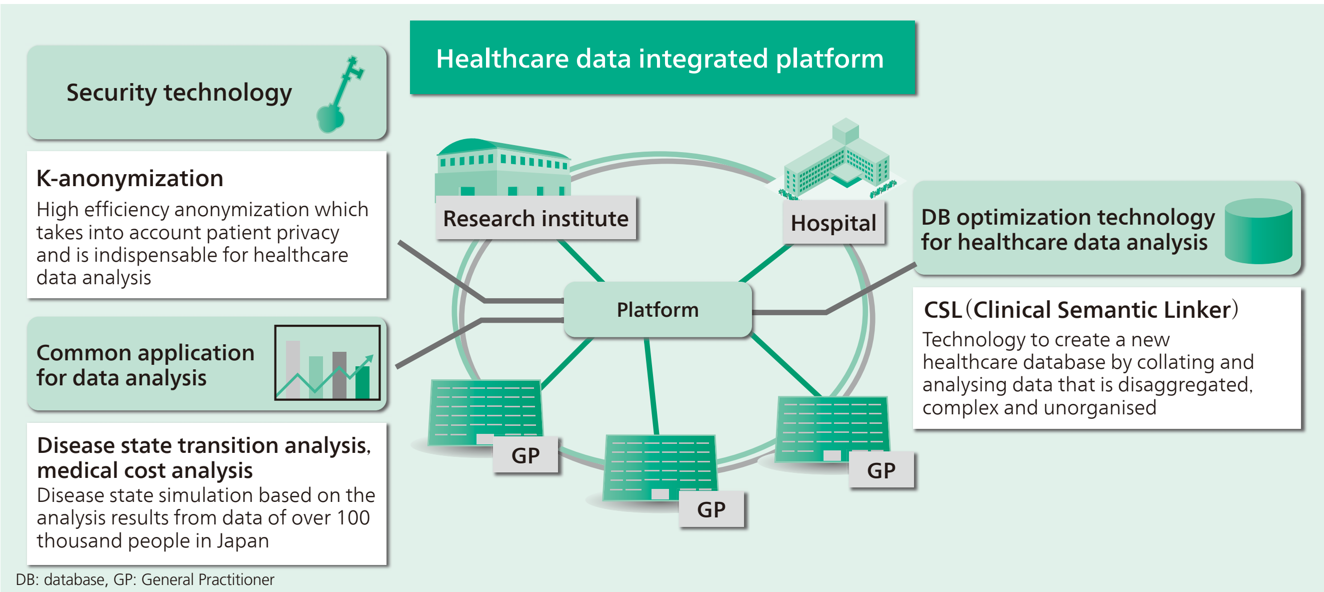


Fig. 5 | UK Healthcare Integrated Platform.

Federated data platform to facilitate the sharing of medical information in order to enable secure analytics technologies to deliver new high quality healthcare services across primary, secondary and social care.

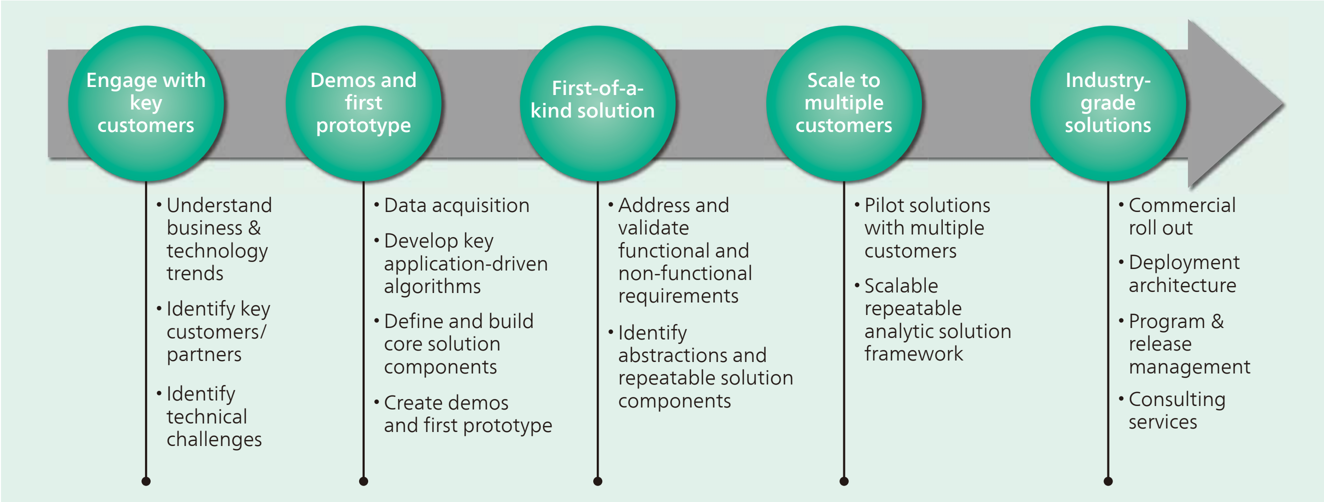


Fig. 6 | **Solution Methodology.**
The stages of our solution methodology consist of engaging key customers in a specific vertical, building a first demo/prototype, testing and validating the first prototype, scaling to multiple customers, and finally commercially rolling out as an industry grade solution.

three primary goals:

(1) Expanding the coverage of secure data-sharing from 0.25 million people in Salford to 2.87 million people across Greater Manchester.

(2) Developing innovative IT systems to tackle the new challenges of increasing complexity in security and, information governance whilst promoting the data analysis that will support service development and assurance, while keeping patient data privacy paramount.

(3) Creating new services that improve patient outcomes reduce variation in treatment, making high quality services available to larger populations and allowing the secure sharing of information between all of the organizations and institutions involved in health and wellness, helping to keep patients healthy for longer.

Based on the results from these initial PoC projects, there are plans for taking the knowledge learned and products and services developed to expand Hitachi's healthcare business to other regions within the U.K., followed by markets in other commonwealth states, North America, Japan and Europe.

5. CONCLUSIONS AND FUTURE ACTIVITIES

Hitachi will continue to work with its partners to build out big data platforms, analytics services, and solutions with the objective of establishing its position as a top-tier player in the IT × OT convergence space. **Fig. 6** shows the process Hitachi is using for moving beyond the PoC stage to select customer deployments, and then to full roll out of industry solutions.

REFERENCES

*"Hitachi Announces to Establish Hitachi Global Center for Innovative Analytics to Accelerate Global Expansion of Big Data-Related Businesses" - Hitachi, Ltd., Tokyo, June 6, 2013.

ABOUT THE AUTHORS


Umeshwar Dayal, PhD
Big Data Lab, Hitachi America Ltd., R&D.
He is Vice-President and Senior Fellow of Information Research at Hitachi America Ltd., R&D. Dr. Dayal is a Fellow of Association for Computing Machinery (ACM) and a member of the IEEE.


Masaharu Akatsu, PhD
Yokohama Research Laboratory, Hitachi Ltd.
He is General Manager of the Information Service Research Center. Dr. Akatsu is a Fellow of The Institute of Electrical Engineers of Japan (IEEJ) and a member of the IPSJ and Society for Serviceology.


Chetan Gupta, PhD
Big Data Lab, Hitachi America Ltd., R&D.
He is Principal Scientist and Architect at Hitachi America Ltd., R&D Big Data Lab. He is currently leading an effort in software reference architecture and analytics for Big Data Analytics Solutions. Dr. Gupta is a member of ACM.


Ravigopal Vennelakanti
Big Data Lab, Hitachi America Ltd, R&D.
He is Sr. Director and Strategist at Hitachi America Ltd., R&D Big Data lab. He is currently leading efforts in applying big data analytics to the oil and gas industry.


Massimiliano Lenardi, PhD
European Big Data Laboratory, Hitachi Europe Ltd, R&D. He is Senior Researcher at Hitachi Europe and is managing EU Big Data Lab as well as the Information and Communication Technology Laboratory (ICTL, in France). He represents the Hitachi Group in ETSI (EU Telecommunication Standardization Body) and is an IEEE Member. Dr Lenardi holds a PhD in Communication Systems from EPFL (Switzerland).