

CNPC E&P Data Management System (A1 Project)*

Tao Ma^{1,2}, Tiecheng Wang^{1,2}, Jingyi Liu^{1,2}, Wenjun Huang¹, and Fusen Yang¹

Search and Discovery Article #70342 (2018)**

Posted August 13, 2018

*Adapted from extended abstract based on oral presentation given at the GEO 2018 13th Middle East Geosciences Conference and Exhibition, Manama, Bahrain, March 5-8, 2018

**Datapages © 2018. Serial rights given by author. For all other rights contact author directly.

¹Richfit Information Technology Co., Ltd., Beijing, P. R. China (yangfusen@cnpc.com.cn)

²Information Technology Center, BGP Inc., CNPC, Zhuozhou, Hebei Province, P. R. China

Abstract

CNPC E&P Data Management System V2.0 (referred to as ‘A1 2.0’) was implemented from 2013 and put into operation in 16 oil & gas companies of CNPC in 2016. Based on CNPC integrated E&P Data Model (EPDM V2.0) derived from PPDM, A1 2.0 is developed and supports configurable and customized data collection, integration, storage management and service. It provides functionalities of data entry/import, load and ETL to realize data collection, synchronization and integration; Provides methods of data parsing, quality rule definition, scanning engine, quality scanning, visualization to carry out data quality control based on life-cycle data management and quality control mechanism; Provides functionalities of meta data management, master data management, data model management, security management to realize centralized data management; Provides functionalities of online data query/search, browse, display, order, download.

Introduction

China National Petroleum Corporation (CNPC) is one of the largest international energy corporations in China, with business covering upstream, midstream and downstream, owning 16 oil & gas companies in China, operating business in 37 countries and regions overseas. In 2016, its oil & gas equivalent reaches 1,466.6 million barrels. Over the past seven decades, CNPC has invested a lot of money in the petroleum E&P and a vast amount of E&P technical data has been acquiring stored in different and in different environments (e.g., archives, E&P research institutes, etc.) before the implementation of A1 project, without unified data standards, without data quality assurance, without efficient data sharing facility. How to protect these data free from spoiled and explore more values from these data conveniently and efficiently becomes a serious challenge and a robust solution is urgently required.

In 2001, CNPC planned an E&P Data Management System project (A1 1.0), aiming to establish A1 for centralized, high-qualified and standardized management of E&P data. We realized data asset management and service, with data scope covering 11 E&P domains, including geophysics, drilling, mud logging, logging, hydrocarbon test, well test and production test, sample experiment, downhole operation, geology

and reservoir, and comprehensive research. A1 1.0 was constructed from 2004 and put into operation in 2008 in 13 oil and gas companies of CNPC. In 2011, A1 2.0 was planned aiming to develop a brand new data management system to replace the old version A1 1.0, and put into operation in 16 oil and gas companies in 2016. In the implementation of A1 2.0, objectives and expected results are achieved.

Objectives and Challenges

The overall goal of A1 2.0 project is, based on the achievement of the A1 1.0 project, to build a new integrated exploration and production technical data management platform. This will optimize the process of data acquisition and service, improve capacity of data acquisition fast and accurate and the quality of E&P data, to strengthen the fast response ability of E&P data application, to expand the scopes of management and application of the E&P data, to realize data exchange, sharing and transmission with existing information systems, to enhance data service abilities for E&P researching environment, E&P management and decision-making supports, to further meet application needs of comprehensive geological research, E&P management and other business, to realize,

- 1) Unified standards: through combing with data standards and management standards of geophysics, drilling, mud logging, well logging and well test, to promote the unification of standards in the fields of E&P and engineering technology, laying the foundation for business collaboration and data sharing.
- 2) Improvement of data quality: based on combing data sources and data process, to establish an integrated E&P data acquisition platform, improving the timeliness and quality of data for A1.
- 3) Combination with business applications: aiming at comprehensive research and business management requirements, to study and develop functionalities such as wellbore visualization, exploration management and so on to improve service capabilities.

The work scopes of the A1 2.0 project include,

- 1) To develop a new, integrated E&P technical data management platform, supporting data loading, data management and data services;
- 2) To apply and expand EPDM, to improve management of data of non-seismic, core, special well logging, pre-stack seismic, to integrate the existing logging database, structured wellbore database and research archive database into single system, A1 2.0;
- 3) To optimize data acquisition process, to realize data exchange, sharing & transmission with CNPC Engineering Technology Operation Management System (A7) so that to realize the change of wellbore engineering data acquisition manner from manual to automatic;
- 4) To improve the data channel from A1 database to E&P researching environment, enhancing the convenience and automation of data applications;
- 5) To increase data services of E&P management and decision-making supports, expanding application scope of the A1;
- 6) To upgrade E&P researching environment to OpenWorks R5000, realizing integration with existing E&P research environments of oil companies;
- 7) To expand applications to exploration management, integrated wellbore etc.;
- 8) To construct advanced and applicable infrastructure and information security solutions, realizing deployments of A1 2.0 in CNPC's 16 oil and gas companies, Research Institute of Petroleum Exploration and Development and headquarters data center.

In the implementation of A1 2.0, main challenges include,

- 1) To further improve unified data standards applied in CNPC.
- 2) To deal with complexity of data integration with distributed data sources of multi-discipline and multi-data-type.
- 3) To further improve E&P data quality, including normalization, accuracy, integrity, consistency, timeliness.
- 4) To provide efficient facilities for data exchange of various types of storage media.
- 5) To achieve efficient data sharing between A1 2.0 and mainstream E&P applications, such as OpenWorks, Petrel, GeoEast, etc.

Solution

The complete solution for the A1 2.0 project actually includes several sub-solutions, mainly including the E&P data management solution, the infrastructure solution and the information security solution. In this paper, only the first solution is discussed. The E&P data management solution introduced for A1 2.0 is able to achieve life-cycle management of E&P data, in compliance with international and CNPC standards. It enables us to achieve data integration using methods like data input, data load, data synchronization, data link which follows data exchange standard. We achieve centralized storage management of static data and real-time data using methods like master data management, data quality control, data model management, security control; to provide web-based functionalities of data query, browse, visualization, download and provide deep level applications (e.g., big data mining) using open data service interfaces and application service interface.

Architectures

Combining with the objectives and requirements of A1 2.0 project, the overall architecture ([Figure 1](#)) includes four layers, i.e., data acquisition, data management, data service and business platform.

Data acquisition layer: according to the features of data sources, it provides different ways for data acquisition, including integrated data acquisition, synchronization, migration, import, input, load, etc., with the unified data acquisition platform.

Data management: the unified data model is applied for the centralized and hierarchical management of bulk data, structured data, documents, E&P operation data, etc. The structured data is stored in database for online management; the index information of seismic data, well logging data and documents is also stored in database, and unstructured files are stored in database, near-line or offline storage devices as needed. It also supports metadata management, data maintenance, authorization, etc. to enhance the security and reliability of data assets.

Data service: it provides different functions for data service, such as GIS navigation, data browsing /downloading, statistics and analysis, data query, etc., supports one-click data loading from A1 2.0 to several kinds of E&P software (Openwork R5000, Petrel, GeoEast), or data downloading or pushing data onto a temporal storage zone to improve the efficiency of data service in this integrated researching environment; additionally, it also supports the data service for E&P management and decision-making, data transmission and sharing with other information systems.

Business platform: it supports different business applications, covering Exploration Management, Engineering Operation Management, and Comprehensive Research Support.

Based on the SOA architecture and Enterprise Application Integration methodology, the solution applies EPAI, an integrated, open integrated E&P Application Infrastructure platform, to support the integration of existing functions and development of new functions, provides various business components, including enterprise service bus, business process management(BPM), authorization, application service interface(ASI), and supports the customized development for the special requirements and features of different oil companies. A1 2.0 technical architecture is shown in [Figure 2](#).

EPAI (E&P Application Infrastructure), provides software development platform or environment and kinds of component packages, including,

- 1) EPAI-RT: a middleware for a service-oriented architecture, which provides system runtime management, data access service, unite service, event service, coordinate service and meta-model service, etc.
- 2) EPAI-DO: a component for management of domain model, which provides scalable business entities, and supports application of XML and other development environments, like J2EE, Dot net, etc.
- 3) EPAI-CC: common components including log management, data access permission, unit transform, multi-language supporting, etc.
- 4) EPAI-VC: web-based visualization components for E&P data, which provides wellbore diagram, well logging, 2D/3D seismic data, geologic & reservoir model, GIS-based map, etc.
- 5) EPAI-DC: a component for E&P data exchange, which provides major E&P applications data, formats analysis and converting tools.
- 6) EPAI-EC: a component for event management on E&P real-time data that provide basic functions including event listening, process, etc.
- 7) eBPM: a middleware for business process management.
- 8) eBUS: a middleware for enterprise application integration.
- 9) EPAI-Portal: a component for portal development in B/S model, which provides UI, login, registration, application portlets, and so on.

A distributed deployment architecture is applied in A1 2.0 project, database servers, application servers and related applications are deployed in different oil companies for their requirements and application; meanwhile, it also supports the data integration and backup in CNPC headquarters (CNPC Changing data center).

Key Techniques

The key techniques in E&P data management solution for A1 2.0 include E&P Data Model, Data Integration, Data Parsing, Data Quality Control (QC), Web-based Data Visualization, and so on.

E&P Data Model: The study and design of E&P Data Model (EPDM) is a core of unified data standards of CNPC. The design of EPDM 2.0 follows the basic design philosophy of ‘first concept model, then logical model, and last physical model’. Based on business modeling, EPDM 2.0 is realized in use of Object-Oriented system design, which is inherited from EPDM 1.0. Compared with EPDM 1.0, EPDM 2.0 has the following major changes:

- 1) The concept of E&P Master Data Management (E&P MDM) is added into EPDM 2.0 and the scope of basic entity is extended.
- 2) The business data specification is improved, all data specifications of different E&P domains consist of the business process as well as data specification.
- 3) The construction of data standard support measures is enhanced, including data model design specification, data model management specification and data model management system construction, supporting model acquisition, model browsing, model evaluation, model comparison, model evolution, model change, instance management, service management, attribute code management, etc.
- 4) The EPDM 2.0 is able to support the hierarchical management; the basic entity is taken as the top business object to carry on the unified management. The scope of business activities and data (of technical and operation management) generated from can be extended according to the specified rules, and the scope can be extended to the whole E&P activities.

EPDM 2.0 is composed of three parts: basic entity, technical data and operation management data. The basic entity includes two management entities of project and organization, and seven technical data entities of survey, geological unit, production unit, station, well, wellbore and equipment. The business scope of technical data includes geophysics, drilling, mud logging, logging, hydrocarbon test, well test and production test, downhole operation, sample experiment, regional geology and single well geology. The business scope of operation management data includes oil and gas production, production test, production stimulations, production engineering, ground engineering, equipment management, material management, team management, personnel management, supervision management.

Data Integration: The purpose of data integration is to solve the problem generated by multi-source data acquisition, transmission and load to ensure that the source data acquisition in A1 2.0 is timely and efficient. The data integration has different methods.

Data Entry/Import: A1 2.0 provides functionalities of web-based data entry and import to realize data remote submission, and support the process of data parsing, quality control, quality audit, loading and storing.

Data Load: A1 2.0 provides functionalities of web-based data load to realize remote submission of bulk data such as seismic data (except for pre-stack seismic data), logging data, and support the process of dump, parsing, quality control, auditing, loading and so on.

Data Synchronization: Based on technologies of model mapping, data transmission (supporting WITSML) and so on, ETL (Extract, Transfer, Load) method is used to realize data synchronization from third party database (production database, engineering operation management database and other related databases), including heterogeneous database, to A1 2.0. This method is also used for data migration from A1 1.0 to A1 2.0.

Data Parsing: In E&P data, a lot of data are unstructured data, such as seismic data, well logging data and comprehensive research results, and is saved in files of different formats. These formatted files are needed to be data parsed to realize data browsing, import / export, visualization, format conversion and so on, which is frequently used in data loading, data quality control and data visualization of A1 2.0. Hence, Data Parsing is one of key techniques in A1 2.0 and necessary data parsing tools are provided to support: Data parsing of seismic data in SEG-Y, SEG-D, RODE, etc.; Data parsing of measurement data in SPS, UKOOA, etc.; Data parsing of VSP in DLIS format; Data parsing of logging

data in ASCII, CLS, DLIS, LAS, LIS, WIS, XTF, 716, etc. In addition, parsing tools can be customized development on individual demands to support other formats (for 2D/3D gridding file, specific ASCII files, etc.) to support data parsing of seismic velocity, seismic interpretation, geological / reservoir models and so on.

Data Quality Control: The main elements of data quality include normalization, integrity, accuracy, consistency, timeliness and so on. A1 2.0 covers a wide range of domains and involves many business processes. The objects of data quality management (data content) cover 11 E&P domains, in which the data can be classified in two major types, i.e., structured data and unstructured data. Based on the life-cycle data management and quality control mechanism, A1 2.0 provides data parsing, quality rule definition, scanning engine, quality scanning, visualization and other means to implement data quality control.

For structured data in A1 2.0, according to the EPDM standard, combined with the business characteristics and needs of oil companies as well as the data quality management requirements of A1, corresponding quality control strategies, schemes and business rules are configured ([Figure 3](#)) to meet individual needs by applying the Informatica data quality control six-step method. The main idea of this method is, according to business rules and related technical standards, to establish the rule base of quality and management to implement the quality control of A1 data during the stages of pre-storing, storing and post-storing, and generate the quality report, monitoring report, quality communiqué, to achieve the data quality control requirements.

For unstructured data in A1 2.0, it is closely related to its disciplines or professional domain in terms of data content, associated information, execution standards, file format, etc., therefore, in the construction of A1 2.0, the corresponding quality control methods and technologies are developed according to their own characteristics, technical standards.

Web-based Data Visualization: In the life cycle of E&P data management, data acquisition, data quality control, data service and so on cannot be separated from data visualization. A1 2.0 provides rich Web-based data visualization functions, so that users can realize the data display in the browser environment, including: seismic section, wellbore trajectory, mud logging bar graph, logging curve and image, sample experiment cross-plot, geological & reservoir map, etc.

Implementation Results

From August 2013 to December 2016, with the process of different project stages, including site investigation, requirements analysis, details design, software development, testing, pilot (Dagang, Changqing), and implementation in different companies of CNPC, A1 2.0 project has been completed fully, and the goals of this project have been archived as well.

Results

- 1) Established the specifications and standards of E&P data model, including data specifications of 6 disciplines, EPDM V2.0 (18793 data items in 562 data tables totally), and the portal of E&P data model for long-term promotion of data model.

- 2) According to the concept of E&P data governance, set the E&P data management system for data collection, management and service to support the workflow of E&P business data management, and got the certification of ISO9001.
- 3) According to the goal of IT systems integration in CNPC 12th Five-year plan, set up the cloud application environment, and realized different functions including data query, statistics & report, knowledge management and seamless data service for E&P software in research environment.
- 4) Realized the centralized data management for huge, distributed legacy E&P data, and enhanced the ability of E&P technical data assets management: 700,000,000 records of wellbore data, 1,080,000 files of well logging data, more than 7,000 seismic surveys, and more than 3,670,000 copies of documents.
- 5) Established a full life-cycle workflow for data quality control, to ensure data quality including completeness, correctness, uniqueness and normalization.
- 6) Built the A1 2.0 integration system in CNPC headquarters as the remote backup and recovery system, to support the requirements of integrated application in CNPC group level and enhance the data security.

Effects

With the implementation of A1 2.0, it brings with the following main effects:

- 1) The distribution way of data loading and management of A1 1.0 is changed and improved, and an integrated, unified and extensible data platform is formed to support the whole workflow from data acquisition, data management to data service.
- 2) Evident application effects are shown: By the end of November 2017, registered users of A1 2.0 are more than 10,000, and the data size in A1 2.0 is more than 1,600TB, with seismic surveys more than 7,000, well loggings more than 300,000, wellbores more than 480,000. A1 2.0 brings lots of benefits, such as effective protection of core data assets, efficient data service for comprehensive research, and improvement of E&P research quality and effect.
- 3) A1 2.0 data platform has been becoming the E&P data asset repository in different oil companies, strongly supporting the comprehensive geological research and E&P operation management, and enhanced the data integration and application. In 2017, A1 2.0 provided data service for 695 E&P research projects in 16 oil companies, including more than 470TB seismic data, 430,000 times of wellbore data and more than 500,000 documents.

Conclusions

The objectives of A1 2.0 project have been achieved through the implementation of more than three years and benefits show evident: A brand-new, integrated CNPC E&P technical data management platform is built up and it make CNPC get rid of dependence on the similar products only from abroad; The deployment of A1 first covers all oil & gas companies of CNPC; Integrated data asset is up to 1.6PB, three times of the size stored in A1 1.0; The registered users are more than ten thousands, five times of the scale of A1 1.0; A1 2.0 is gradually becoming the unified E&P data asset management system of oil & gas companies of CNPC and provides a strong support for E&P applications.

Selected Reference

Yan, Y.S., T. Ma, T.C. Wang, and Z.K. Xu, 2012, Difficulties and strategies of integrated reservoir studies: 74th EAGE Conference & Exhibition incorporating SPE EUROPEC, 4 June 2012, doi:10.3997/2214-4609.20148271

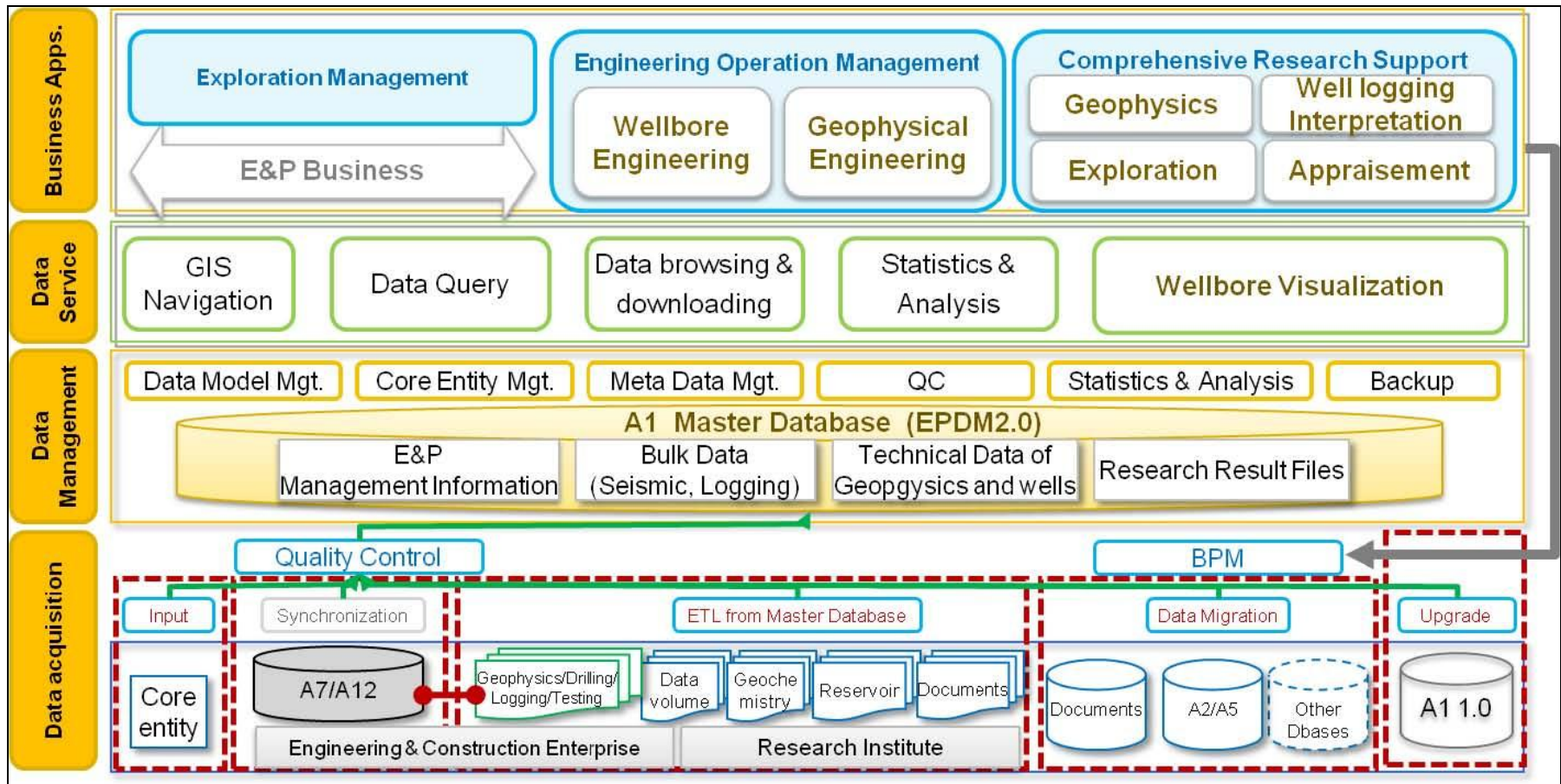


Figure 1. Overall architecture of A1 2.0. Note: A2: CNPC Oil & gas & water Operation Management System; A5: CNPC Surface Engineering Operation Management System; A12: CNPC Engineering Technology IoT System.

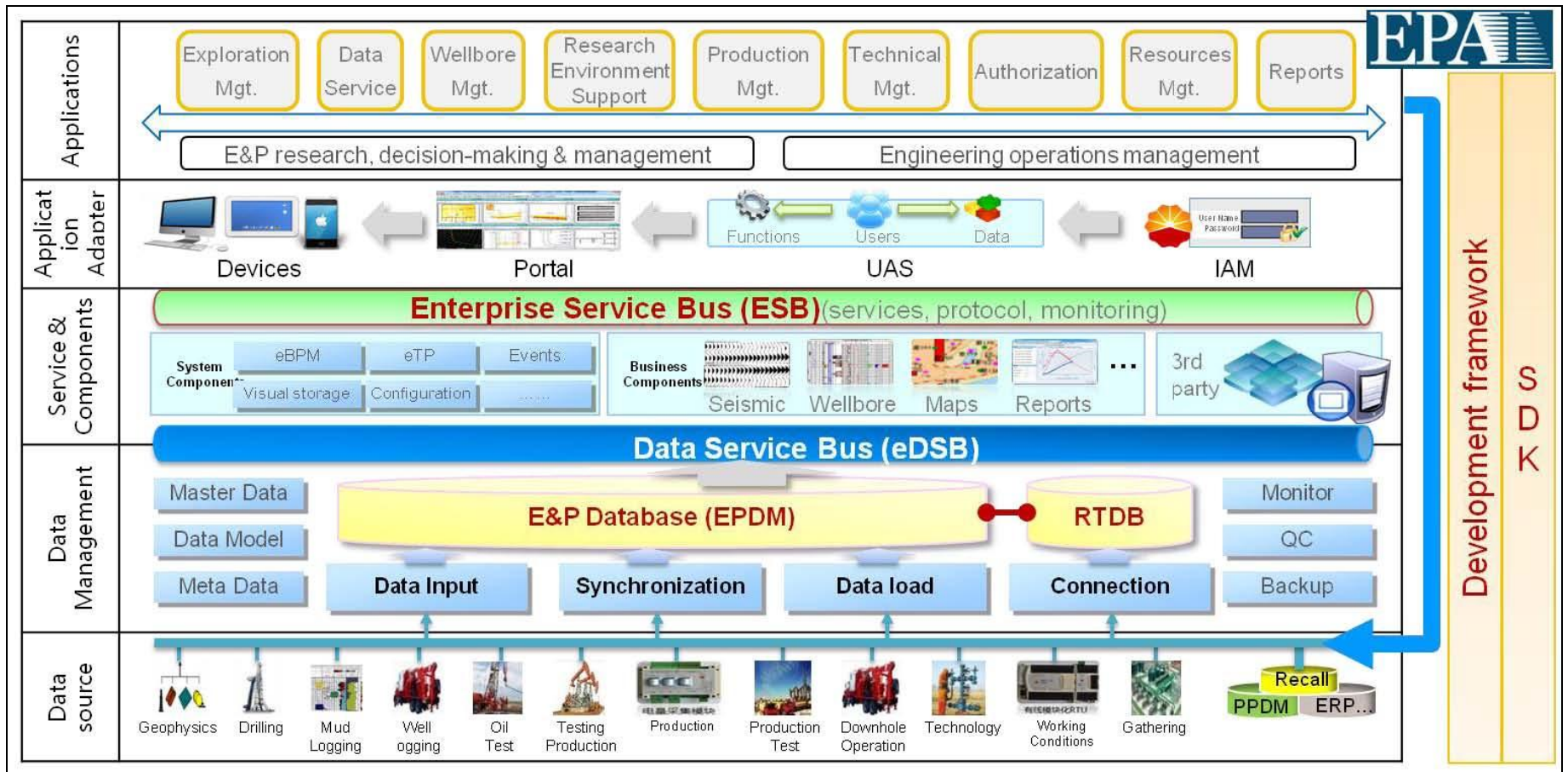


Figure 2. Technical architecture of A1 2.0.

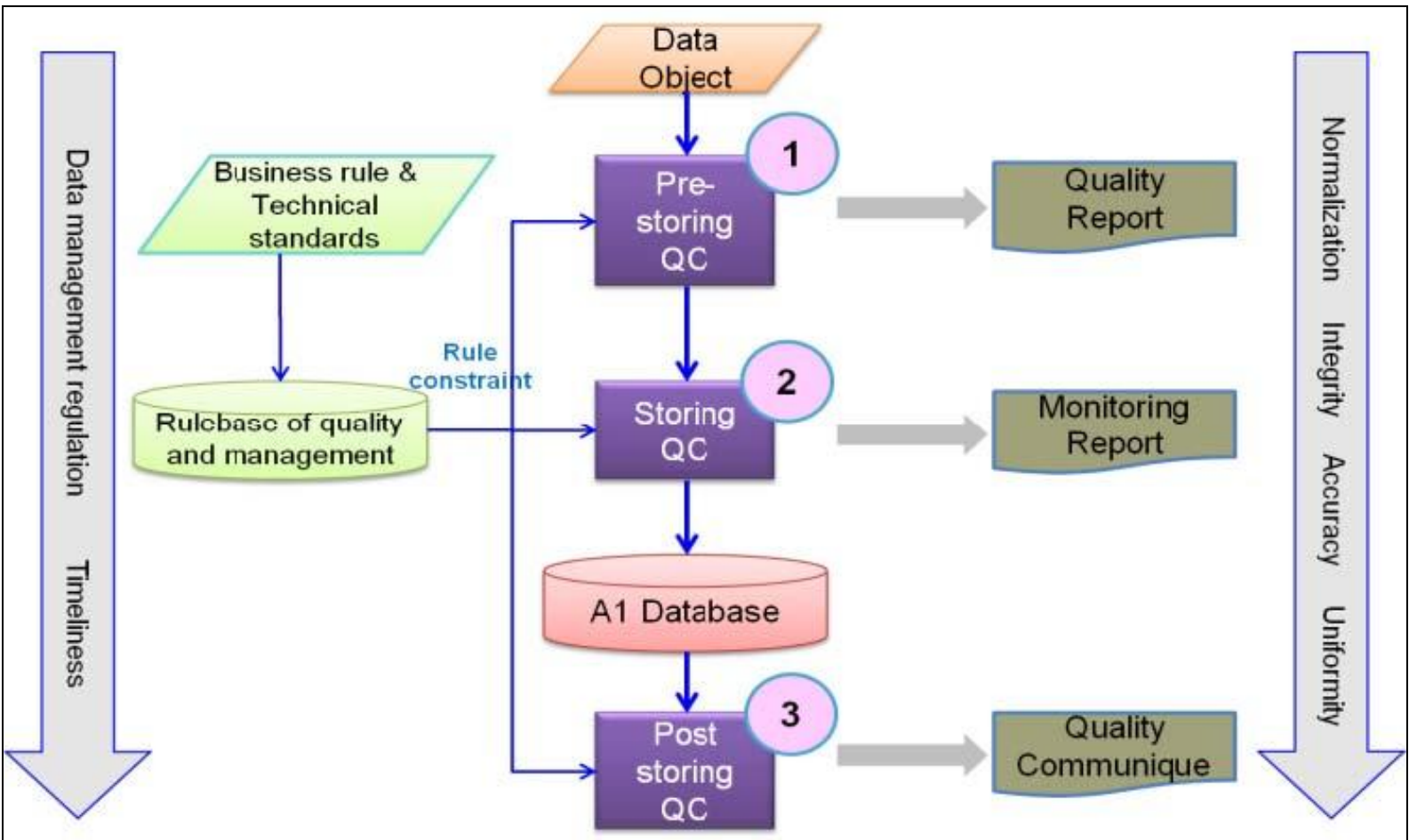


Figure 3. Flow chat for structured data quality control strategy and process.