

데이터 품질진단 솔루션

데이터쿼크

DataQuark

v2.0

2025년 1월



CONTENTS

01

데이터쿼크
제품 소개



제품 개요
제품 용도/목적

02

데이터쿼크
주요 기능



다양한 진단 기능
진단쿼리 자동화 기능

03

데이터쿼크
활용사례



예상수용처
활용사례

데이터 품질진단 솔루션 DataQuark

제품 개요

품질진단 솔루션인 데이터쿼크(DataQuark)는 사용자 친화적인 인터페이스로 누구나 손쉽게 마우스 클릭만으로 데이터 품질진단을 수행할 수 있는 획기적인 품질진단 솔루션입니다.

1 다양한 DB 지원

- 알티베이스, 큐브리드, 마리아DB, 오라클, MySQL, PostgreSQL, MS-SQL, 티베로, SQLite 등 11개 DB 품질진단 지원
- **오픈포맷 지원(CSV, JSON, XML)**

3 다양한 표준사전 지원

- 공공데이터 공통 표준사전 (행안부)
- 빅데이터 연계 표준사전 (NIA)
- DB별 표준사전 자동구축 지원
- 단어, 용어, 도메인 표준사전 구축

2 진단규칙 자동생성

- 7개 지표별 도메인 자동 분류
- 값/BR 진단규칙 자동생성
- 정규표현식 자동 생성
- 제한없는 커스텀 진단 지원

4 비정형 데이터 지원

- 비정형 파일의 메타데이터 자동 추출 (텍스트/이미지/오디오/비디오 파일)
- AI학습용 이미지 어노테이션 검사 (VGG, CVAT, MakeSence, Labellmg 지원)



데이터 품질진단 솔루션 DataQuark

용도 및 목적

정보시스템의 데이터 품질현황을 분석하기 위하여 고객과 사전 합의된 정보시스템에 대하여 DataQuark를 활용하여 값진단, BR(업무규칙)진단, 표준진단, 구조진단을 수행하고 진단결과에 대한 원인분석 및 개선방향을 제시하여 고품질 데이터관리의 근거를 마련합니다.

진단도구



사용 환경

- 제품명 : DataQuark v2.0
- 사용환경 : Win10 이상의 PC 및 노트북
- 행안부(공공데이터) 및 K-DATA 품질 기준 적용

용도 및 목적



주요 용도

- 도메인 기반 품질진단 지원
- 값진단, BR(업무규칙)진단 지원
- 행안부 표준진단, 구조 진단 지원
- 공통코드 및 개별코드 진단 지원

주요 기능



차별화된 품질 진단 기능

- Oracle 등 11개 DB 진단
- CVS, JSON, XML 등 오픈포맷 지원
- 도메인 자동분류 / 진단규칙 자동 생성
- 강력한 스키마 수집 및 프로파일링 지원

기본 품질 진단 지표

- | | |
|---|---|
| <input checked="" type="checkbox"/> 여부유효성 | <input checked="" type="checkbox"/> 범위유효성 |
| <input checked="" type="checkbox"/> 코드유효성 | <input checked="" type="checkbox"/> 명칭유효성 |
| <input checked="" type="checkbox"/> 날짜유효성 | <input checked="" type="checkbox"/> 내용유효성 |
| <input checked="" type="checkbox"/> 형식유효성 | |

세부 품질 진단 지표

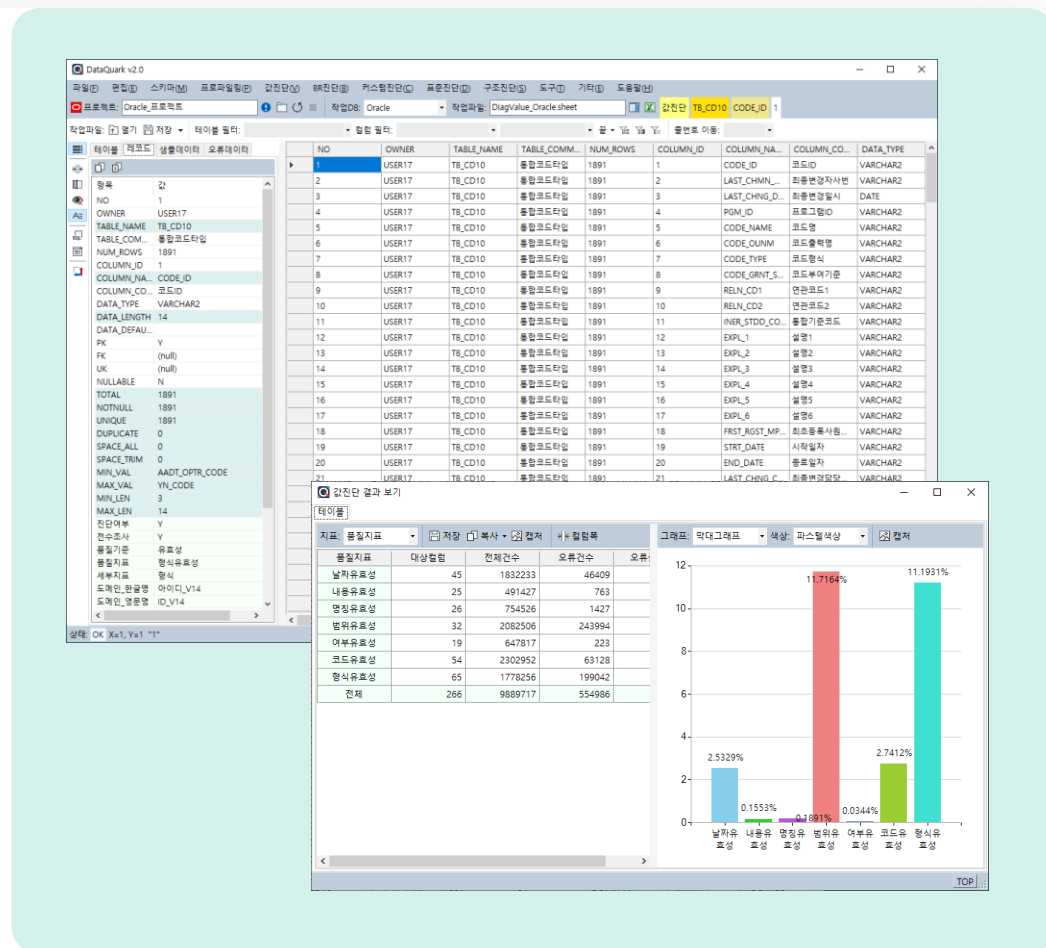
- | | | | |
|--|--|--|--|
| <input checked="" type="checkbox"/> 여부 | <input checked="" type="checkbox"/> 날짜 | <input checked="" type="checkbox"/> 금액 | <input checked="" type="checkbox"/> 내용 |
| <input checked="" type="checkbox"/> 유무 | <input checked="" type="checkbox"/> 시간 | <input checked="" type="checkbox"/> 비율 | |
| <input checked="" type="checkbox"/> 코드 | <input checked="" type="checkbox"/> 형식 | <input checked="" type="checkbox"/> 수량 | |
| <input checked="" type="checkbox"/> 분류 | <input checked="" type="checkbox"/> 번호 | <input checked="" type="checkbox"/> 명칭 | |

데이터 품질진단 솔루션 DataQuark

품질진단 내용

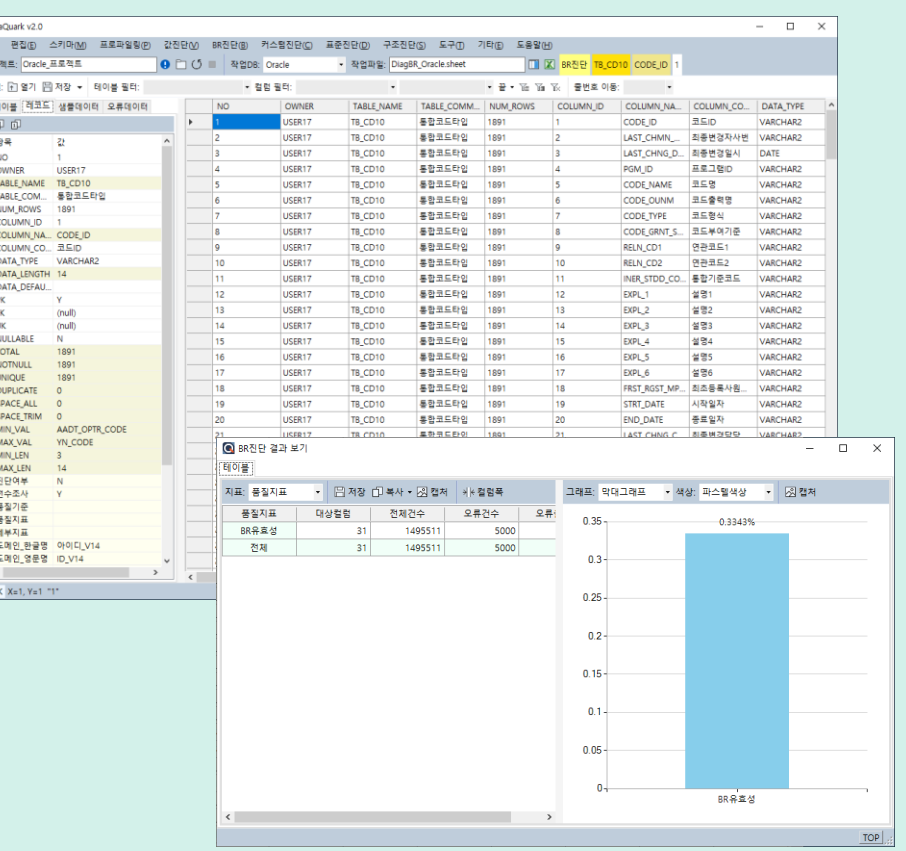
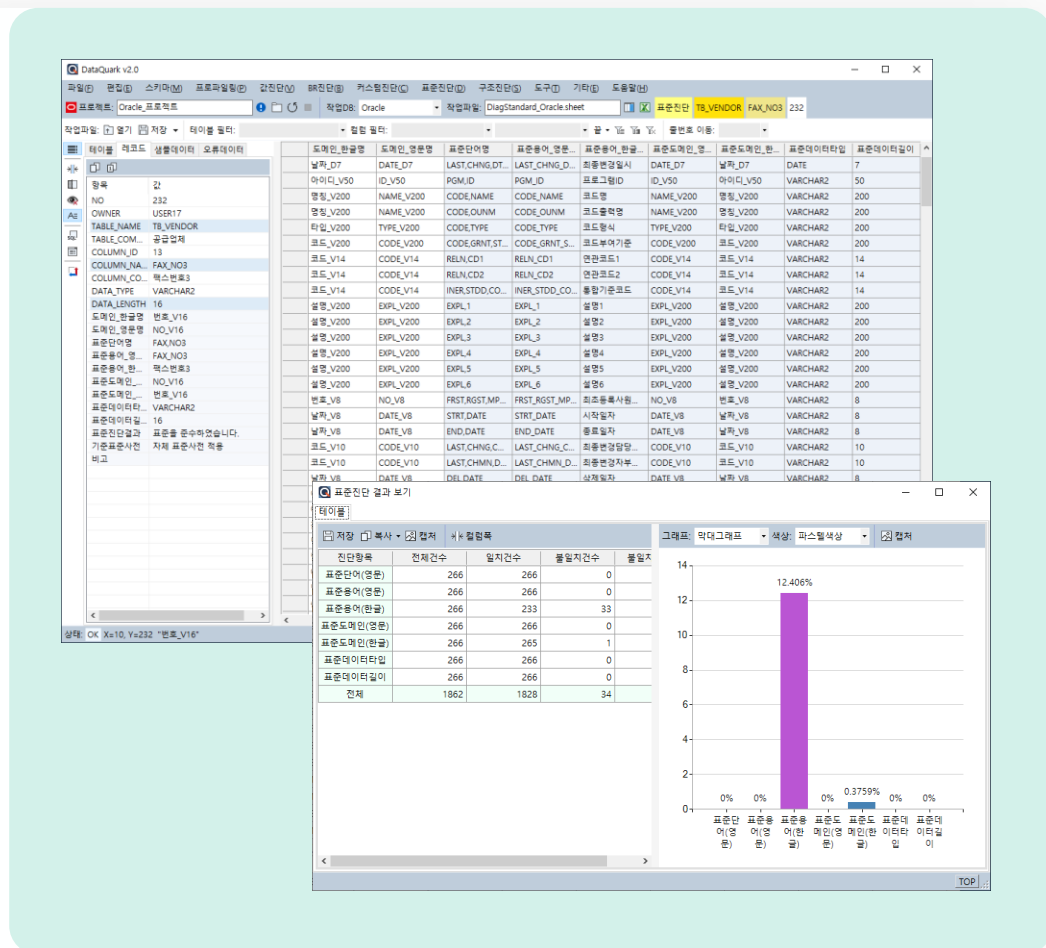
값 진단

- 날짜, 여부, 수량 등 범정부 진단 기준에 맞는 도메인 선정
- 진단기준에 따른 진단규칙 생성
- 검증규칙에 맞게 진단수행 후 오류율 산출



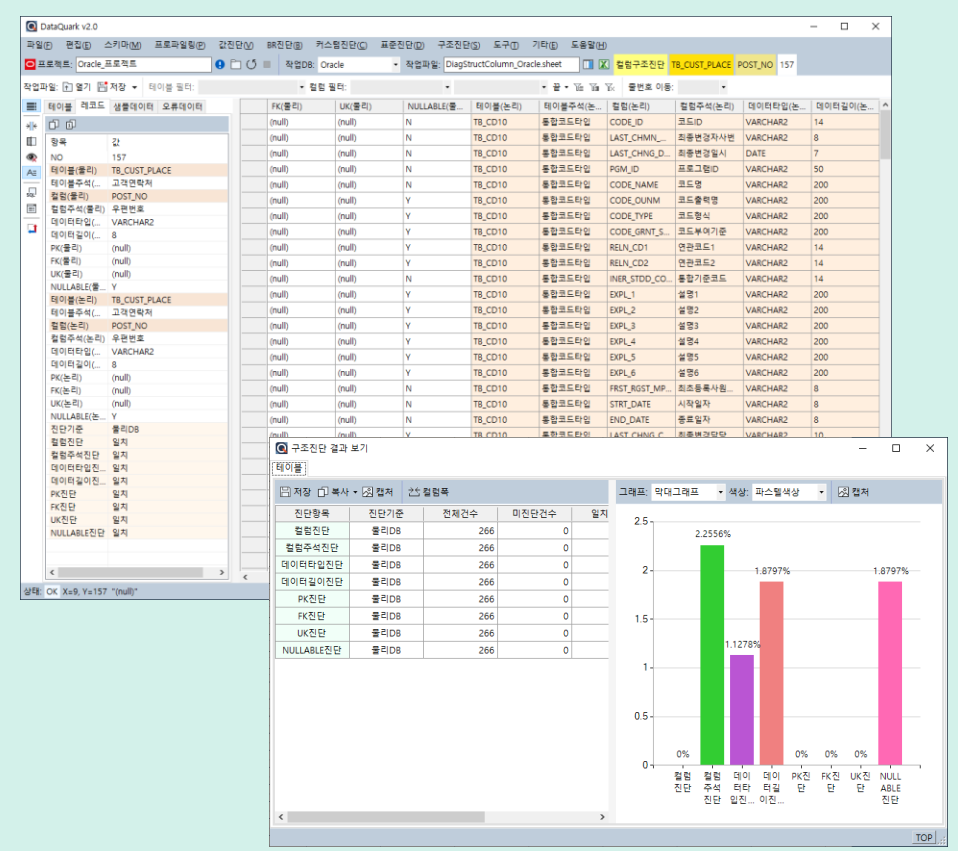
표준 진단

- 현행 DB의 컬럼명의 표준용어 적용률 진단
- 표준용어에 따른 표준도메인 적용률 진단
- 표준적용률을 산출하여 표준수준 진단



업무규칙(BR) 진단

- 선후, 수식, 동등, 논리관계 등 기준에 맞는 업무규칙 선정
- 날짜, 시간, 금액 등 업무규칙 자동생성
- 업무규칙에 맞게 진단수행 후 오류율 산출

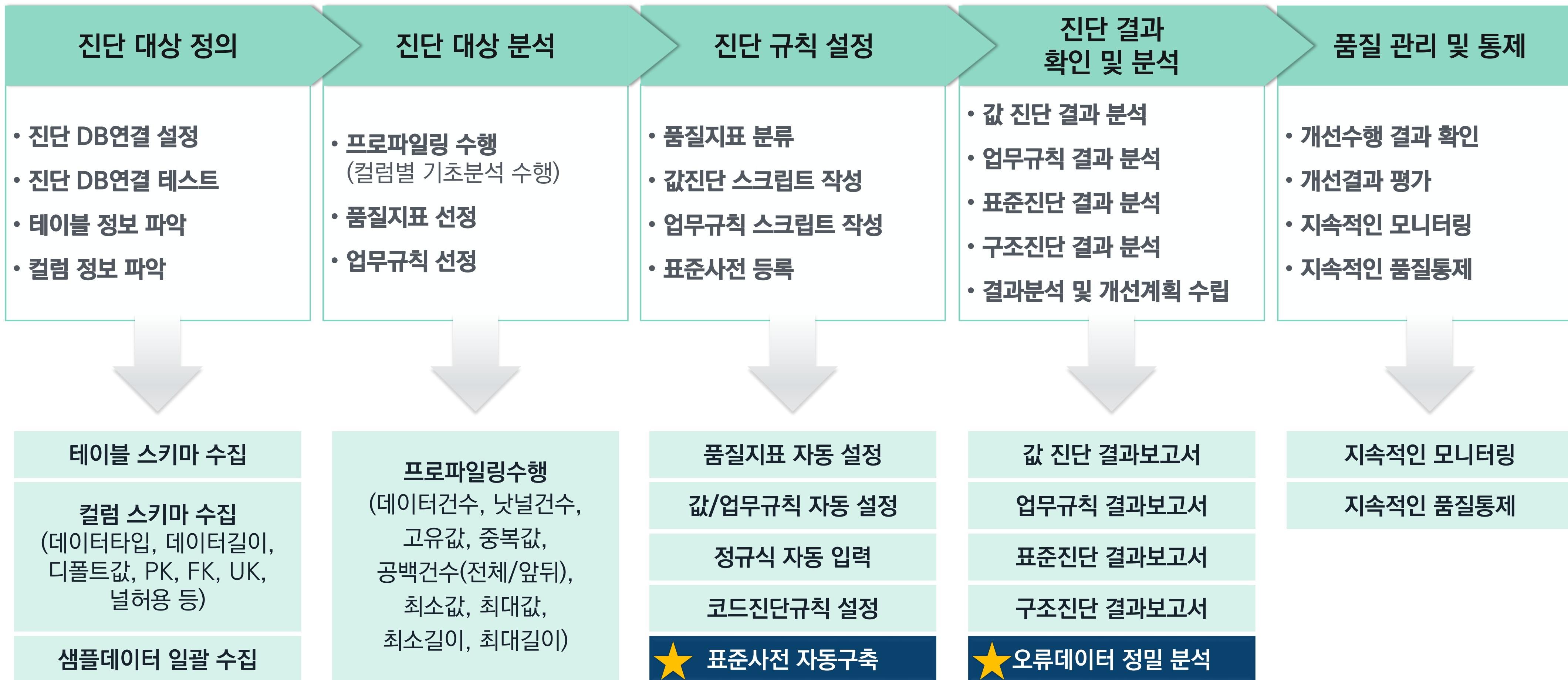


구조 진단

- 테이블정의서와 현행 DB의 교차검증비교
- 칼럼정의서와 현행 DB의 교차검증비교
- 적용률을 계산하여 가장 낮은 항목이 점수로 산출

데이터 품질진단 솔루션 DataQuark

진단 프로세스



데이터 품질진단 솔루션 DataQuark

자동화 기능

DataQuark는 품질진단 도메인 자동 선정, 진단규칙 자동 입력, 표준사전 자동 구축 등 품질관리체계 구축에 있어서 필요한 사전 작업들을 자동화 기능을 통해 빠르게 수행할 수 있도록 다양한 기능을 지원하고 있습니다.

도메인에 따른 진단 규칙이 자동으로 입력됩니다.

진단 수행후 오류건수에 해당하는 실제 오류를 곧바로 확인할 수 있습니다.

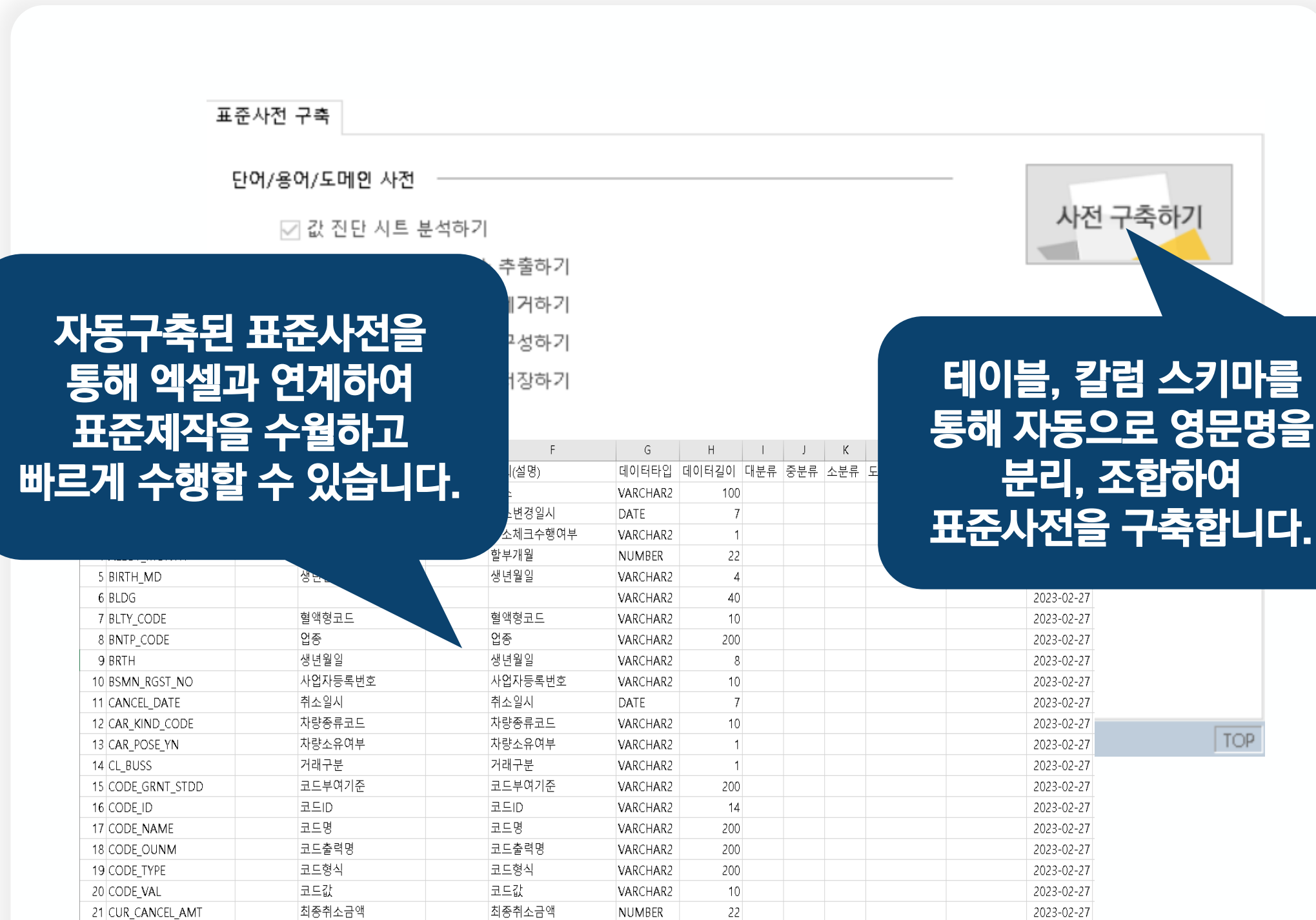
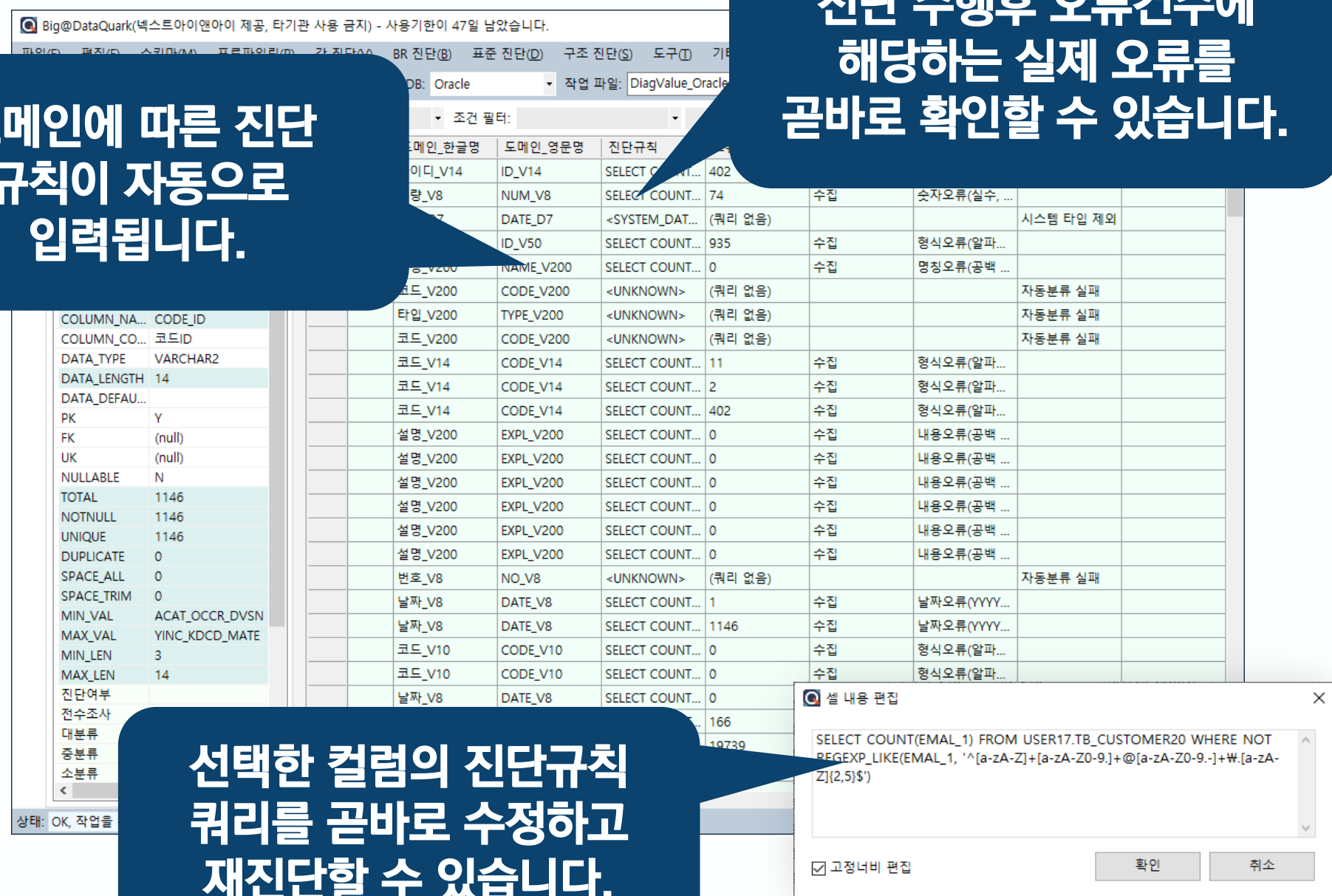
선택한 컬럼의 진단규칙 쿼리를 곧바로 수정하고 재진단할 수 있습니다.

[품질진단 기준 자동등록]

자동구축된 표준사전을 통해 엑셀과 연계하여 표준제작을 수월하고 빠르게 수행할 수 있습니다.

테이블, 칼럼 스키마를 통해 자동으로 영문명을 분리, 조합하여 표준사전을 구축합니다.

[표준사전 자동구축]



데이터 품질진단 솔루션 DataQuark

다양한 진단규칙 내장

DataQuark는 그 동안 제안사에서 품질진단을 수행하면서 경험했었던 다양한 진단규칙을 정규식 형태로 제공합니다. 이를 통하여 사용자는 자동으로 진단쿼리를 즉시 생성할 수 있습니다.

The screenshot shows the DataQuark configuration interface. At the top, it indicates the target database is Oracle. Below, there are various configuration options like '진단여부' (Y), '전수조사' (Y), and '대중소분류' (유효성). A dropdown menu is open, listing various built-in rules such as '여부(YN)', '유무(01)', '남녀(WM)', '숫자코드', '문자코드', '알파벳(소문자)', '알파벳(대문자)', '알파벳(숫자 포함)', '숫자(Digit)', '자연수(1,2,3,...)', '자연수(0,1,2,3,...)', '정수', '실수(소수점 포함)', '실수(소수점, 1자리)', '실수(소수점, 1~2자리)', '실수(+, 소수점 포함)', '날짜(년월일)', '날짜(년-월-일)', '날짜(년/월/일)', '날짜(년월)', '날짜(년-월)', '시간(시분초)', '시간(시:분:초)', '시간(시:분)', and '시간(시:분)'. A callout box highlights the '정규식' (Regular Expression) tab, showing a text input field with '^[YN]\$' and a dropdown menu with '여부(YN)'. Below this, there are sections for '샘플 값을 입력하십시오.' (Enter sample values), '자동으로 생성된 정규식입니다.' (Regular expression automatically generated), and '생성된 정규식을 테스트 합니다.' (Test the generated regular expression).

정규식 자동 작성기를 제공하며 테스트도 가능합니다.



날짜 유효성

- YYYYMMDD
- YYYY-MM-DD
- 기타 날짜 형식 등



번호 유효성

- 전화번호
- 주민등록번호
- 기타 번호패턴 등



범위 유효성

- 값 범위
- 길이 범위
- 목록 범위
- 기타 범위 설정 가능



여부 유효성

- Y/N
- 0/1
- 기타 여부 및 리스트형
- 데이터 등

데이터 품질진단 솔루션 DataQuark

빅데이터 플랫폼 진단

DataQuark는 자동화된 품질진단 솔루션으로서 21개 빅데이터 플랫폼 및 80여개 센터 등 100여곳에서 진단을 수행하였으며 국방기술품질원, 시험인증 빅데이터 플랫폼, 한국기술교육대학 등에서 성공적으로 품질진단을 수행하는데 기여를 했습니다.

감염병 한국과학기술연구원 국립중앙의료원 이화여자대학교 용인시산업진흥원 더브레인	공간융합 한국국토정보공사 (주)지앤티솔루션 공간정보기술(주) 한컴어셈텍 제타력시스템	부동산 한국부동산원 한국교통연구원 공간정보산업진흥원 스페이스워크 피타그래프	스마트팜 네이버클라우드 한국산업기술시험원 경남테크노파크 웰니스파머(주) (주)유비엔	연안 한국해양과학기술원 국립해양조사원 국립수산과학원 인하대학교 (주)해양정보기술	농식품 한국농수산식품유통공사 와이즈레이크 NICE지니데이터(주) 한국무역통계진흥원 케이플러스	디지털산업혁신 한국산업기술시험원 에프엔가이드 한국M&A거래소 벅스인테리전스 광개토연구소
라이프로그 원주세브란스기독병원 고려대의료원 대한청각협회 한림대병원 LG유플러스	소방안전 소방청 서울소방본부 부산소방본부 한국소방산업기술원 한국화재보험협회	스마트치안 경찰대학치안정책연구소 수원시청 서울신용한국토지주택공사 보증재단 지란지교데이터	해양수산 한국해양수산개발원 해양교통안전공단 광주과학기술원 울산과학기술원 올시데이터	금융 비씨카드 닐슨코리아 부산광역시 여기어때컴퍼니 한국감정평가사협회	환경 한국수자원공사 국립생태원 한국지질자원연구원 한국해양과학기술원 국립중앙과학관	문화 한국문화정보원 국립중앙도서관 국민체육진흥공단 한국문화예술위원회 제주관광공사
교통 한국교통연구원 울산정보산업진흥원 포항테크노파크 한국부동산원 한국철도기술연구원	통신 케이티 한국인터넷진흥원 인천테크노파크 소상공인연합회 경기대학교	유통소비 매일방송 NICE디앤알 경동도시가스 코리아크레딧뷰로 한국우편산업진흥원	헬스케어 국립암센터 삼성서울병원 건양대학교병원 아주대학교병원 서울대학교산학협력단	중소기업 비즈온 한국무역정보통신 녹색기술센터 한국경영인증원 한국물가정보	지역경제 경기도청 한국기업데이터 한국생산기술연구원 한국신용데이터 국토연구원	산림 한국임업진흥원 한국산림복지진흥원 데이터사이언스랩 한국한의학연구원 삼아항업(주)

데이터 품질진단 솔루션 DataQuark

활용사례

DataQuark의 자동화된 품질진단 기능은 데이터 품질진단을 처음 수행하거나 마이그레이션, 차세대 시스템 구축 등 기존의 DB를 재구축하는 과정에서 예방적 품질진단을 수행하는데 경제적, 시간적 비용을 크게 절감하였습니다.



별첨. 데이터 품질진단 솔루션 DataQuark

경쟁제품 비교

구분		S제품	W제품	D제품	DataQuark
사용환경	윈도우 지원	O	O	O	O
	리눅스 지원	O	O	O	X
기능성	스키마 수집	O	O	O	O
	프로파일링 지원	O	O	O	O
	값진단 지원	O	O	O	O
	BR진단 지원	O	O	O	O
	표준진단 지원	O	O	O	O
	구조진단 지원	O	O	O	O
	코드진단 지원	O	O	O	O
	커스텀진단 지원	X	X	X	O
	사용성	정형 다양한 DB 지원	O	O	O
반정형 오픈포맷 지원	X	X	X	O	
비정형 메타데이터 추출	X	X	X	O	
	어노테이션 검사	X	X	X	O
편의성	도메인 자동분류	△	△	△	O
	진단규칙 자동생성	△	△	△	O
	SQL 편집기 제공	△	△	△	O
	외부쿼리 가져오기	△	△	△	O
	외부데이터 가져오기	△	△	△	O
품질인증	외부기관 인증	GS인증 1등급	GS인증 1등급	GS인증 1등급	GS인증 1등급
사업실적	금융 차세대 및 공공기관	다수의 차세대 사업 수행 및 금융권 사업수행 실적	공공데이터 수준평가용 도구로 활용	모델링 등 다양한 연계 자동화로 국내 최대 고객 보유	23' 빅데이터 플랫폼 및 24' 개방데이터 진단수행

별첨. 데이터 품질진단 솔루션 DataQuark

제품버전 비교

구분		DataQuark 1.0	DataQuark 2.0 Standard	DataQuark 2.0 Professional
	스키마 수집	O	O	O
	프로파일링 지원	O	O	O
	값진단 지원	O	O	O
	BR진단 지원	O	O	O
	표준진단 지원	O	O	O
	구조진단 지원	O	O	O
	도메인 자동 분류	O	O	O
	진단규칙 자동 생성	O	O	O
	외부쿼리 가져오기	X	O	O
	외부데이터 가져오기	X	O	O
	오픈포맷	X	X	O
	커스텀진단 지원	X	X	O
	표준사전 구축	X	X	O
	코드진단 지원	X	X	O
	메타데이터 추출	X	X	O
	어노테이션 검사	X	X	O
	DB 개수	4개	8개	11개
지원 DB	DB 종류	MariaDB, MS-SQL, MySQL, Oracle	Altibase, CUBRID, MariaDB, MS-SQL, MySQL, Oracle, SQLite, Tiberio	Altibase, CUBRID, DB2, MariaDB, MS-SQL, MySQL, Oracle, PostgreSQL, SQLite, SAP(Sybase), Tiberio

별첨. 데이터 품질진단 솔루션 DataQuark

DataQuark 커스터마이징 지원

사업기간내 진단도구 무제한 제공
(DataQuark Professional + 1년 무상 제공)

통합 엑셀문서 지원(*.xlsx)
(모든 진단결과를 통합된 엑셀 문서로 저장)

공공기관 대상 진단도구 배포 및 기술지원
(설치프로그램 및 사용 매뉴얼 제공)

오픈포맷(JSON, XML) 진단 지원
(데이터 형식 및 파싱트리 검사, 데이터 추출 검사)

진단도구의 원활한 사용을 위한 교육지원
(온오프라인 교육: 원격 및 방문교육)

미지원 및 비정형 DB 진단 지원(하둡)
(Export 데이터를 로컬 DB로 이관하여 진단)

감사합니다

Create the Future with
Trust Technology