

CHECKLIST

BIG DATA MANAGEMENT REQUIREMENTS

Big Data management is the genus to which Big Data processing and analytics are species. For developers, data scientists, and other data professionals, keeping every data management requirement in mind is not an easy task. This checklist will help you navigate critical requirements for consistently and flexibly curating raw sources of Big Data into trusted assets for next-generation analytics.

BIG DATA INTEGRATION

HIGH PERFORMANCE INGESTION	SCALABLE PROCESSING	FLEXIBLE DEPLOYMENT
<input type="checkbox"/> Purpose-built adapters for wide variety of data sources	<input type="checkbox"/> Parameterized pipeline development	<input type="checkbox"/> Deployment-agnostic pipeline abstraction
<input type="checkbox"/> Mass ingestion of high-volume and changed data	<input type="checkbox"/> Purpose-built components for data parsing and transformation	<input type="checkbox"/> Resource-optimized pipeline execution
<input type="checkbox"/> Real-time streaming for low-latency data	<input type="checkbox"/> End-to-end pipeline processing and data provisioning	<input type="checkbox"/> Deployment on cloud or on-premise

BIG DATA GOVERNANCE + QUALITY

COLLABORATIVE GOVERNANCE	360 VIEW OF RELATIONSHIPS	FIT-FOR-PURPOSE DATA CONFIDENCE
<input type="checkbox"/> Role-based data stewardship	<input type="checkbox"/> Universal metadata catalog	<input type="checkbox"/> Anomaly detection through data profiling
<input type="checkbox"/> Keyword and faceted data search	<input type="checkbox"/> Data matching and record linking at scale	<input type="checkbox"/> Data quality using purpose-built components
<input type="checkbox"/> Team-based data publishing and sharing	<input type="checkbox"/> Data relationship tracking and inference	<input type="checkbox"/> Data provenance through end-to-end lineage

BIG DATA SECURITY

360 VIEW OF SENSITIVE DATA	RISK ANALYTICS OF SENSITIVE DATA	POLICY-BASED PROTECTION OF SENSITIVE DATA
<input type="checkbox"/> Understanding of sensitive data protection and cost	<input type="checkbox"/> Risk modeling and score trending by residency, usage, and proliferation	<input type="checkbox"/> Non-intrusive de-identification of sensitive data
<input type="checkbox"/> Automated policy-based profiling and discovery	<input type="checkbox"/> Data proliferation tracking and monitoring	<input type="checkbox"/> Centralized security policy management
<input type="checkbox"/> Visibility of sensitive data by residency, usage, and proliferation	<input type="checkbox"/> Detection, analysis, and alerting of high risk users	<input type="checkbox"/> Support for on-premise and cloud

IN08_0915_02979