Data Management and Analytics

Previously two separate KSA lists, Data Management and Analytics have been combined as of 2026. This is in direct response to feedback from members of industry, who shared that with the onset of generative AI, most entry-level data roles will require literacy in skillsets across both of these two specializations.

Employer Signaling System by the Greater Washington Partnership

Generated: November 20, 2026

Data Management, Governance, and Foundations

Label	KSAC Description	KSA	Bloom's Taxonomy Level
a	Define and appropriately use key data concepts including data engineering, data repositories, metadata, data governance, and data lifecycle	knowledge	2
b	Define and appropriately use AI-related data concepts including machine learning, model drift, training data, inference data, and data leakage, particularly as they relate to analytics and automated decision-making	knowledge	2
c	Explain the purpose and structure of data governance programs, including how governance supports data quality, consistency, compliance, and trust across analytics and AI-enabled use cases	knowledge	3
d	Explain the role of change management in data governance, including how policies, standards, and definitions are introduced, maintained, and adopted over time	knowledge	2
e	Describe roles within data governance frameworks (e.g., data owner, steward, producer, consumer), including how responsibilities differ across data creation, maintenance, and use	knowledge	2
f	Explain the importance of metadata, lineage, and provenance across analytics and AI workflows, including their role in reproducibility, auditability, and model reliability	knowledge	2
g	Identify governance risks introduced by AI-assisted data use, such as loss of transparency, inappropriate reuse of data, or misalignment with original data purpose	knowledge	3

Label	KSAC Description	KSA	Bloom's Taxonomy Level
h	Explain the concept of a system of record and golden source data, and why these distinctions matter for analytics, reporting, and AI training data	knowledge	2
i	Identify basic governance issues across the analytics lifecycle, including data access, versioning, reuse, and downstream dependency management	knowledge	3

Data Storage, Platforms, Operations, and Access

Label	KSAC Description	KSA	Bloom's Taxonomy Level
a	Explain the role of data storage systems in organizational decision-making	knowledge	1
b	Define key storage concepts including databases, data lakes, data warehouses, and lakehouses	knowledge	2
c	Differentiate structured, semi-structured, and unstructured data storage	knowledge	3
d	Describe on-premises, cloud, and multi-cloud environments and their use cases	knowledge	2
e	Explain how cloud platforms separate storage and compute and the implications of this separation for scalability, cost control, and analytics performance	knowledge	3
f	Describe how data architecture affects analytics performance	knowledge	2
g	Identify common DBMS platforms and their roles	knowledge	2
h	Explain database performance considerations and tuning concept	knowledge	2
i	Demonstrate the ability to access data using SQL or similar query interfaces to retrieve, filter, and aggregate data	skill	3
j	Modify and refine AI-generated queries for correctness, efficiency, and alignment with business logic	skill	3

Label	KSAC Description	KSA	Bloom's Taxonomy Level
k	Identify inefficient, incorrect, or risky query patterns produced by automation, such as unnecessary scans or unintended joins	ability	4

Data Modeling and Design

Label	KSAC Description	KSA	Bloom's Taxonomy Level
a	Describe the systems development life cycle (SDLC) and its relevance to data and analytics systems	knowledge	2
b	Explain the purpose of data modeling in analytics systems, including how models translate business requirements into technical structures	knowledge	2
c	Identify relational, NoSQL, and analytical data models and the types of use cases each supports	knowledge	2
d	Explain conceptual, logical, and physical data models and how they are used at different stages of system and analytics design	knowledge	3
e	Identify core data modeling components (entities, attributes, relationships, keys) and how they define data structure and integrity	knowledge	3
f	Describe normalization and denormalization, including their benefits and tradeoffs for analytics and performance	knowledge	2
g	Explain dimensional modeling concepts (facts and dimensions) and their role in analytical querying and reporting	knowledge	3
h	Differentiate common warehouse schemas (e.g., star and snowflake) and their implications for usability and performance	knowledge	3
i	Review AI-generated schemas for accuracy and completeness against defined business and analytical requirements, and flag potential concerns for escalation	skill	3

Label	KSAC Description	KSA	Bloom's Taxonomy Level
j	Identify common data modeling issues in automated outputs, such as inappropriate keys, incorrect relationships, or misleading abstractions, when compared to expected patterns	ability	3

Data Integration, Pipelines, and Automation

Label	KSAC Description	KSA	Bloom's Taxonomy Level
a	Describe ETL and ELT approaches and when each is most appropriate	knowledge	2
b	Explain batch versus streaming processing and the constraints associated with each approach	knowledge	2
c	Explain API-based data integration and how APIs are used to move data between systems	knowledge	2
d	Describe data synchronization, latency, and reconciliation concepts and their impact on analytics reliability	knowledge	2
e	Identify common pipeline stages and dependencies, from ingestion through transformation and delivery	knowledge	2
f	Explain how automation and AI are used in modern pipelines, including schema inference, transformation, and orchestration	knowledge	3
g	Identify pipeline failure modes such as schema drift, duplication, delayed data, and silent data loss	knowledge	3
h	Interpret pipeline logs, alerts, and monitoring outputs to diagnose data flow issues	skill	3
i	Assess pipeline reliability, data freshness, and downstream impact on analytics or AI use cases	ability	4

Data Quality, Validation, and Monitoring

Label	KSAC Description	KSA	Bloom's Taxonomy Level

a	Explain data quality dimensions (accuracy, completeness, consistency, integrity, timeliness)	knowledge	3
b	Identify common data quality issues such as missing values, outliers, duplication, and inconsistent definitions	knowledge	3
c	Perform data profiling and summarization to understand distributions, trends, and anomalies	skill	3
d	Perform data validation and sanity checks to confirm data meets expected rules and constraints	skill	3
e	Review AI-generated data summaries or transformations for accuracy and alignment with source or expected data	skill	3
f	Explain continuous data quality monitoring concepts, including why proactive monitoring is necessary in automated environments	knowledge	3
g	Identify data leakage risks in analytics and AI pipelines, including unintended use of future or restricted data	knowledge	3
h	Assess data fitness for specific analytical or AI-driven use cases, considering quality, scope, and limitations	ability	4

Data Analytics, Statistics, and Interpretation

Label	KSAC Description	KSA	Bloom's Taxonomy Level
a	Differentiate descriptive, diagnostic, predictive, and prescriptive analytics	knowledge	2
b	Explain probability concepts and common statistical distributions used in analytics	knowledge	2
c	Explain sampling methods and sources of bias, including how bias can be introduced through automated processes	knowledge	2
d	Explain statistical inference, uncertainty, and confidence in analytical results	knowledge	2
e	Interpret analytical outputs in the context of the original business or policy question	skill	3
f	Interpret AI-generated analytical results, including understanding assumptions and limitations	skill	3

Label	KSAC Description	KSA	Bloom's Taxonomy Level
g	Identify misleading, incomplete, or invalid AI-generated analysis and explain why it is problematic	ability	4
h	Articulate the limits of statistical inference and automation, including when human judgment is required	knowledge	2

Data Visualization and Communication

Label	KSAC Description	KSA	Bloom's Taxonomy Level
a	Explain the role of visualization in analytics, discovery, and decision-making	knowledge	3
b	Identify common visualization types and appropriate use cases for each	knowledge	2
c	Evaluate visualization choices for clarity, accuracy, accessibility, and audience appropriateness	ability	4
d	Create basic static visualizations that accurately represent data	skill	3
e	Create interactive or narrative visualizations to communicate insights to varied audiences	skill	3
f	Communicate analytical findings clearly to non-technical stakeholders, including explaining uncertainty and limitations	skill	3
g	Ensure visualizations meet accessibility and compliance requirements	ability	3

Ethics, Privacy, and Responsible AI Use

Label	KSAC Description	KSA	Bloom's Taxonomy Level
a	Explain legal and regulatory constraints on data use	knowledge	2
b	Explain ethical risks in analytics and AI systems, including unintended consequences of automated decisions	knowledge	2

Label	KSAC Description	KSA	Bloom's Taxonomy Level
c	Identify bias in data, models, and AI-generated outputs and understand potential impacts	knowledge	3
d	Explain privacy considerations in data collection, storage, and usage	knowledge	2
e	Identify risks of AI hallucination, misuse, and overreliance on automated outputs	knowledge	3
f	Apply ethical and governance principles to evaluate whether AI-assisted data use is appropriate, responsible, and compliant	ability	4