DATA LAKES FOR EDUCATION

MCG Midtown Consulting group

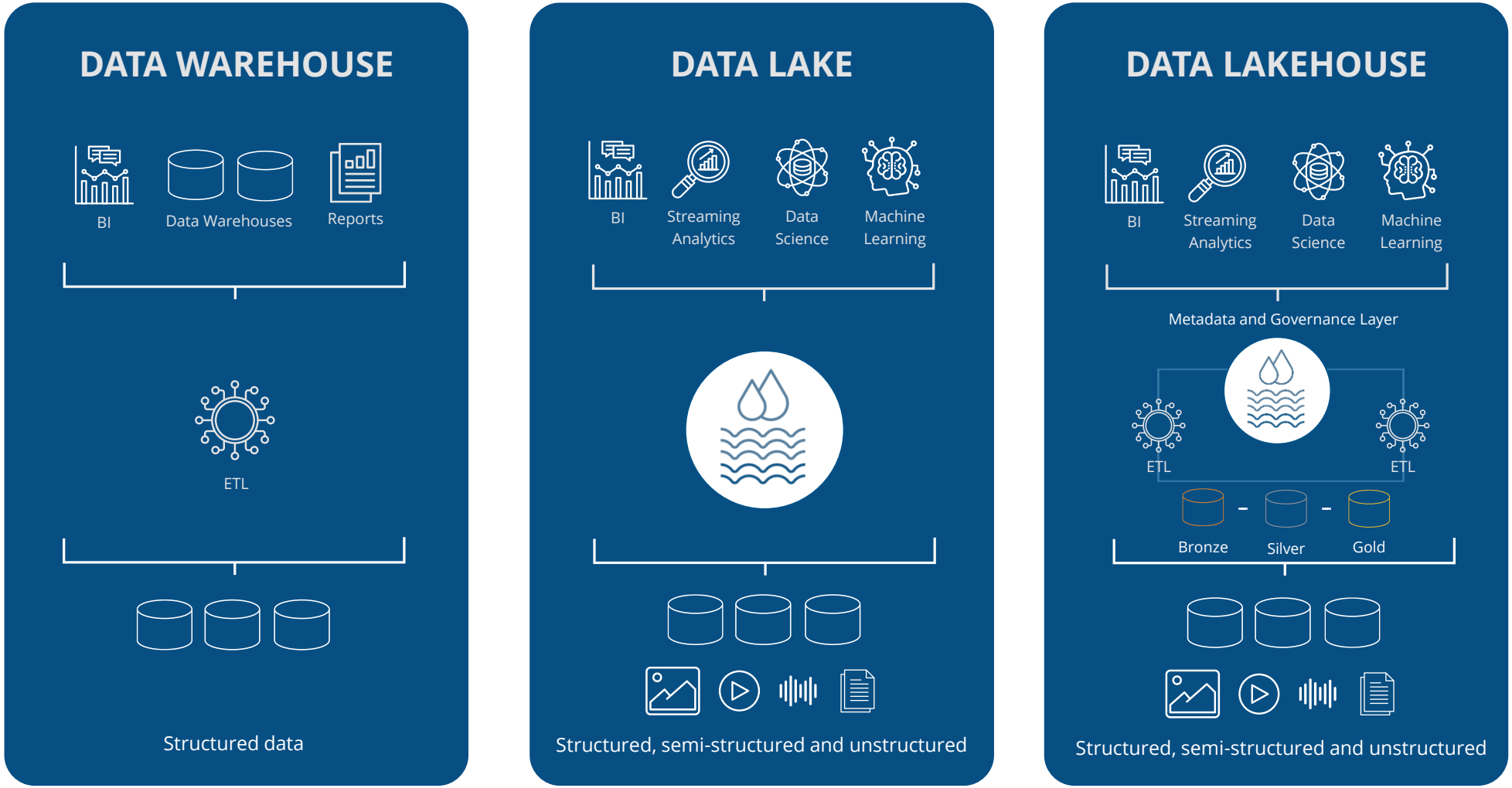
Topics

Microsoft Solutions Partner
Data & AI
Azure

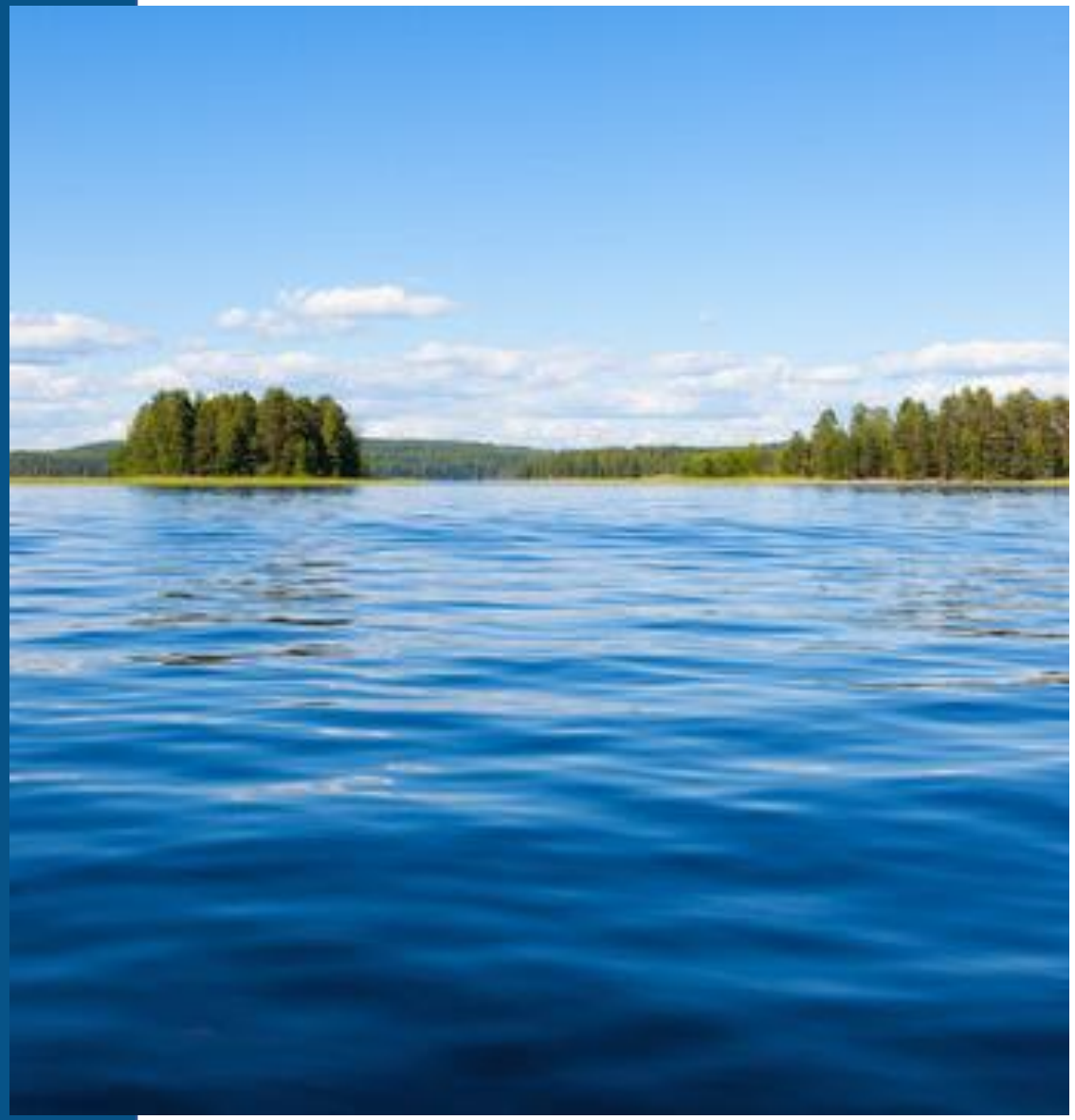
Objective: Review data lake solution benefits and its fitness for Analytics and AI

1. Data Lake
2. Data Lakehouse
3. End to End Data & AI

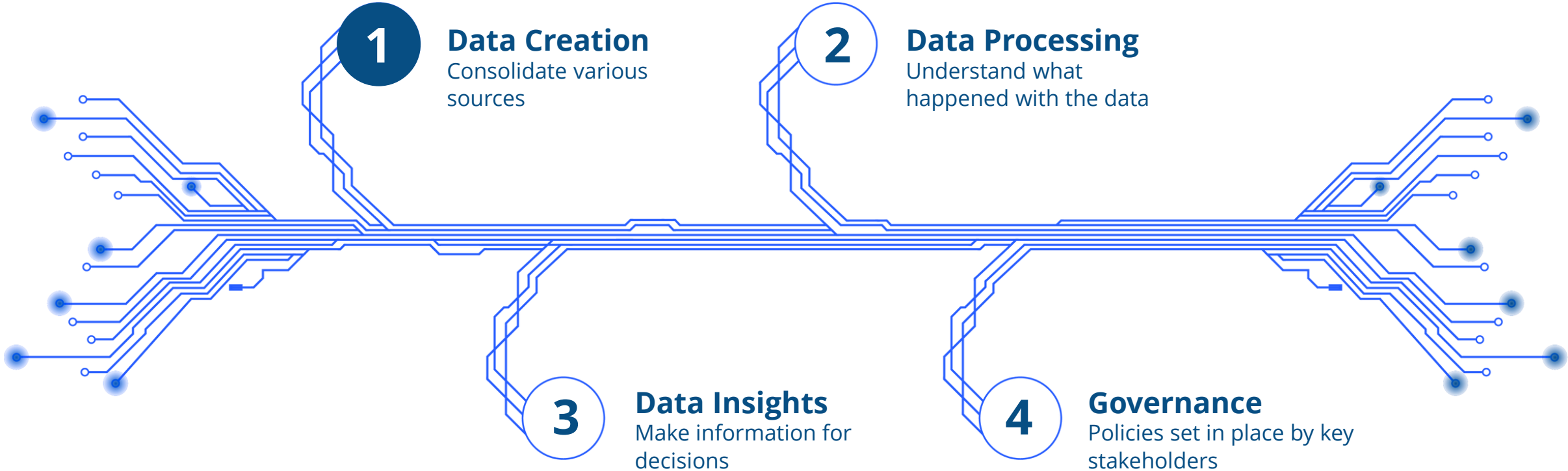
Architectures



Data Lake



Traditional Data Cycle

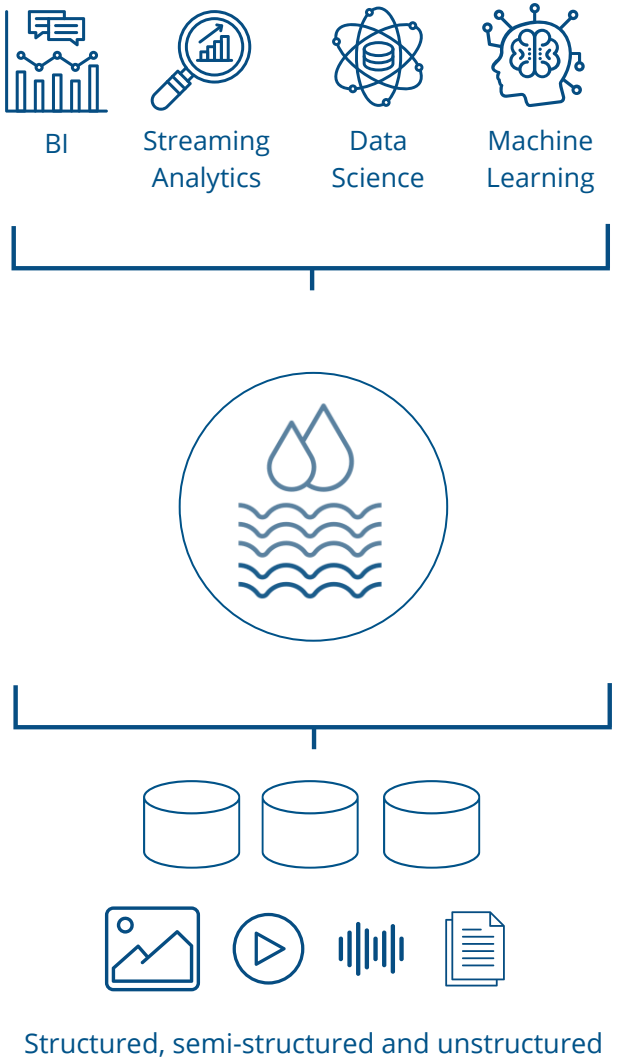


Data Management

1990s	2005	2010
Data Warehouse (EDW)	Data Lake	Lakehouse
• First Response To Growth In Data 1990s • Oracle, SQL Server, Teradata • After A Data Lake (Linear View) • Single Source Of Clean Data • Formatted • Highly Structured	• Second Response To Data Growth • Hadoop • After The Original Source (Linear View) • Dump In Many Data Formats - Dirty • Original Format • Original Structure	• Response To Cost Effective Information • Delta Lake • On Top The Lake, Before The Warehouse • Parquet Data Format • More Formatted, Lightly Structured • Feeds | BI, ML, Streaming Analytics

Data Lake

1. File System Or Repository Data Stored In Its Natural/Raw Format
2. One Place To Land Data | Regardless Of Source, Structure Or Type
3. Supports Structured, Semi-structured & Unstructured Data
4. Stores Transformed Data For Analysis, Visualization, And AI
5. Users: IT and Some Data scientists



Data Lake Benefits

TIME

Faster Time To Consumption

STORAGE

Scalable, Format Friendly, Tiered Performance

COST

Low & Tiered Cost Storage

PARQUET

Columnar, Compressed & Indexed

AUTOMATE

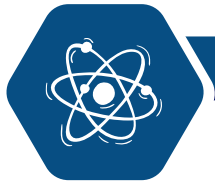
Automate Data Evaluation and Dissemination

Data Lakehouse

Data Lakehouse

+ ACID Transactions
+ Schema Enforcement: Versioning, Time Travel, Schema Evolution
+ Governance: Auditing, Retention, And Lineage
+ Medallion Structure For Data Evolution | Reliable
+ Expanded Usership| Data Analysts, Data Scientists, Machine Learning Engineers

Data Lakehouse Benefits



TIME

Faster Time To Consumption For More Users

STORAGE

Elastic & Tiered Performance

COST

Retains Data Lake Cost Structure

DELTA LAKE

Columnar, Compressed, Indexed, Organized

AUTOMATE

Automate Zone Transformations for Value

Data & AI Solution

End to End Data & AI Architecture