Choosing the Model Type Based on Data Access and Storage Criteria

Patrick Pichler
DATA & ANALYTICS SPECIALIST

Overview

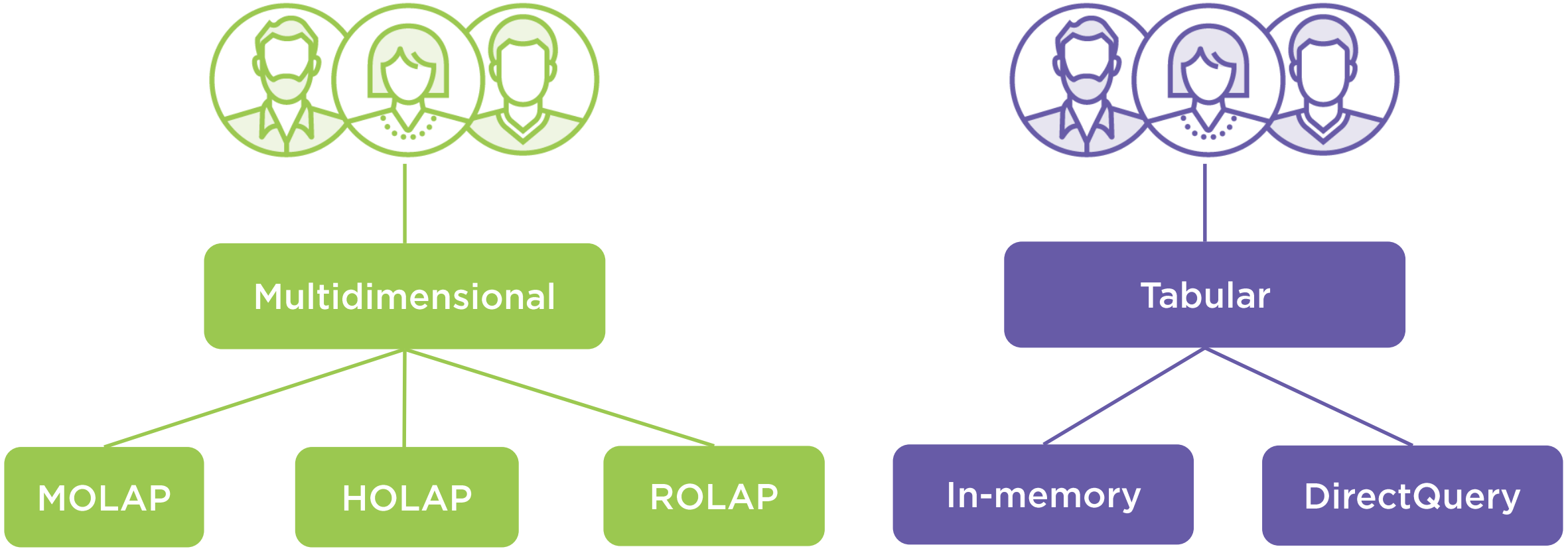
- Analyzing supported data sources
- Contrasting storage modes
- Comprehension of query engines
- Taking into account performance and scalability

Supported Data Sources

Supported Data Sources

	Multidimensional	Tabular
Databases	SQL S., Access, Oracle, Teradata, IBM	SQL S., Access, Oracle, Teradata, IBM, SAP
Data Warehouses/OLAP	Azure DW, PDW/APS	Azure DW, PDW/APS, SSAS, SAP BW
Azure storages	-	Blob, Data Lake, Cosmo DB, HDInsight
Files	-	Excel, Folder, Text/CSV, JSON, XML Table
Online services	-	D365, Salesforce, Exchange, SharePoint
Others	-	AD, Exchange, OData Feed,...

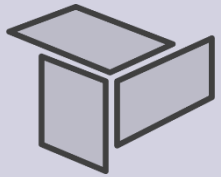
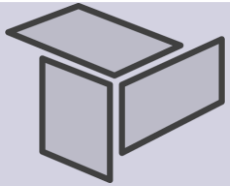
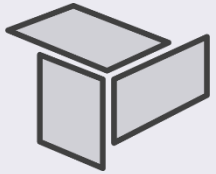
Storage Modes/Data Access



Multidimensional

	MOLAP	HOLAP	ROLAP
Cube structure			
Aggregate data			
Detail data			
Benefits	+ Fast response + Complex calc.		+ Huge data + Data freshness
Drawbacks	- Processing time - Cube size (disk)		- Slow response - Limited calc.

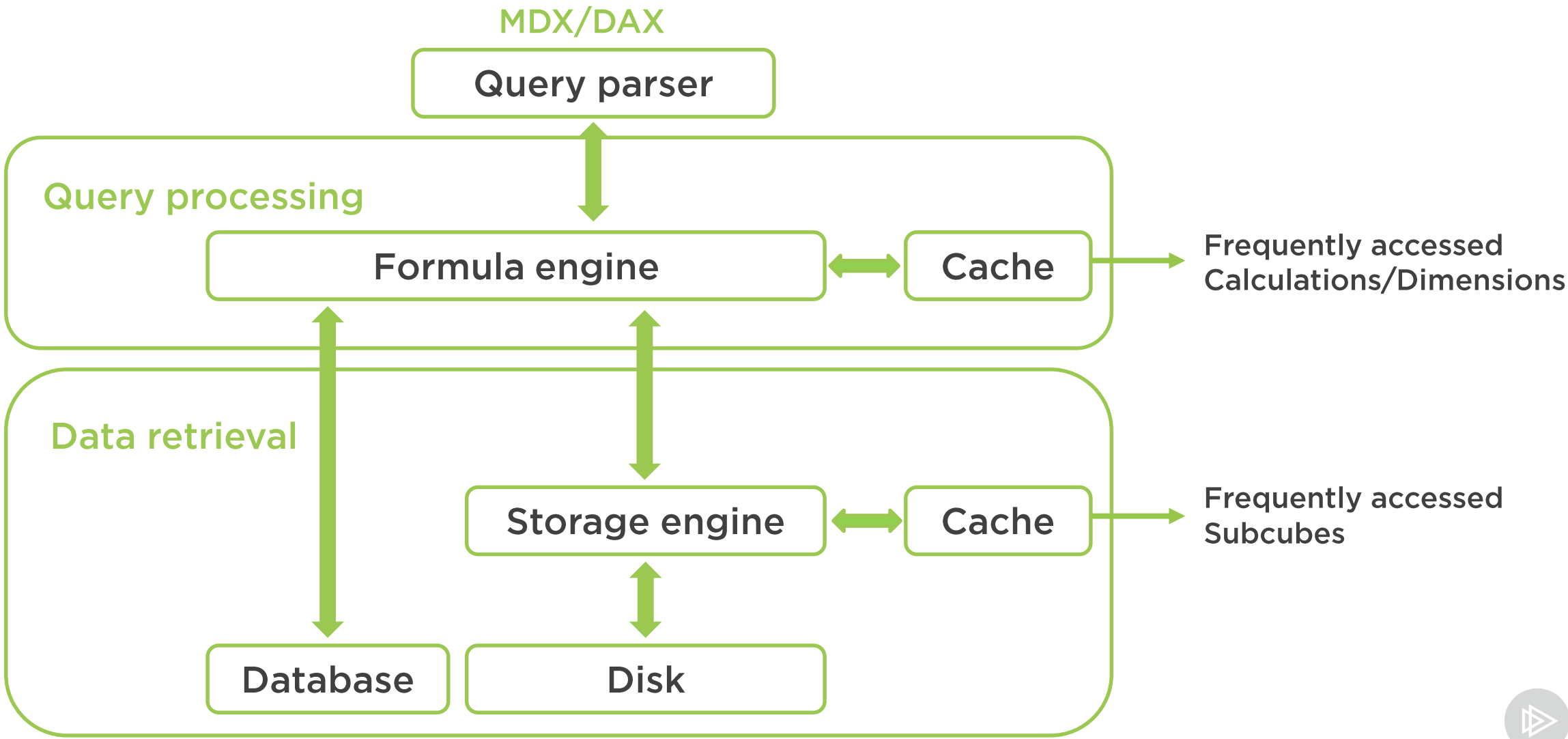
Tabular

	In-memory	DirectQuery
Model structure		
Detail data		

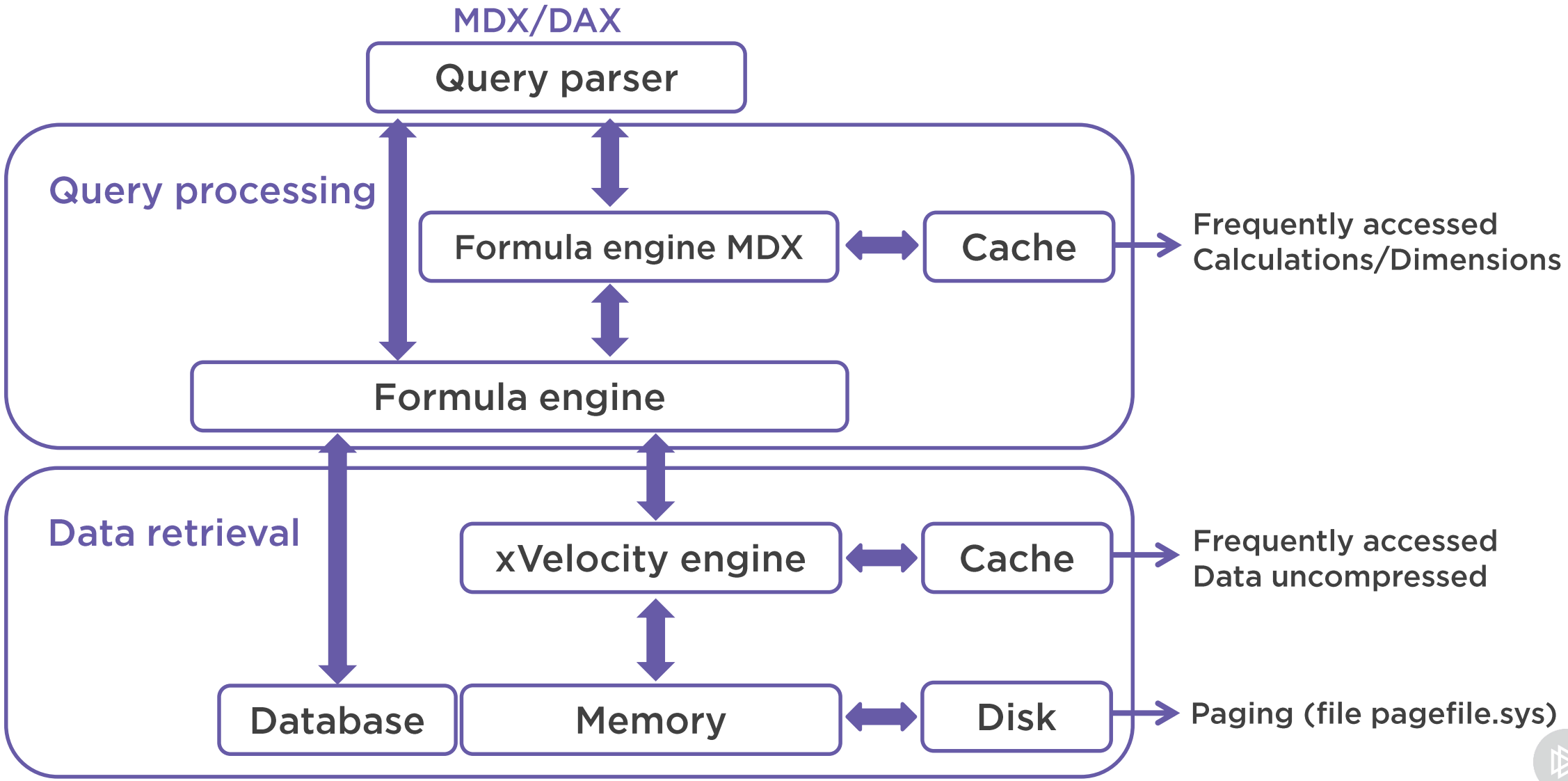
Benefits	+ Fast response + Multiple data sources	+ Huge data + Data freshness
Drawbacks	- Model size (memory) - Processing time	- Single data source - Limited functions/logic

Query Engines

Multidimensional Query Engine



Tabular Query Engine



Performance and Scalability

Performance and Scalability

	Multidimensional (MOLAP)	Tabular (In-memory)
Data compression ratio	Approx. 3:1	Approx. 10:1
Data storage	Row storage	Columnar storage
Data aggregation	Pre-aggregated data from disk	Raw data from memory (query time)
Data volume	Limited by disk (multi-terabyte)	Limited by memory
Concurrent queries	Disk I/O intensive	Memory intensive
Query response	Medium to fast	Very fast

Scalability

Performance

Demo

- Creating a new project from scratch
- Importing data into the model
- Creating a simple semantic data model
- Writing queries and analyzing query behavior

Summary

Discussing various potential data sources of both model types

Analyzing the different ways of storing and accessing data

Taking apart the query engine

Pointing out major characteristics concerning performance and scalability