
MODULE 6

EMERGING TRENDS



06

CHALLENGES IN TRADITIONAL WAREHOUSES

Traditional warehouses face challenges such as limited space, inefficient layout, manual inventory management, high labor costs, inefficient order picking, lack of real-time visibility, inventory loss, and limited scalability. These challenges can be overcome by adopting modern technologies and practices like warehouse management systems, automation, robotics, and data analytics.



01

SCALABILITY

Growth in data volume, Number of users and applications

[LEARN MORE](#)



02

CONCURRENCY

Concurrency issue occurs when number of users increase

[LEARN MORE](#)



03

PERFORMANCE

Slow running queries

[LEARN MORE](#)



04

RESILIENCE

Data backup/retention and node failure

[LEARN MORE](#)



05

SUPPORT OF VARIETY OF DATA

Lack of support for semi-structured data(json,xml) and un-structured data, streaming data

[LEARN MORE](#)

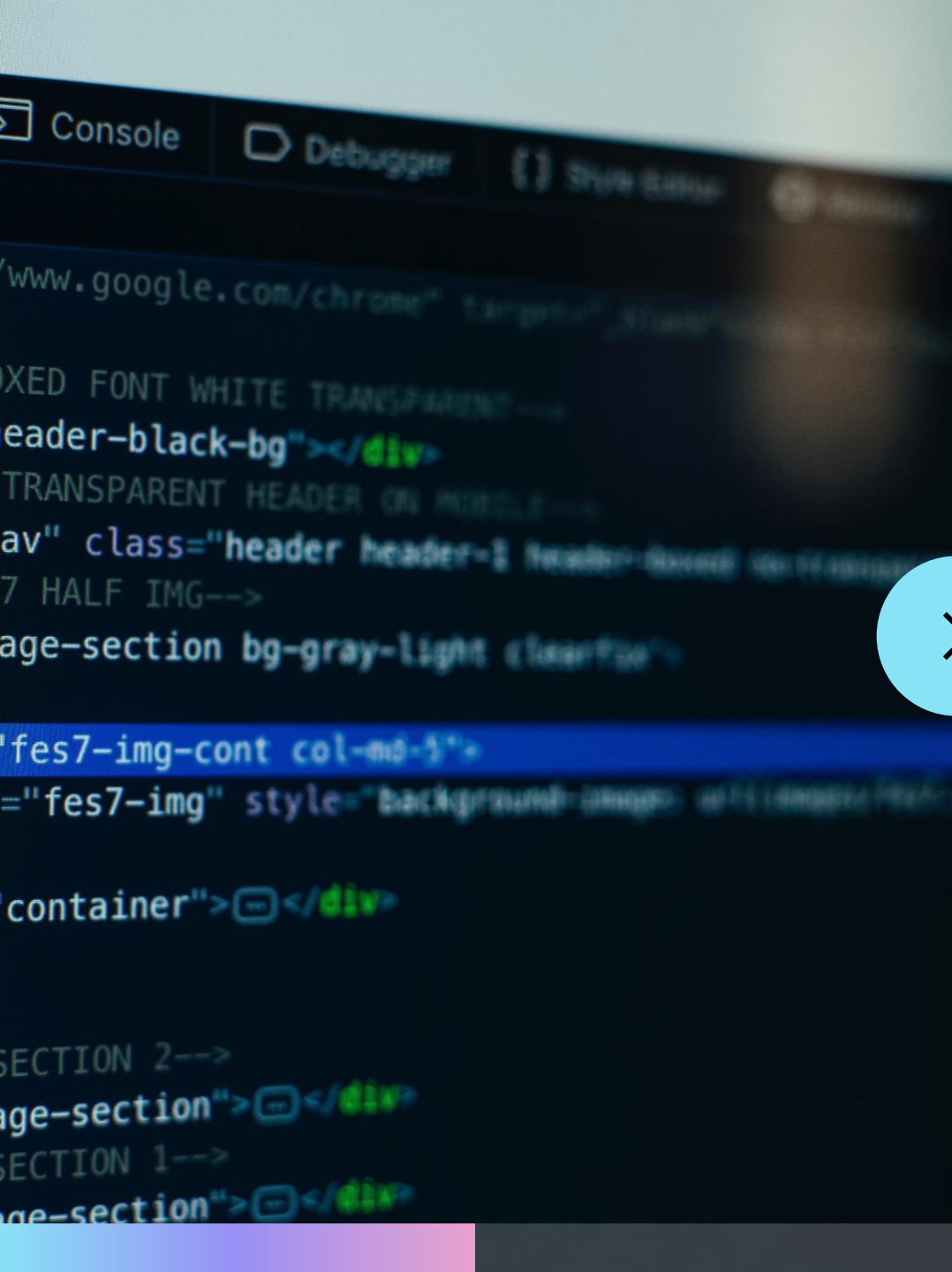


06

HIGH MAINTENANCE

Overhead for tuning, indexing, sorting

[LEARN MORE](#)



MODERN DATA WAREHOUSE SOLUTIONS

Scalability – Scale up, down or off quickly without delay

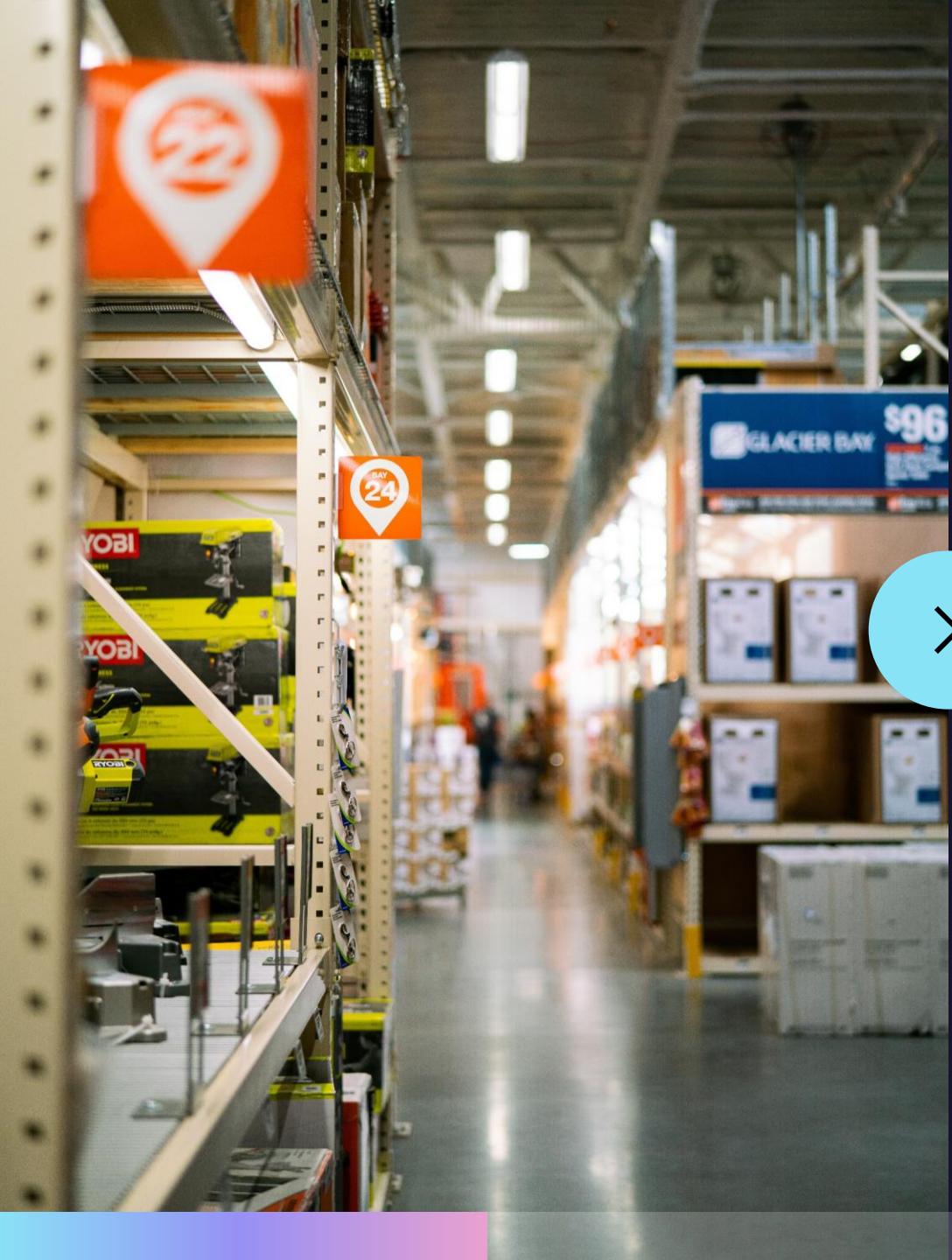
Concurrency- Many users can operate simultaneously

High performance – Processing bottlenecks and delays, slow running queries (Serverless)

Resilience – Data backup/ retention and node failure protection

Support for semi-structured data(json,xml) and un-structured data, streaming data

[READ MORE](#)



MODERN DATA WAREHOUSE SOLUTIONS

Maintenance – Pay as you go, less maintainability, no wastage of resources

Handling workload failure – High elasticity, cloud data warehouse adopt to workload peaks

Cost – No upfront cost, pay as you go, no infrastructure creation cost

[READ MORE](#)



A laptop is shown from a front-three-quarter view, displaying a web page titled "COMPARISON OF CLOUD DATA WAREHOUSES". The page contains a table with five columns: Service, Configuration (100 GB), Cost, Configuration (1 TB), and Cost. The table lists five services: Amazon Redshift, Snowflake, Microsoft Azure SQL Data Warehouse, Google BigQuery, and IBM Db2 Warehouse on Cloud. The costs are listed in US dollars.

	Configuration (100 GB)	Cost	Configuration (1 TB)	Cost
Amazon Redshift	8x dc2.large	\$2.0	4x dc2.8xlarge	\$19.20
Snowflake	X-Small	\$2.0	Large	\$16.0
Microsoft Azure SQL Data Warehouse	DW200	\$2.42	DWI500c	\$18.12
Google BigQuery	On-Demand		On-Demand	
IBM Db2 Warehouse on Cloud				

POPULAR CLOUD DATA WAREHOUSES

- Amazon Redshift (AWS): Fully managed, scalable, and high-performance data warehousing service.
- Google BigQuery: Serverless, highly scalable, and cost-effective data warehouse for large dataset analysis.
- Microsoft Azure Synapse Analytics: Cloud-based analytics service combining data warehousing and big data processing capabilities.
- Snowflake: Cloud-based data warehousing platform with scalability and separation of compute and storage resources.
- IBM Db2 Warehouse on Cloud: Fully managed cloud service offering in-memory processing and advanced analytics.

DATA LAKES

01

DEFINITION

A Data lake uses a flat architecture to store a huge amount of raw data in it's native format until it is needed

02

SIZE

There is no fixed limit on account size or file

03

ELEMENTS

The different data elements in data lakes are assigned unique identifiers and tagged with extended metadata tags

04

USED BY

Data lakes are useful for data scientist because they allow experimentation on massive data sets

DATA LAKES

01

USE CASE

The users of data lakes are usually people who want to do a thorough analysis of data

02

ACCESS

A data lake, because it stores all kinds of data in its raw form, is easily available for access to any user.

03

DATA SCHEMA

Users are able to explain data in novel ways

04

CHEAP

Data lakes is a cheaper way to store/ manage data

DATA WAREHOUSE

01

DEFINITION

A hierarchical data warehouse stores data in files or folders with a defined schema

02

SCHEMA

The information in a data warehouse is stored by subject in order to assist management make quick decisions

03

COMPONENTS

The measures and dimensions are conformed to curable components which are consistent, governed and easier for an ever-scalable audience to consume

04

USERS

80% of users of data warehouses are business users who need refined and systematic data

DATA WAREHOUSE

01

SET UP

A data warehouse takes a fairly long period of time to set up.

02

DEVELOPMENT

During its development, a lot of time is dedicated to analysing the sources of data and how it can be tuned to meet the needs of a particular business

03

COST

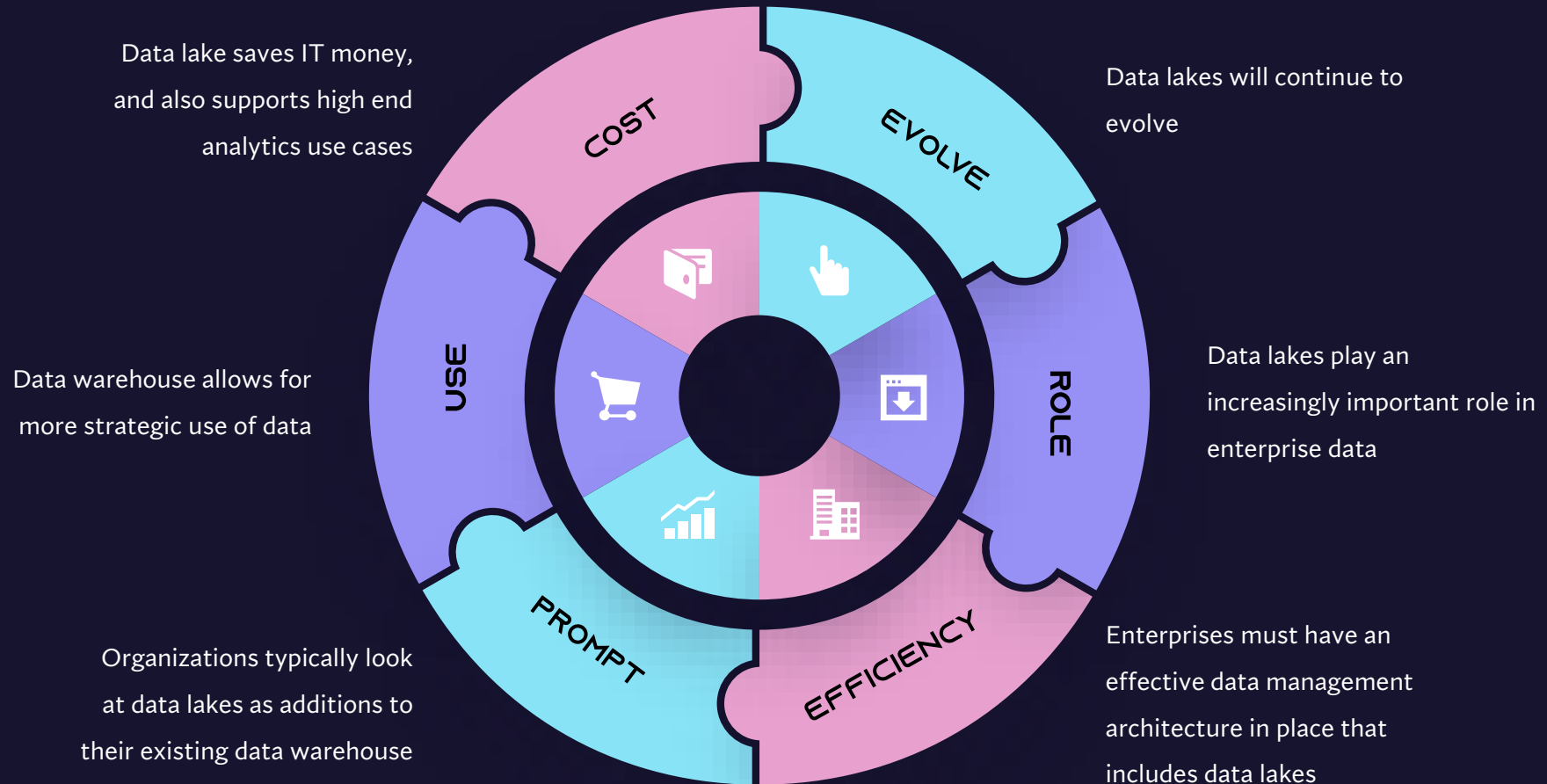
Data warehouse is a costlier way to store.
Manage data

04

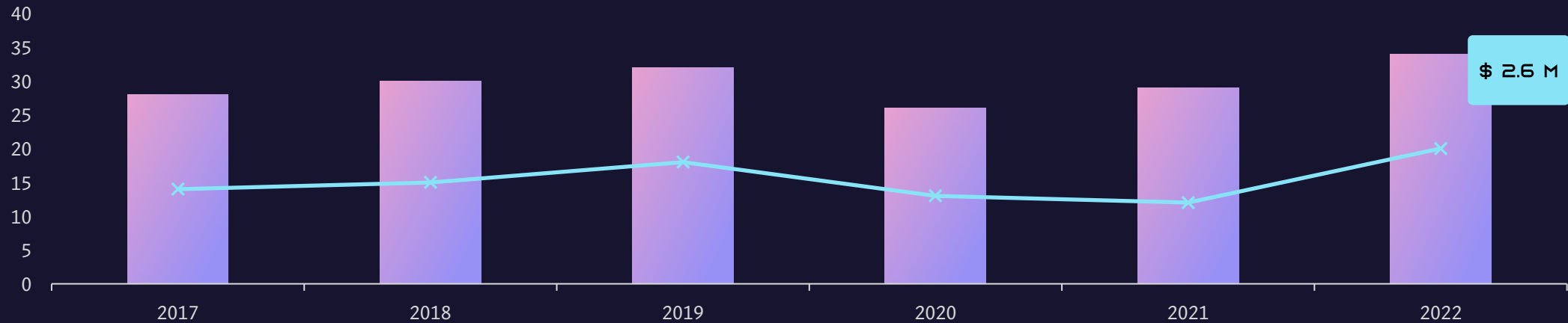
USERS

Database administrators and Database Engineers and sometimes Maintenance crew

THE MAIN USE CASES



MANAGED DATA WAREHOUSE SERVICES



+56%

Managed data warehouse services offer the benefits of scalability, agility, cost-effectiveness, and integration with the broader cloud ecosystem, empowering organizations to leverage their data assets effectively and drive data-driven decision-making.

DEFINITION

Managed data warehouse services refer to cloud-based solutions that provide organizations with fully managed and scalable data warehousing capabilities. These services handle tasks such as infrastructure management, database administration, data storage, and maintenance, allowing businesses to focus on data analysis and decision-making rather than the underlying technical aspects.

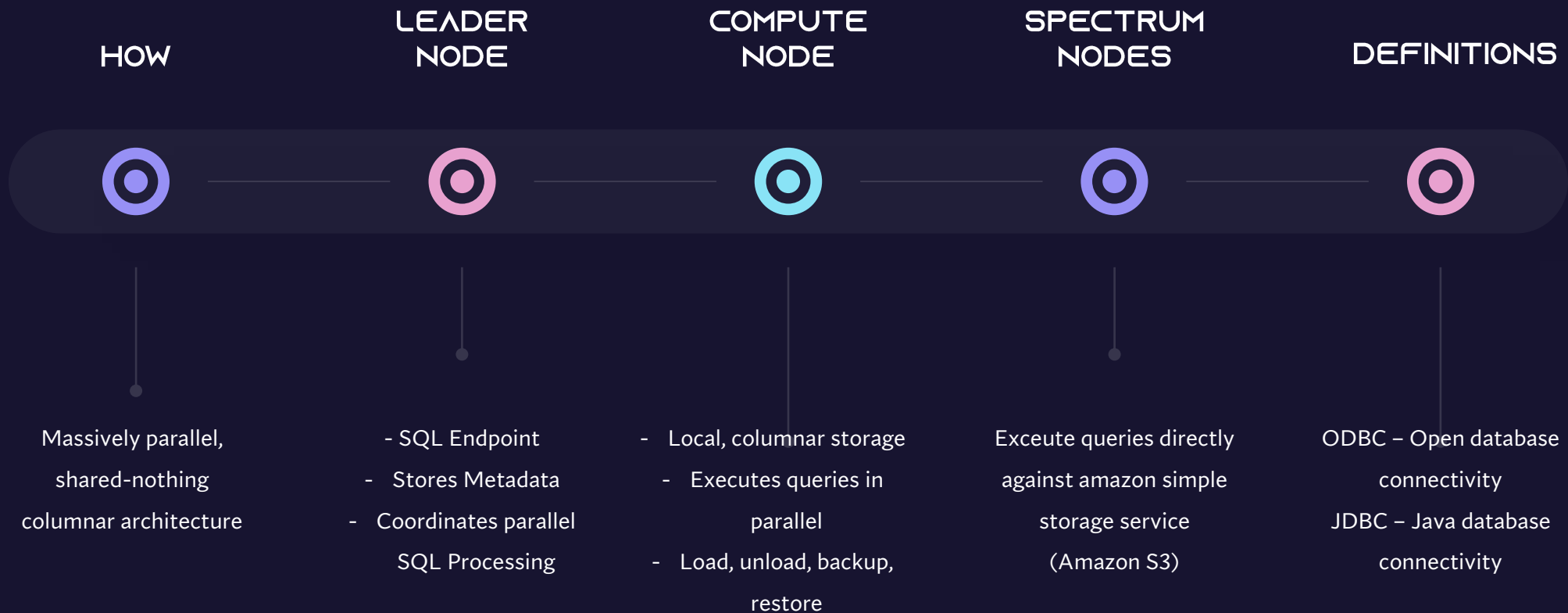


AMAZON REDSHIFT ARCHITECTURE

Types of nodes:

1. Leader node
2. Compute node
3. Amazon redshift spectrum nodes

BREAKDOWN OF THE ARCHITECTURE



LEADER NODE

01

TERMS

Parser & Rewriter

Planner & Optimizer

02

ACTS

Code generator

Compiler

03

WHY

Task Scheduler

Workload management

04

HOW?

Admission + Scheduling

PostgreSQL Catalog
tables

COMPUTER NODE

01

TERMS

Query execution
processes

02

ACTS

Backup & Restore
processes

03

WHY

Replication processes

04

HOW?

Local storage:
Disk+Slices+Tables
Columns & Blocks

**01**

CLUSTERS

A cluster is composed of one or more compute nodes. If a cluster is provisioned with two or more compute nodes, an additional leader node coordinates the compute nodes and handles external communication. Your client application interacts directly only with the leader node. The compute nodes are transparent to external applications.

**02**

LEADER NODE

The leader node plays a crucial role in managing communication between client programs and compute nodes in a database system. It is responsible for handling all communication tasks and parsing commands from clients. Additionally, it develops execution plans for database operations, which involve a series of steps required to obtain results for complex queries.

[LEARN MORE](#)**03**

COMPUTE NODES

The compute nodes, each equipped with its own dedicated CPU and memory, execute the compiled code and send back intermediate results to the leader node.

[LEARN MORE](#)

A DOPE MAP TO DISTRACT YOU FROM THE
FACT THAT 7.45L IS GONE LIKE MY EX



THANK YOU

That joke bit- “Redefining leading talents”

