



Db2 Cloud Update

Les King

lking@ca.ibm.com

Christopher Hutchings


hutchcj@amazon.co.uk

September 2024

Data Server Day – Stockholm, Sweden

Agenda

- Db2 RDS
- Db2 Warehouse on Cloud (SaaS) Gen 3
- Db2 pureScale on AWS Marketplace
- Db2U Update

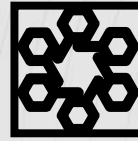


Db2 RDS

IBM Db2 + Amazon RDS

Bringing together one of the **LARGEST WEB SERVICE** in the world with a database that runs the most **COMPLEX TRANSACTIONAL WORKLOADS** in the World

Released at re:Invent 2023



- Automates time-consuming tasks like provisioning, patching and backups.



- Highly available across availability zones with managed snapshot backups and point in time recovery.

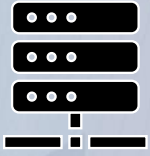


- Built in IBM and AWS technical expertise integrated and accessible via standard RDS interfaces



- Built-in compliance and tested for security

Self managing databases is time consuming, complex, and expensive



Hardware & software installation, configuration, patching, backups



Performance and high availability issues



Capacity planning and scaling



Security and compliance

Introducing Amazon RDS for Db2

RUN FULLY-MANAGED IBM Db2 DATABASES ON AWS



Increase efficiency

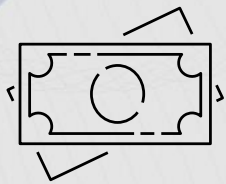
Automates undifferentiated Db2 tasks, such as provisioning, backups, patching, and monitoring



Focus on innovation

Easily migrate existing IBM Db2 databases

Launch Db2 Version 11.5 databases in minutes and enable high availability with RDS Multi-AZ deployment



Reduce costs

Bring your own IBM software licenses, supporting Standard and Advanced Editions & sub capacity licensing policy

Supports transactional, mixed and analytics workloads, including Oracle compatibility

Overview of Amazon RDS for Db2

Easy to administer



- Create database with few clicks in few mins
- No infrastructure provisioning, software installation, or patching
- Built-in monitoring

Performant and scalable



- Power your database with push-button compute scalability
- Auto scale your storage

Available and durable



- Achieve high availability with Amazon RDS Multi-AZ deployments
- Automated backup, snapshots, and failover

Secure and compliant



- Protect data with encryption at rest and in transit
- Achieve compliance with key industry compliance programs

Amazon RDS for Db2 – Key Attributes

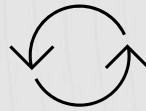
Increase
Efficiency

Focus on innovation

Reduce Costs



Grow Storage up to **64 TiB** and scale storage performance independent of storage size up to **256K IOPS** and **4000 MiB/s**



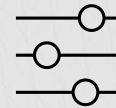
Automatic patching with customizable maintenance windows and policies



Db2 11.5.9 - Standard and Advanced Edition (BYOL & private offer AWS Marketplace)



Scalable up to **128 vCPU** and **4 TB of memory**.



Customizable configuration parameters and ability to manage shared configurations.



Storage replication with automatic health and repair capability.

What aspects does AWS RDS Manage for you?



Provisioning – infrastructure, OS, database, configuration



Patching – provides packages for updates, orchestration and auto apply



Backups – automated snapshots, configurable



Recovery – restore, PITR to a new RDS system



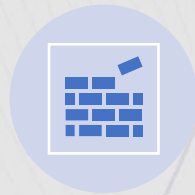
Failure detection – health detection, failover to another AZ



Security – Encryption at REST, in transit, security protocols, elevated controls (Db2 will be the most secure offering on RDS)



Architecture – system design and implementation



High Availability/Replication/Data Durability – storage replication

Responsibility Matrix

	End User	Amazon	IBM
Schema Design	X		
Query Optimization	X		
Workload Management & Sizing	X		
Authentication and Authorization	X	X	
Backup and Recovery		X	
Monitoring Capabilities	X	X	X
Patching and upgrades		X	X
Automated Patching		X	
Support		X	X
Scaling		X	
Security Monitoring		X	
Durability Management		X	
Db2 Feature Development			X
RDS Feature Development		X	
Industry Compliance		X	X

Amazon RDS for Db2 Restrictions

- Customers are only given DBADM authority
 - Actions like Tablespace and Bufferpool management which are still needed are done via RDSADMIN stored procedures
- No support for read replicas currently – coming soon
- No support for multiple databases currently – coming soon
- Only SQL and Java Stored Procedures are available
- No support for unfenced procedures
- No support for heterogenous federation at present
- Custom Security Plugins are not allowed
- No support for Db2 Native Encryption
- No Support for HADR from Prem Environments
- Text Search is currently not available
- No support for reserved instances – coming soon

RDS for Db2 2024 Releases since launch

03/20/2024

AWS License Manager

[AWS License Manager now allows you to track IBM Db2 licenses on Amazon Relational Database Service \(RDS\)](#)

03/11/2024

Support for M6i and R6i in additional AWS Regions

[Amazon RDS for Db2 expands support for M6i and R6i in additional AWS Regions](#)

03/11/2024

Support for X2iedn instances in additional AWS Regions

[Amazon RDS for Db2 expands support for X2iedn instances in additional regions](#)

03/06/2024

Support for io2 Block Express for sub-millisecond latency and 99.99% durability

[Amazon RDS now supports io2 Block Express for consistent sub-millisecond latency and 99.999% durability](#)

02/15/2024

Support for Audit Logging

[Amazon RDS for Db2 now supports audit logging](#)

01/29/2024

Support for EBCDIC Collation Sequence

[Amazon RDS for Db2 now supports EBCDIC collation sequence](#)

01/19/2024

Support for Cross-Region Automated Backups

[Amazon RDS for Db2 now supports Cross-Region Automated Backups](#)

01/19/2024

Amazon RDS for Db2 now supports local time zones

[Local time zone support for Amazon RDS for Db2](#)


Amazon RDS for Db2 resources

- [AWS News Blog](#)
- [Service page](#)
- [Technical Documentation for RDS Db2](#)

Contact for questions, deep dive session

Karthik Gopalakrishnan

Senior Product Manager
carthik@amazon.com



Db2 Warehouse on Cloud (SaaS) – Gen 3

Db2 Warehouse (SaaS) – Gen 3 – Based on Db2 11.5.9



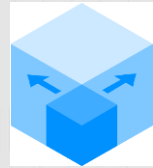
Fully managed / SaaS

Focus on the analytics, we'll take care of the rest



Blazing-fast

Columnar-organized, memory-optimized data warehouse



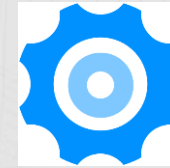
Scalable & elastic

Independently scale and manage compute & storage



Continuously available

Managed compute, highly available storage, cross-cloud replication



Reliable

Double protection with disaster recovery & self-service backup/restore



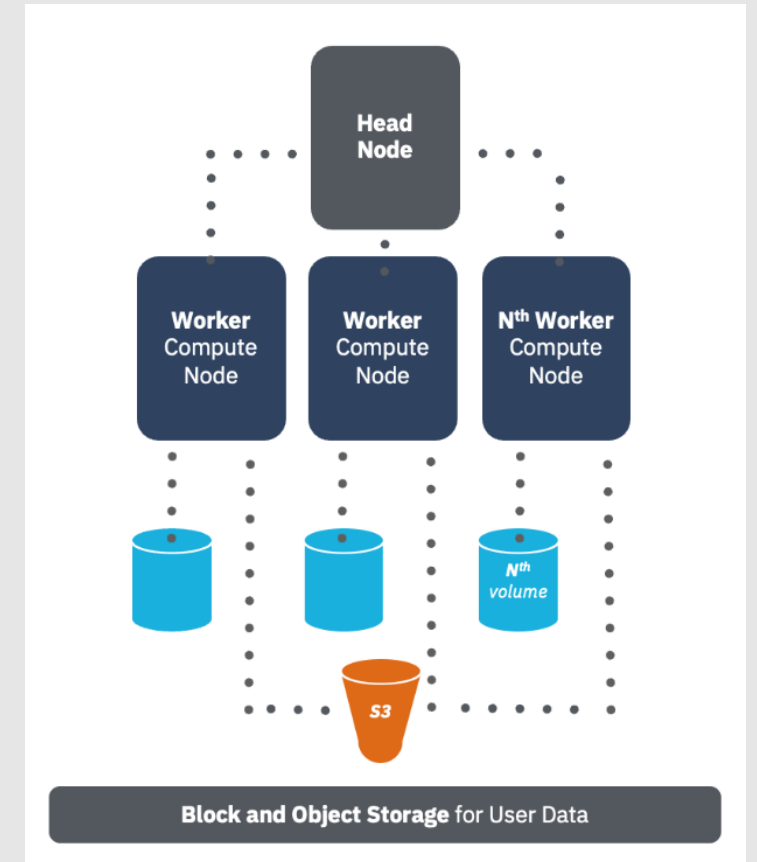
Secure

Encrypted at the storage level by the vendor and customer can protect data using their Key Protect keys.

Deploy as fully-managed DBaaS on **IBM Cloud** and **Amazon Web Services**

Db2 Warehouse (SaaS) – Gen 3 – Cloud Object Store

- Db2WoC now supports Amazon S3 object storage for database table storage, where customer data resides within the database
- Customer saves cost by using object storage instead of block storage
 - Allows customers to store data across a mix of block storage and object storage, based on business or technical requirements
 - Db2WoC uses different mechanisms in order to facilitate reads and writes to object storage
- No applications and workload changes necessary in order to use this feature
 - Db2WoC handles all the necessary interfacing to object storage, thus existing applications and warehouse workloads do not have to be changed in order to make use of this
 - Specific Db2 tablespace available backed by S3 for customer use
 - Insert, Update, Delete data as needed into and out of tables within object storage
 - Move and copy data to and from column-organized tables residing in block storage and object storage
 - Query data seamlessly no matter where it resides (in block or object storage), in isolation or in combination with each other



Db2 Warehouse (SaaS) – Gen 3 – Cloud Object Store - Motivation

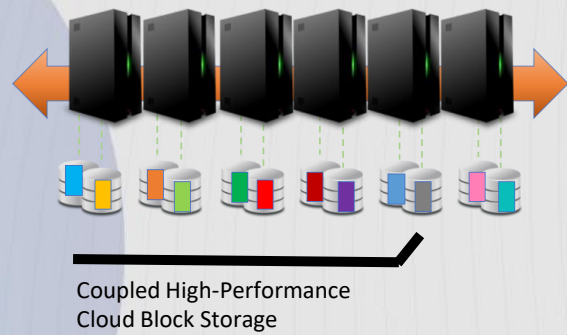
- ✓ Low Cost: 34X cheaper than network-attached cloud storage
- ✓ Near unlimited scalability
- ✓ Extreme durability + reliability: 11 9s in cloud environments
- ✓ High throughput: only limited by network bandwidth of compute
- ⊠ High latency: 10X or more than network-attached cloud storage



Cloud Object Storage (COS)

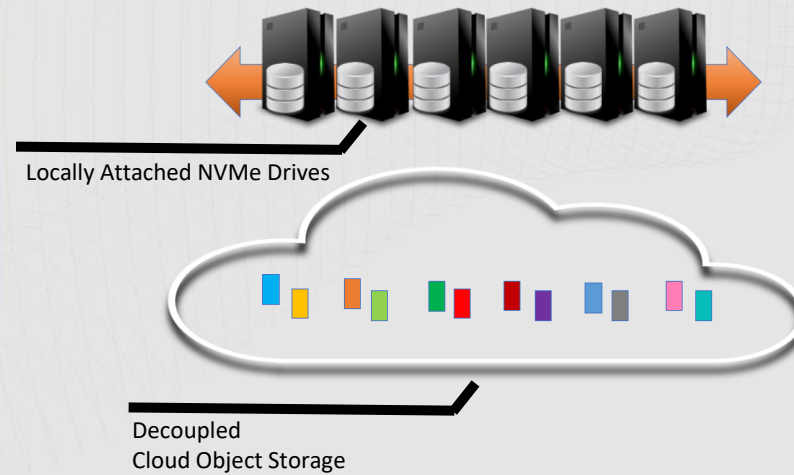
Db2 Warehouse (SaaS) – Gen 3 – Storage Architecture

Db2 Warehouse on Cloud Gen 2



@ 10-30ms latency each (6 IOPS/GB)

Db2 Warehouse On Cloud Gen 3

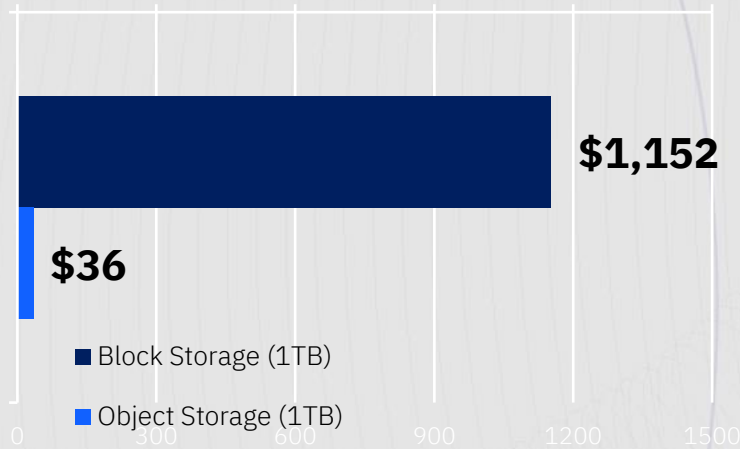


@ 100-300ms latency per operation

Db2 Warehouse (SaaS) – Gen 3 - Cloud Object Storage - Value

Storage Cost:

34x
Less expensive to host Db2 data on object vs block storage¹

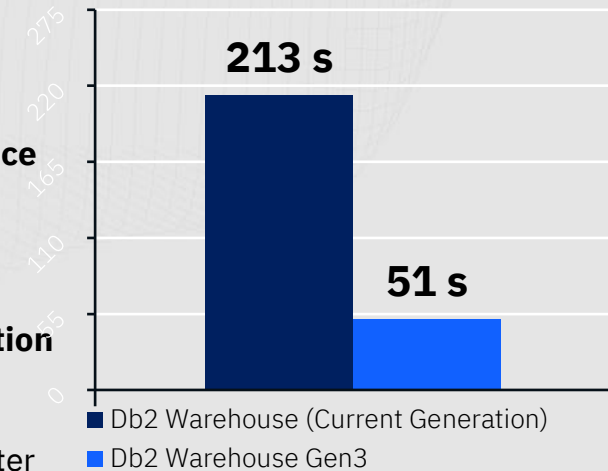


- Cost of Db2 Warehouse block vs. object storage.

Performance:

4x
Faster query performance

When Gen3 is compared against the prior generation



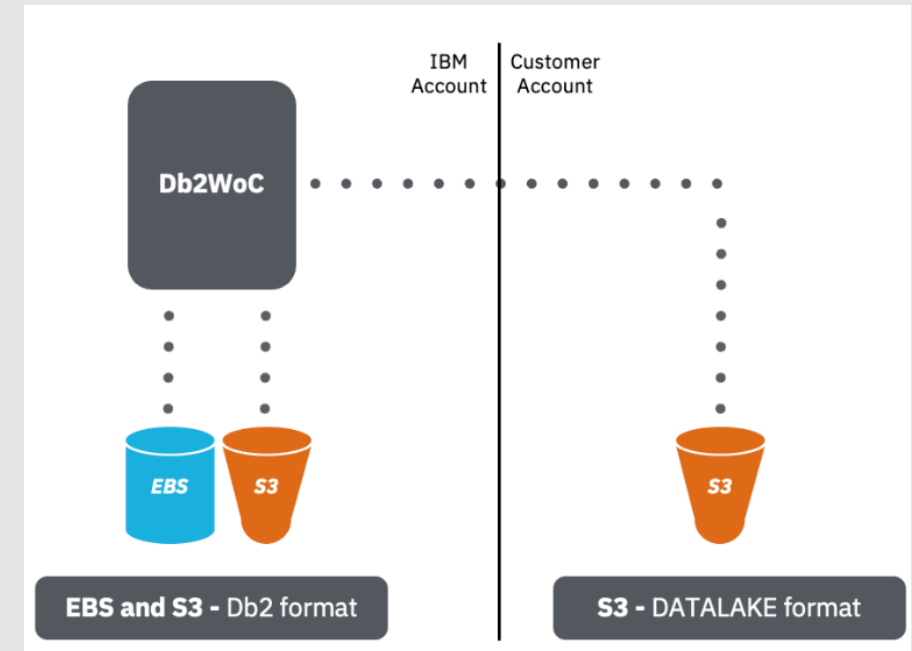
Note: Lower number is better

- IBM Big Data Insight (BDI) Benchmark simulates real-world deep analytics, reporting, and dashboard queries
- 10TB Db2 data warehouse residing either on block storage (current generation) or object storage (Gen3)
- 16 concurrent users running a variety of ML, reporting, and dashboard queries
- Cold cache start for both in-memory buffer pools or the NVMe cache

¹Block vs Object Storage comparison depicts difference between published prices for Amazon EBS 1TB of io1 at 6 IOPS/GB (and additional tiers to support Db2 data) vs Amazon S3. This metric is not an indicator of future storage pricing for Db2 Warehouse Gen 3.

Db2 Warehouse (SaaS) – Gen 3 – LAKEHOUSE Table

- Db2WoC now also supports Open Data Formats as DATALAKE tables, allowing for seamless access to other data within the enterprise for integrated workloads.
- Leverage existing compute resources dedicated to the warehouse
 - Facilitate data use to and from the Db2 Warehouse to quickly access a variety of enterprise data
 - Leverage the high performance Db2 engine for queries against enterprise data
- Db2WoC provides interfaces for customers to leverage their enterprise data residing in object storage as DATALAKE tables
 - Supports both regular and Iceberg DATALAKE table types, based on existing data formats or for business/technical requirements such as ACID compliance
 - Browse, explore, and query enterprise data in both Db2 and DATALAKE formats, using either the web-based UI, or through SQL
 - Access data in place within DATALAKE tables, joining as necessary with Db2 based data for queries
 - Access data within DATALAKE tables and import into Db2 formatted tables
 - Create new DATALAKE tables in S3 and export from Db2 formatted tables



Db2 Warehouse (SaaS) – Gen 3 – Based on Db2U

- DB2u containers are the base.
- Every customer deployment maps to a **Formation**. A formation is realized by the Operator model and deploys a complete cluster
 - Gets storage provisioned
 - Gets N/W components created
 - Creates container/pod configurations based on formation specs
 - Db2u process sets up a db2 database with the storages and sets up tablespaces with them
 - Customer gets two endpoints, one to the Console and the other to the database.
- **Recipes**
 - Used for all Operations like scale, backup, addition of users etc.

Db2 Warehouse (SaaS) – Gen 3 – Highly Scalable

- Scale using the Console or APIs
 - The platform provides mechanics to initiate a scale (we create a scale recipe)
 - In the scale recipe, we provide the new core and storage values.
- The operator framework will
 - Gracefully shutdown the engine and scale down the formation
 - Bring it up in the new spec (appropriately setting the correct MLN count per pod)
 - As the new spec is brought up, new nodes are automatically added by the Cluster Auto Scaler
 - If it is a shrink operation, unused nodes are automatically returned by the Cluster Auto Scaler
- Compute Expansion / Shrink incurs a Downtime
- Storage scaling works the exact same way (recipe mechanism)
 - Storage scale up is a totally online operation
 - Block Storage shrink (support date TBD) will be offline in nature

Db2 Warehouse (SaaS) – Gen 3 – Improved HA

- **Multi prong, provided by**
 - Wolverine (our HA component) - Core Db2 container
 - Kubernetes - non Db2 containers
- **Wolverine responsible for and provides**
 - System recovery (when the pods fail)
 - Overall System health and status recording in ETCD
 - Device Manager framework
 - HA loop (for monitoring and handling HA for processes within core Db2 Container)
 - API endpoints
- **Kubernetes**
 - Pod level HA based on liveness probes
- **When a Db2 node goes down,**
 - Kubernetes will automatically schedule the pod to run on a free node from the pool
 - Wolverine will wait for the pod to come back up and take the engine through a recovery process
 - Engine is back running in full capacity within a short time.
- **When a Db2 pod goes down**
 - Kubernetes will restart the pod
 - Wolverine will wait for the pod to come back and take the engine through a recovery process.

Db2 Warehouse (SaaS) – Gen 3 – More Granular Backup/Restore

- Backups - a combination of Snapshot Backups and S3 backup
 - Scheduled every 24 hour period or on demand
 - Scheduled or on-demand both internally create a Recipe
 - The Recipe will go through the following phases
 - Copy s3 metadata to block storage location
 - Do a db2 write suspend (including S3 deletes off)
 - Create volume snapshots
 - Initiate S3 backup process
 - Backup metadata
 - Write resume (but leave S3 delete on)
 - When AWS S3 backup completes (tag backups and turn on S3 deletes)

Restore process, On-demand, internally creates a Recipe.

The recipe will go through the following phases

- Stop Db2
- Initiate meta data restore
- Initiate S3 restore
- Initiate volume snapshot restore
- When S3 restore is ready to use, scales all pods up

Db2 Warehouse (SaaS) – Gen 3 - Widest Analytic Workload Support

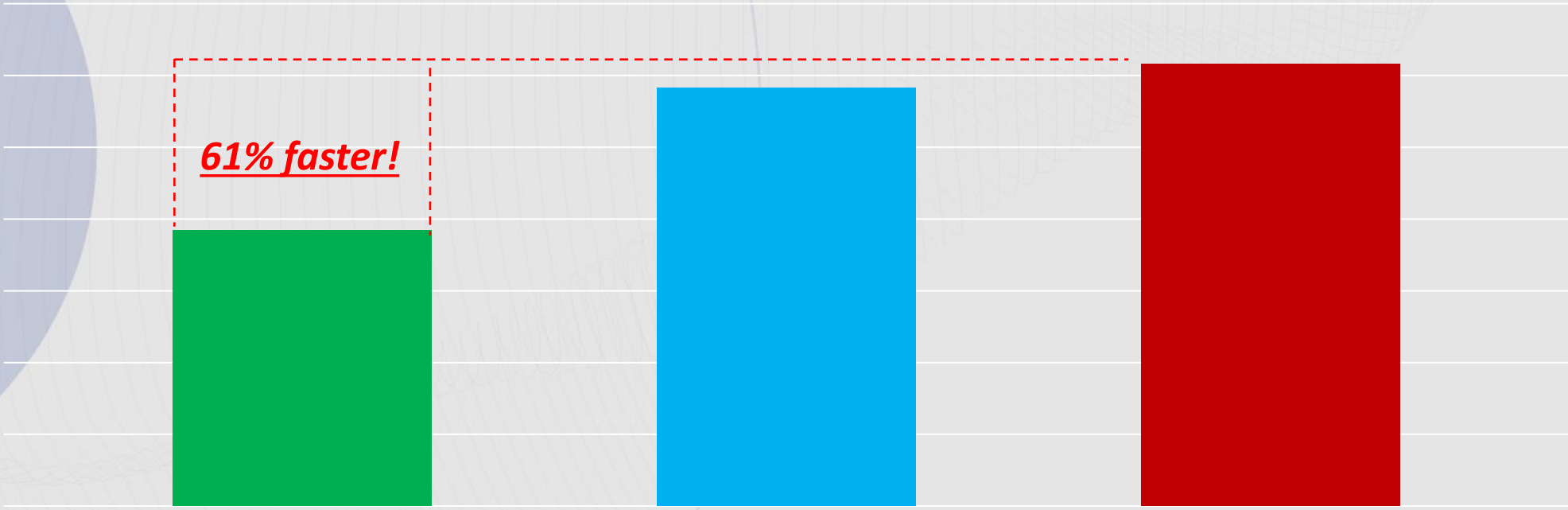
Db2 Warehousing
(Db2 DPF, Db2 Warehouse, Db2 Warehouse on Cloud, IIAS)

Competitor
Cloud Data Warehouse +
Lakehouse

	OLTP	Mixed Workloads		OLAP / Warehousing			
Workload	OLTP	Operational	Extreme Analytics	Operational Data Store	Operational Analytics	Analytics	Data Marts
Use	Transactions	Transactions with operational analytics	All types of analytic workloads and federated sources	Simple queries	Single-record look-up / IUD	Deep analytics – bulk scan	Computationally heavy / mining
Data Types	Traditional structured	Traditional structured + events + JSON + Graph + ...	Traditional structured + events + JSON + Graph + ...	Traditional structured	Traditional structured	Traditional structured	Traditional structured
Performance	1000s TPS	1000s TPS + 10s QPS	1000 QPS, many complex	100s QPS	10s-100s of S/IUD per second	100s QPS – 1000 QPS	Lower volume long running
Inserts	1000/s	1000/s	Continual Data Ingest	snapshots Seconds delay	Continual Data Ingest (Trickle feed) – OR - Batch - hourly/n-times daily/daily/etc		
Schema	Normalized	Normalized	De-normalized	Normalized	De-normalized	De-normalized	De-normalized
Applications	SQL	SQL + NoSQL + NewSQL	SQL + NoSQL + NewSQL	SQL	SQL	SQL	SQL

Db2 Warehouse (SaaS) – Gen 3 – Performance on AWS

TPCDS Derived Analytic Query Workload Runtime (99 queries)



■ Db2 Warehouse (AWS) ■ Cloud Warehouse Competitor ■ Lakehouse Competitor

Db2 Warehouse (SaaS) – Gen 3 – Product Tiers

	Starter Model	Middle-tier model	Large-scale Model
	Storage Optimized An MPP data warehouse ideal for cost-effective data warehousing, dev/test environment, infrequently accessed data. Support for Object Storage + DataLake tables	Compute Optimized An MPP data warehouse ideal for high-performance and production workloads Support for Object Storage + DataLake tables	Compute Optimized An MPP data warehouse ideal for high-performance and production workloads and very large systems Support for Object Storage + DataLake tables
Architecture	MPP (Massively Parallel Processing)	MPP (Massively Parallel Processing)	MPP (Massively Parallel Processing)
Configuration	Dedicated compute & storage	Dedicated compute & storage	Dedicated compute & storage
Cloud Platform Support	Amazon Web Services	Amazon Web Services	Amazon Web Services
Relative Cost	\$\$\$	\$\$\$\$	\$\$\$\$\$
Independently scale storage & compute	Yes	Yes	Yes
Storage for Table Data	Up to 80 TB (Block + Object Storage)	Up to 240 TB (Block + Object Storage)	Up to 600 TB (Block + Object Storage)
Compute Range	16 – 128 cores	48 – 576 cores	288– 2880 cores
Private Endpoint	Available	Available	Available



Cloud Reference Architectures

IBM Db2 pureScale on AWS Marketplace

The screenshot displays the AWS Marketplace interface for the IBM Db2 pureScale product. The product page shows the IBM logo and the text "IBM Db2 pureScale ready data management". A "Subscribe" button is visible. Overlaid on this is the EC2 console showing a list of instances. The instance list includes columns for Name, Instance ID, Instance state, Instance type, Status check, Alarm status, and Availability Zone. One instance, "pS-AWS-demo-1" with ID "i-017abb4e4c6b5b43d", is highlighted as "Running" with "2/2 checks passed". Below the list, the details for this instance are shown, including the Instance ID and IPv6 address.

Name	Instance ID	Instance state	Instance type	Status check	Alarm status	Availability Zone
pS-AWS-demo-2	i-06a37b7db2d0bb6f1	Terminated	c6i.xlarge	-	No alarms	us-east-2a
pS-AWS-demo-3	i-0088daecd9bc13eda	Terminated	c6i.xlarge	-	No alarms	us-east-2a
pS-AWS-demo-4	i-0e525a86b81dc0879	Running	c6i.xlarge	2/2 checks passed	No alarms	us-east-2a
pS-AWS-demo-1	i-017abb4e4c6b5b43d	Running	c6i.xlarge	2/2 checks passed	No alarms	us-east-2a
KT-Test	i-0df626c13bbc7a6e9	Running	c6i.xlarge	2/2 checks passed	No alarms	us-east-2a
psLargePerf-4	i-0f011c5b68ca5e596	Stopped	c6i.16xlarge	-	No alarms	us-east-2a
psLargePerf-2	i-0b429cf0bbc5f84f0	Stopped	c6i.16xlarge	-	No alarms	us-east-2a

Push-button deployment of self-managed pureScale on AWS through the Marketplace. Choose the region, t-shirt size configuration and get going.

Db2 pureScale deployed to customer's own account in **30-40 minutes**.



Db2U Update

Db2 Universal Container - Db2u

- Db2 database deployed as part of your micro-services architecture
- Multi-Container Architecture
- Open Kubernetes platform



Db2U on OCP/K8s – 11.5.9

- Support for Open K8s (no RHOS dependency)
- Support in 11.5.9 for Native COS
- Open Data Format Support
- AWS EKS Support – NCOS/ODF using OpenEBS Framework



Db2u container, operators and reference architecture

Container: Launch a standardized Db2 container across your entire deployment.

Operator: Easily manage the lifecycle of that container and various database functions with a well-defined, well-documented interface.

Reference architecture: Detailed architecture and guide on how to deploy Db2 across AWS, Azure and on-premises.



AWS EKS/ROSA



Azure AKS/ARO



OpenShift



Infrastructure



[Db2 on AWS](#)

[Db2 Warehouse on AWS EKS](#)

[Db2 Warehouse on AWS ROSA](#)



[Db2 on Azure](#)

[Db2 Warehouse on Azure AKS](#)

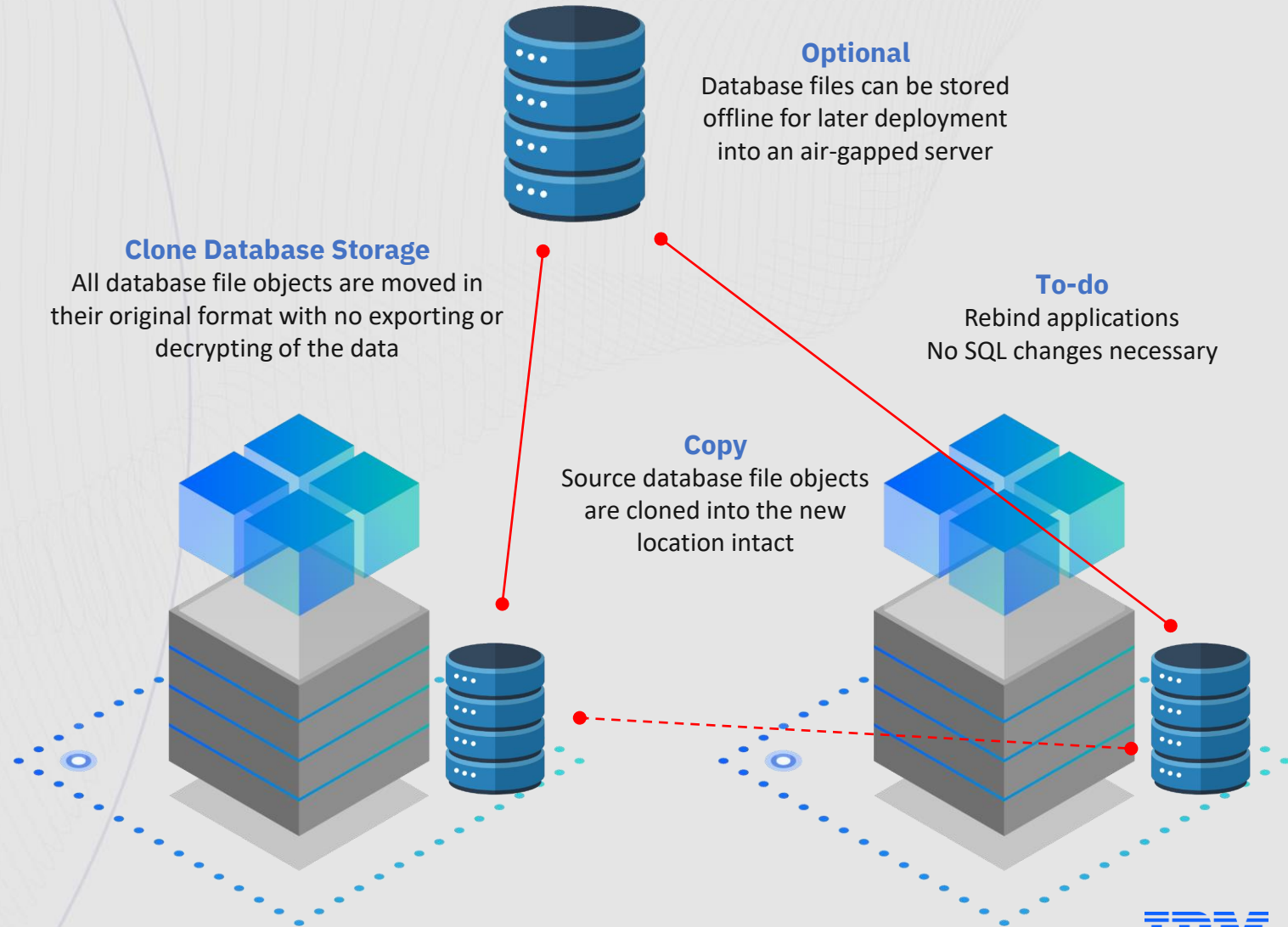
[Db2 Warehouse on Azure ARO](#)

Db2 Click to Containerize Tool

The fastest way to move your data from a traditional on-premise Db2 Linux X64 system to a Cloud Environment

- Move to Db2U on OpenShift, Kubernetes, or Cloud Pak for Data
- Containerize your data or move to a traditional server instance
- Command line or UI tool for flexibility
- With automatic database upgrades to V10.5, v11.1, V11.5+

More info: ibm.biz/c2cguide



Db2 Universal Container "Go" Operator Overview

Db2U embracing cloud-native modernization, leverage a "Go" Operator to manage the lifecycle of Db2.

Operational management

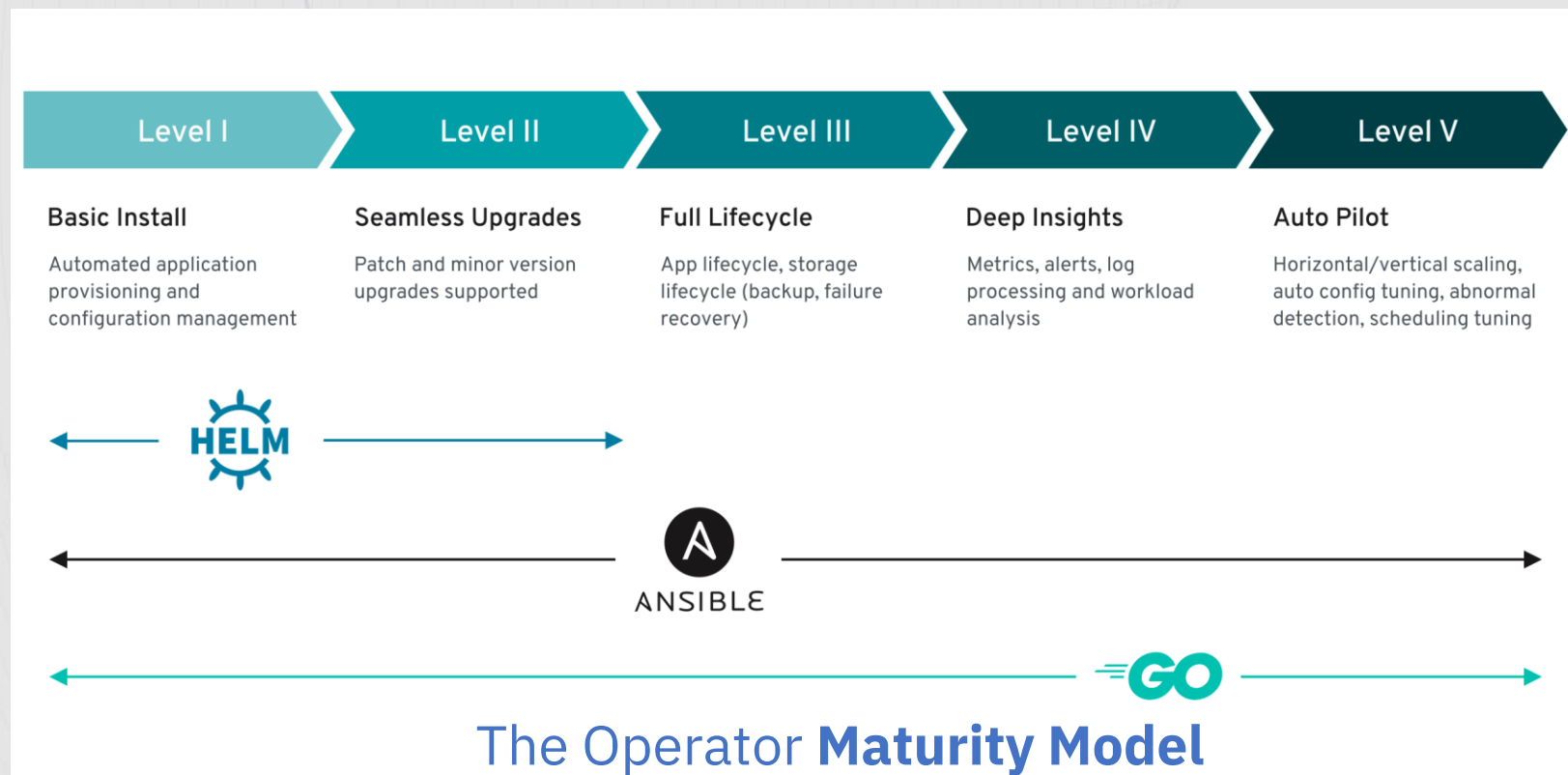
- Package, Configure, Deploy & Manage
- Deployment
- Management

Measured for completeness by a maturity

- Currently expanding Phase III to V

Delivery

- IBM Operator Catalog with support for air-gap deployments
- [Red Hat Marketplace](#)



Db2 Universal Container “Go” Operator capabilities

Db2 Operator is fully compliant with first two levels and have some capabilities in levels 3 and 4.

Day 2 Scaling – 2023 1H GA (Level 3)

- Vertical scaling (scale compute resources up/down)
- Horizontal scaling (scale compute resources out/in)

Day 2 BAR Controller – 2023 1H GA (Level 3)

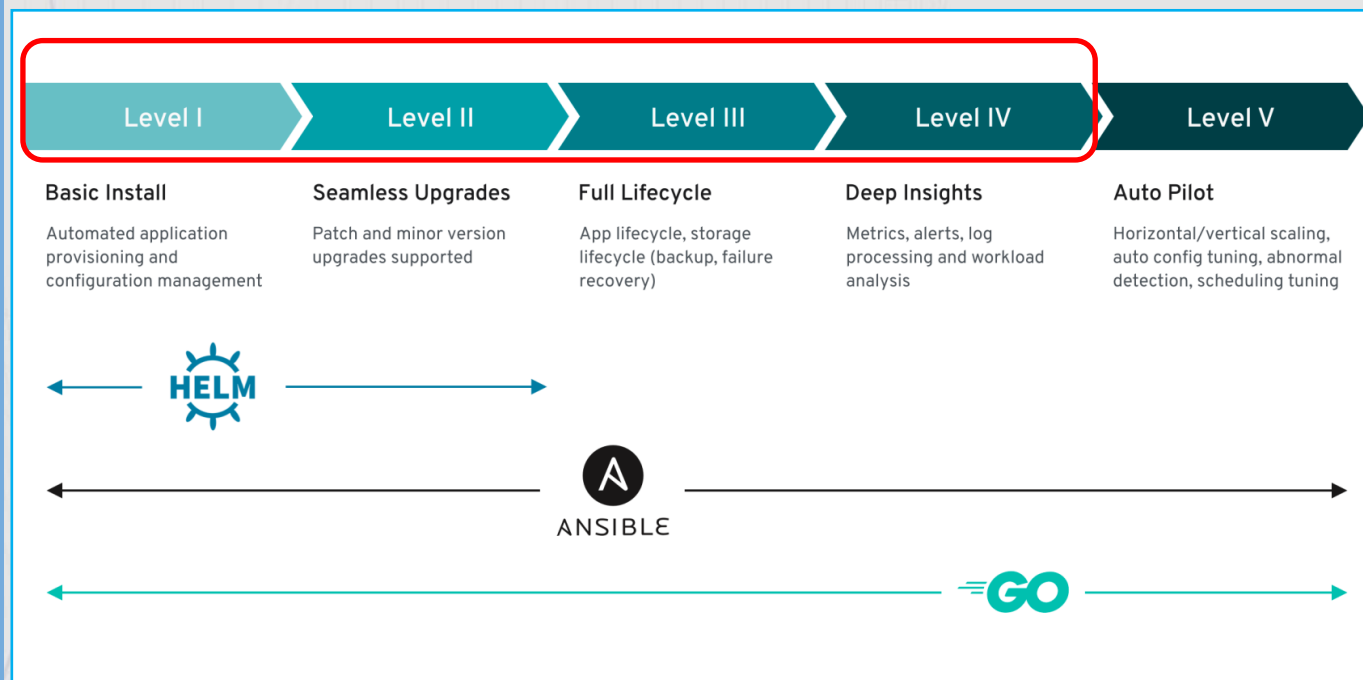
- Db2 native database backups
- Schema-level BAR

Day 2 HADR Controller – 2021+ (Level 3)

- API driven HADR for same cluster/namespace
- [1H 2023] Role-aware load balancer service for Cloud deployments. IPC Key Bank FTM4C is leveraging this.

Day 2 Log Streaming Controller – 2023 1H (Level 4)

- Built-around Fluentd data collector for a unified logging layer
- [CP4D 4.7+] Stream Db2 audit logs to zen-audit service



The Operator Maturity Model

Db2 Universal Container Performance Evaluation

Metric	Db2U on OpenShift
Install time	[Automated] <ul style="list-style-type: none"> ○ 4' – 8' (Db2 OLTP / Warehouse SMP) ○ 10' – 30 ' (Warehouse MPP)
Upgrade time	[Automated] 15' – 30' (Dependent on number of catalog objects)
[Db2 OLTP] HADR Takeover Time	[Automated] 1' – 3'
[Db2 WH MPP] Recovery on Db2 Failure	[Automated] <ul style="list-style-type: none"> ○ Non-catalog POD failure: 3' + Db2 Crash Recovery time ○ Catalog POD: 5' + Db2 Crash Recovery time
Recovery on Hardware Failure	[Semi-automated] OCP/k8s Node failure detection time + Db2 Recovery time
Vertical Scaling Compute	[Fully automated starting CP4D 4.6.0/11.5.8.0-cn1 release] 2 -3' + Db2 autoconfigure time



<https://www.redhat.com/en/resources/IBM-Db2-Warehouse-MPP-on-OpenShift-Container-Storage-detail>

Db2 Cloud Update

Les King

lking@ca.ibm.com

Christopher Hutchings

hutchcj@amazon.co.uk

September 2024

Data Server Day – Stockholm, Sweden