
MariaDB
OPEN WORKS
BE UNSTOPPABLE

MariaDB

OPENWORKS

USING COLUMNSTORE FOR DATA WAREHOUSING AND ANALYTICS WORKSHOP

ALLEN HERRERA, CUSTOMER ENGINEER, MARIADB

https://github.com/mariadb-AllenHerrera/openworks-2023

Analytics Technology Options

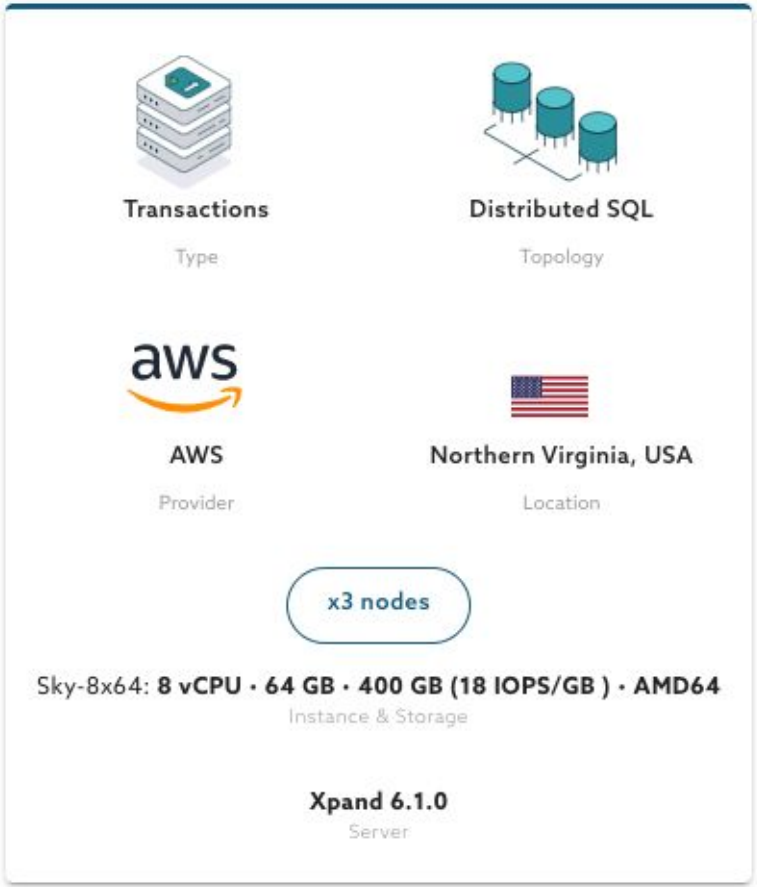
1. ColumnStore
    - Analytics on large datasets
    - Data warehousing functionality
2. Xpand w/ Columnar Indexes
    - Transactional Read/Write at massive scale
    - w/ Adhoc operational analytics
3. Serverless Analytics
    - Spark Integration with SkySQL databases
4. Q & A

MariaDB

Starting with SkySQL

cloud.mariadb.com

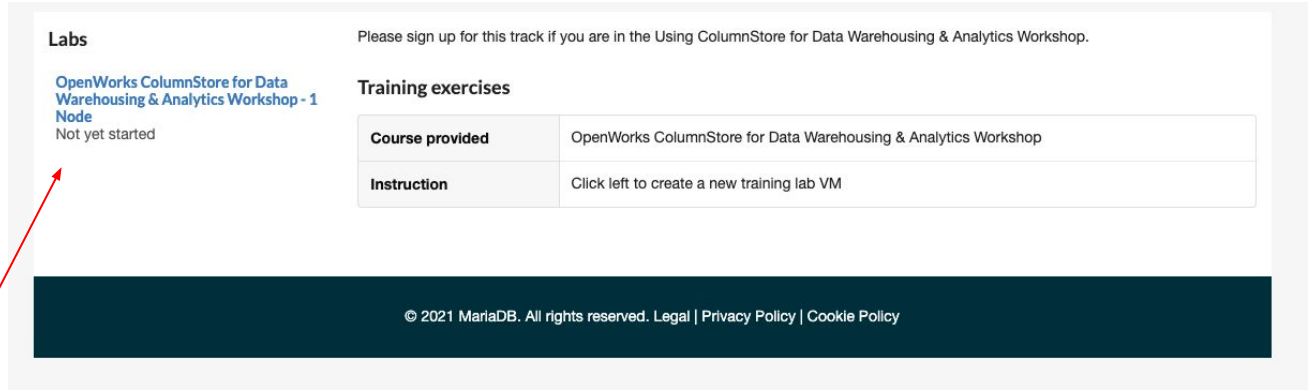
Deploy Xpand



MariaDB

Accessing EC2 Machine

https://training-exercises.mariadb.com/



Lab: OpenWorks ColumnStore for Data Warehousing & Analytics Workshop - 1 Node

This lab will create a one node exercise VM. You will use this lab during the Using ColumnStore for Data Warehousing & Analytics workshop at OpenWorks to use docker to deploy columnstore, access skysql and more ?

You have 1 try to complete this lab

Lab Environment

Your lab will be provisioned after you click start.

Start

Machine Info

Name	Login	Password	Public IP	GUI ID
mariadb-lab-instance-1	centos	********** Show	54.226.6.71	i-094f58b752fbc2cac

ssh centos@54.226.6.71

MariaDB

ColumnStore

ROW-ORIENTED VERSUS COLUMN-ORIENTED

```
SELECT Fname FROM TABLE1 WHERE State = 'NY';
```

ROW-ORIENTED

- Rows stored
- sequentially in a file
- Scans through every record row by row

ID	FNAME	LNAME	STATE	ZIP	PHONE	AGE	GENDER
1	Bugs	Bunny	NY	11217	(718) 938-3235	34	M
2	Yosemite	Sam	CA	95389	(209) 375-6572	52	M
3	Daffy	Duck	NY	10013	(212) 227-1810	35	M
4	Elmer	Fudd	ME	04578	(207) 882-7323	43	M
5	Witch	Hazel	MA	01970	(978) 744-0991	57	F

COLUMN-ORIENTED

- Each column is stored in a separate file
- Scans the only relevant column

ID	FNAME	LNAME	STATE	ZIP	PHONE	AGE	GENDER
1	Bugs	Bunny	NY	11217	(718) 938-3235	34	M
2	Yosemite	Sam	CA	95389	(209) 375-6572	52	M
3	Daffy	Duck	NY	10013	(212) 227-1810	35	M
4	Elmer	Fudd	ME	04578	(207) 882-7323	43	M
5	Witch	Hazel	MA	01970	(978) 744-0991	57	F

MariaDB

ColumnStore

EXTENT ELIMINATION

STORAGE ARCHITECTURE REDUCES I/O

- Only touch column files that are in filter, projection, group by, and join conditions
- Eliminate disk block touches to partitions outside filter and join conditions

EXTENT 1
ShipDate: 2016-01-12 - 2016-03-05

EXTENT 2
ShipDate: 2016-03-05 - 2016-09-23

EXTENT 3
ShipDate: 2016-09-24 - 2017-01-06

```
SELECT Item, sum(Quantity) FROM Orders
WHERE ShipDate
BETWEEN '2016-01-01' AND '2016-01-31'
GROUP BY Item;
```

Horizontal Partition
8 Million Rows
Extent 1

Horizontal Partition
8 Million Rows
Extent 2

Horizontal Partition
8 Million Rows
Extent 3

ID	ORDERID	LINE	ITEM	QUANTITY	PRICE	SUPPLIER	SHIPDATE	SHIPMODE
1	1	1	Laptop	5	1000	Dell	2016-01-12	G
2	1	2	Monitor	5	200	LG	2016-01-13	G
3	2	1	Mouse	1	20	Logitech	2016-02-05	M
4	3	1	Laptop	3	1600	Apple	2016-01-31	P
...	...	...	...	...	...	...	...	...
8M							2016-03-05	
8M+1							2016-03-05	
...	...	...	...	...	...	...	...	...
16M							2016-09-23	
16M+1							2016-09-24	
...	...	...	...	...	...	...	...	...
24M							2017-01-06	

ELIMINATED PARTITION

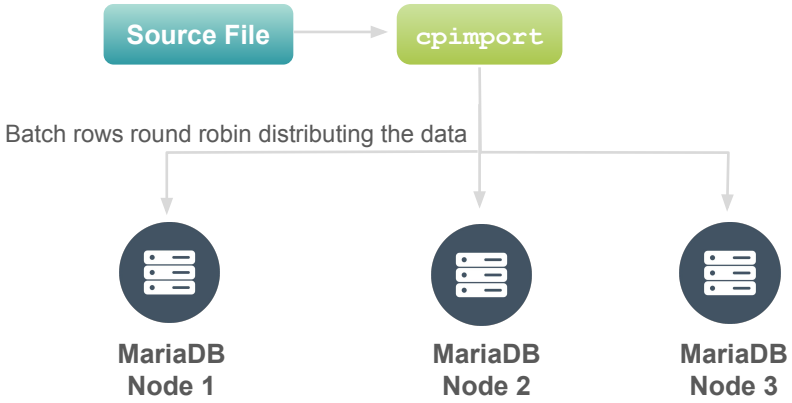
ELIMINATED PARTITION

MariaDB

ColumnStore

Cpimport

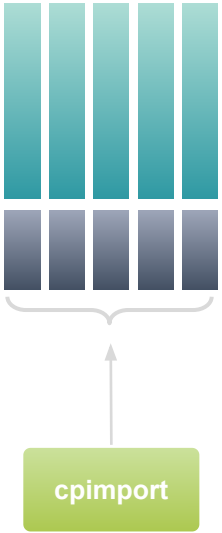
```
cpimport -m1 db1 table1 filename.tbl
```



Existing Data Stays Readable while data is being appended by the import

Appends Data to existing CS columns

s



MariaDB

Xpand

Tables and Indexes

Table - Base representation

id	col1	col2	col3
1	16	36	JANUARY
2	17	35	FEBRUARY
3	18	34	MARCH
4	19	33	APRIL
5	20	32	MAY

Index col(2)

col2	id
32	**5**
33	**4**
34	**3**
35	**2**
36	**1**

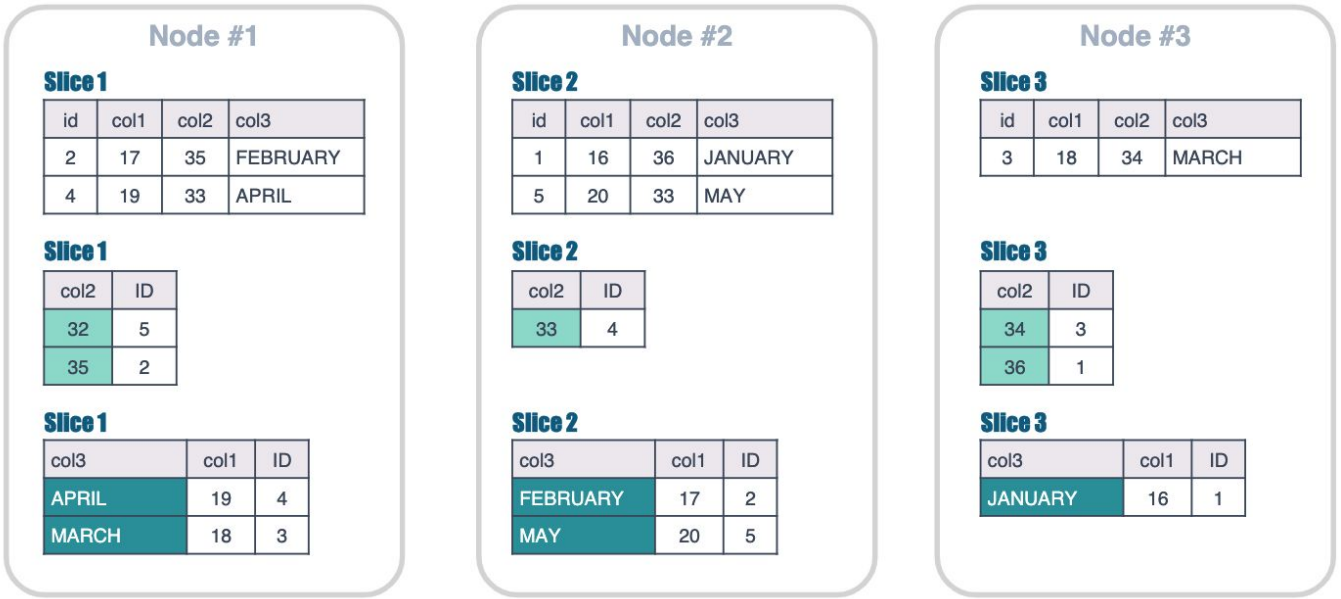
Index col(3, 1)

col3	col1	id
APRIL	19	**4**
FEBRUARY	17	**2**
JANUARY	16	**1**
MARCH	18	**3**
MAY	20	**5**

MariaDB

Xpand

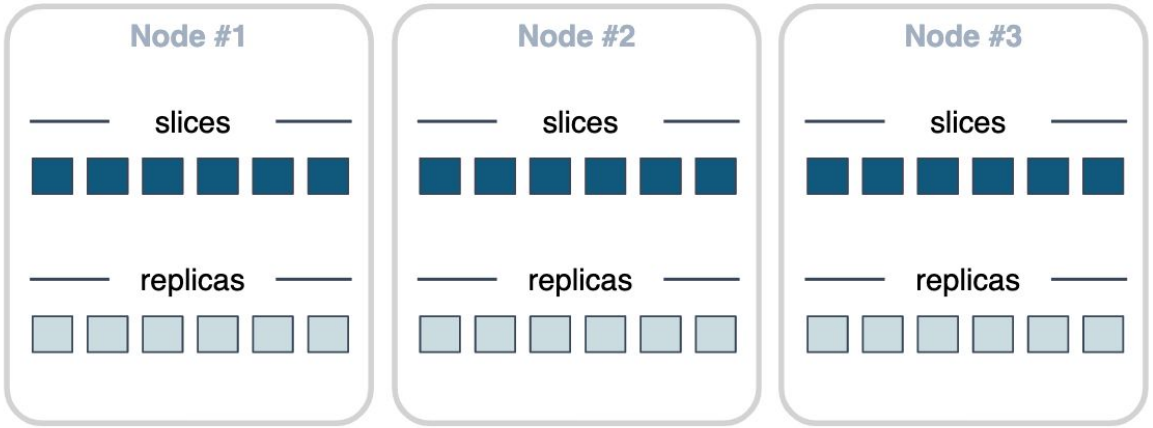
SLICES



MariaDB

Xpand

Self-healing

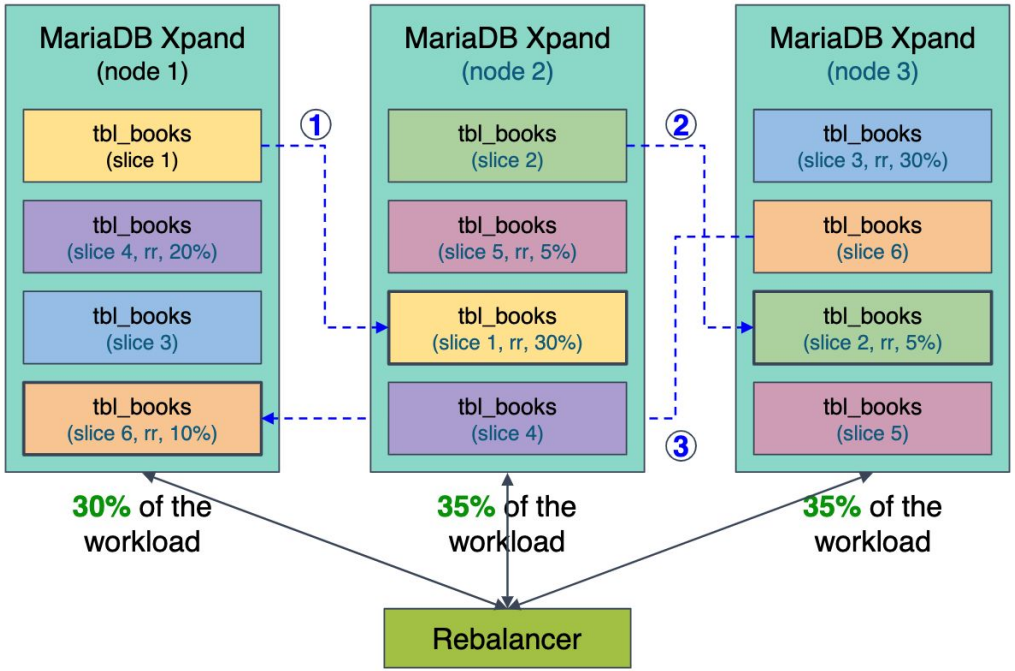


MariaDB

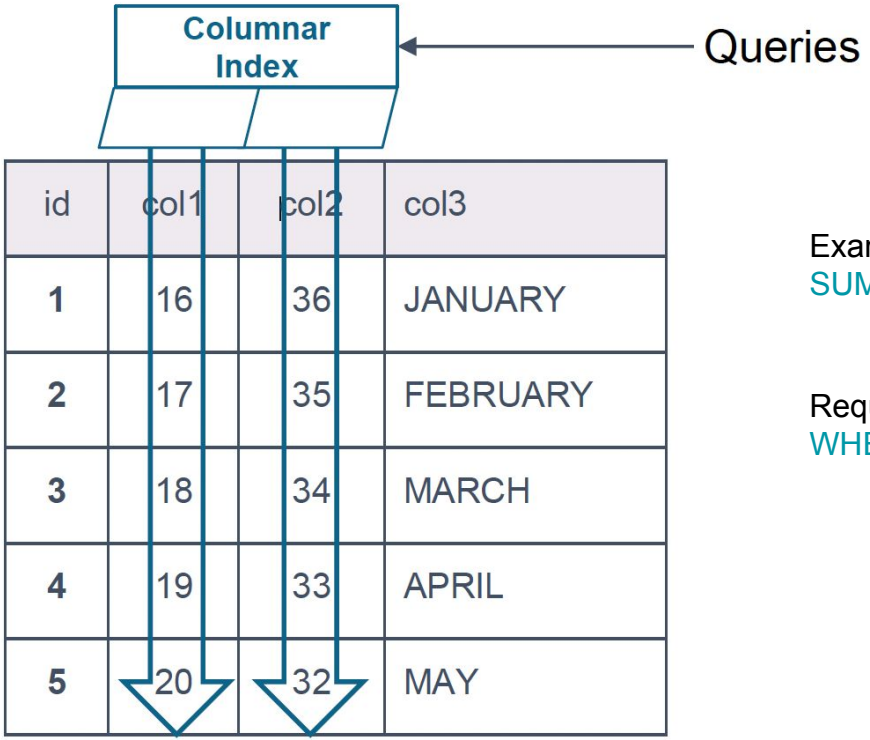
Xpand

Balance Nodes by Reranking

★ No transactions are blocked
★ No data movement occurs



Xpand Columnar Index

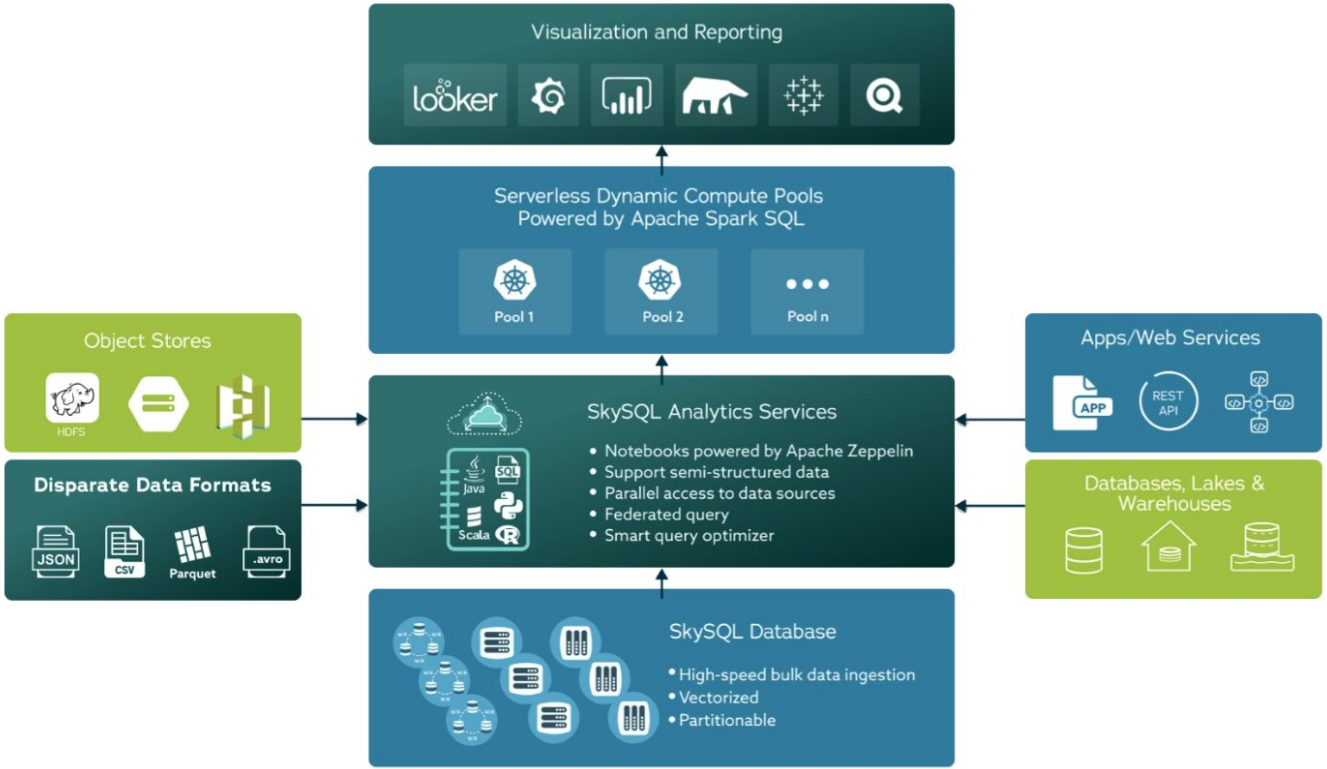


Example:
SUM, AVG, COUNT, etc.

Requirement:
WHERE =

MariaDB

Serverless Analytics



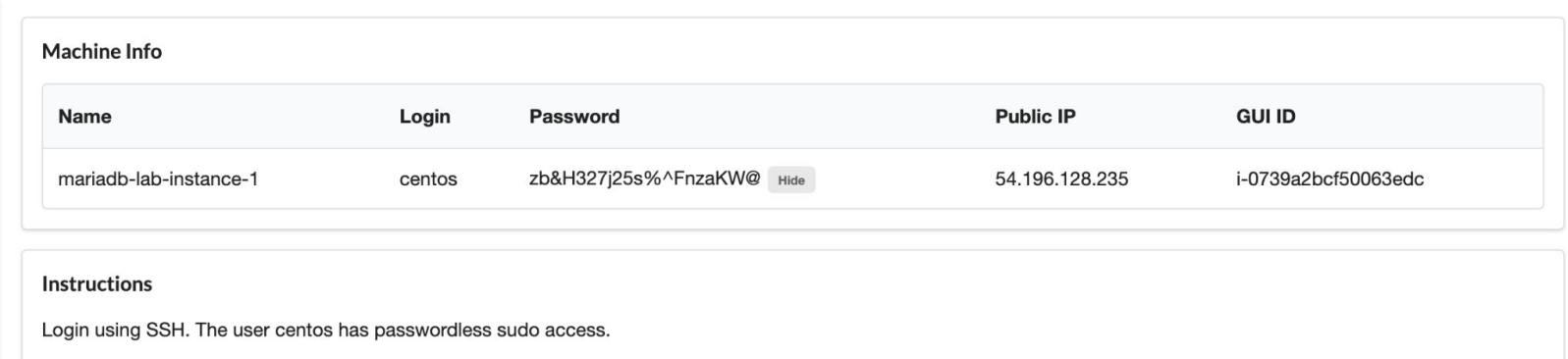
MariaDB

Hands on time!

https://training-exercises.mariadb.com/

Find course **OpenWorks ColumnStore**

Example EC2 Credentials:



Machine Info

Name	Login	Password	Public IP	GUI ID
mariadb-lab-instance-1	centos	zb&H327j25s%^FnzaKW@ Hide	54.196.128.235	i-0739a2bcf50063edc

Instructions

Login using SSH. The user centos has passwordless sudo access.

Any issues raise your hand, have your email address handy.

MariaDB

Columnstore Hands On

- SSH to supplied VM
- Install docker
- Clone git repo
- Setup env file
- Startup columnstore
- Import flights data
- Run sample queries
- Review extents
- Test extent elimination improvement
- Review query execution plan

MariaDB

Xpand Hands On

- Begin deploying serverless analytics
- Connect to skysql xpand
- Create table without index
- Import flights data
- Baseline queries
- Alter table w/ columnar index
- Retest queries

MariaDB

Serverless Analytics Hands On

- Access SkyBook
- Learn about catalogs
- Create connection to EC2 Columnstore
- Create a table in SA, load the data from columnstore
- Query flight data from xpand
- Run queries against SA

MariaDB


MariaDB
OPEN WORKS
THANK YOU
ANY QUESTIONS?

NEXT STEPS

Check out these resources to learn more about MariaDB

- **Other sessions**
  - Using MariaDB ColumnStore with Power BI for Visualization and Reporting
  - SkySQL Serverless Analytics Powered by SparkSQL - PART 1 & 2
  - Plug Into Analytics: Connecting ColumnStore to Source Databases with Spider Engine
- **More Resources**
  - https://github.com/mariadb-AllenHerrera/opennworks-2023
  - https://github.com/mariadb-corporation/mariadb-columnstore-docker
  - https://skysql.mariadb.com/dashboard


MariaDB
OPEN WORKS
BE UNSTOPPABLE