Big Data Analytics on the Cloud: Deploy big data applications with Redshift, Amazon S3 and Google Big Query
This presentation may include statements that constitute “forward-looking statements” for purposes of the safe harbor provisions under the Private Securities Litigation Reform Act of 1995, including descriptions of technology and product features that are under development and estimates of future business prospects. Forward-looking statements inherently involve risks and uncertainties that could cause actual results of MicroStrategy Incorporated and its subsidiaries (collectively, the “Company”) to differ materially from the forward-looking statements.

Factors that could contribute to such differences include: the Company’s ability to meet product development goals while aligning costs with anticipated revenues; the Company’s ability to develop, market and deliver on a timely and cost-effective basis new or enhanced offerings that respond to technological change or new customer requirements; the extent and timing of market acceptance of the Company’s new offerings; continued acceptance of the Company’s other products in the marketplace; the timing of significant orders; competitive factors; general economic conditions; and other risks detailed in the Company’s Form 10-K for the year ended December 31, 2016 and other periodic reports filed with the Securities and Exchange Commission. By making these forward-looking statements, the Company undertakes no obligation to update these statements for revisions or changes after the date of this presentation.
Agenda

• Redshift
• Google Big Query
• S3
MicroStrategy Sends Analytical Queries to Amazon Redshift

Characteristics of an Analytical Query
• Access large data volumes
• Process large data volumes

Typical Challenge:
• Achieve interactive response time

Redshift Key features
Massively Parallel Processing
Columnar Database
Compression
Zone maps
Amazon Redshift Demo

There are numerous ways for MicroStrategy to interact with Amazon Redshift:

- Ad-hoc Schema
- Live Connect
- In-Memory Cube
- Modeled Schema
- JDBC/ODBC drivers
Data Distribution Styles to Enhance Parallel Processing

**Pro Tips:**

- DISTRIBUTION Key → Frequently joined tables
- ALL → Small dimension tables
- EVEN → Large tables that don’t have common join keys

---

**Distribution Key**
Same key to same location

**All**
All data on every node

**Even**
Round robin distribution

---

Slice 1 | Slice 2 | Node 1
---|---|---
Key 1 | Key 2 | | Slice 3 | Slice 4 | Node 2

Slice 1 | Slice 2 | Slice 3 | Slice 4 | Node 1
---|---|---|---|---
| Key 1 | Key 2 | | Node 2

Slice 1 | Slice 2 | Slice 3 | Slice 4 | Node 1
---|---|---|---|---
| Node 2

Slice 1 | Slice 2 | Slice 3 | Slice 4 | Node 1
---|---|---|---|---
| Node 2
MicroStrategy Generates Redshift-Specific SQL Syntax

MicroStrategy integrates with Redshift’s broad list of database functions and SQL functionalities to improve analytical performance.

- 128 Function patterns pushed down
- 23 Data types supported
- 26 Unique VLDB properties
MicroStrategy is Optimized When Creating Temporary Tables
Tables created with implicit syntax are compressed, distributed and sorted

Implicit Table Creation Syntax
Example:
create temporary table ZZTNW2U6VGQPO000 as
select distinct a12.year_id
year_id,
anl.item_nbr WJXBFS1,
anl.class_nbr WJXBFS2
from store_item_93 all
join lookup_day a12
on (a11.cur_trn_dt = a12.cur_trn_dt)

Data is compressed
Data is distributed
Data is sorted
Customize Intermediate Result Sets Distribution for Improved Performance

Tables created using explicit syntax can be customized

```sql
create temporary table ZZZZTXTBGZYY6PO00000000 (  
  year_id  NUMERIC(10, 0),  
  WJXBFS1  NUMERIC(10, 0),  
  WJXBFS2  NUMERIC(10, 0))  
DISTKEY (year_id)

insert into ZZZZTXTBGZYY6PO00000000  
select distinct a12.year_id, year_id,  
  a11.item_nbr WJXBFS1,  
  a11.class_nbr WJXBFS2  
from store_item_93 a11  
join lookup_day a12  
on (a11.cur_trn_dt =  
  a12.cur_trn_dt)
```
Managing MicroStrategy Workload in Redshift

Redshift creates multiple queues to prioritize queries at runtime based on their user group or query group.

Set and reset report specific Redshift query groups in MicroStrategy

<table>
<thead>
<tr>
<th>VLDB Category</th>
<th>VLDB Property Setting</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pre/Post Statements</td>
<td>Report Pre Statement</td>
<td>set query_group to ‘MSTR_High=!o;’</td>
</tr>
<tr>
<td>Pre/Post Statements</td>
<td>Clean-up Post statement</td>
<td>reset query_group;</td>
</tr>
</tbody>
</table>
Managing MicroStrategy Workload in Redshift

Amazon Redshift

MicroStrategy

Redshift WLM
Directs queries to queue based on user and query group

MSTR_HIGH
Queue
MSTR_MED
Queue
MSTR_LOW
Queue
Super User Queue
ETL & Other Apps Queue

Redshift Super User Queries
ETL and Other Applications

MSTR_HIGH = Exec Dashboard
MSTR_MED = Analyst Report
MSTR_LOW = Developer Report

ETL and Other Applications
Connect Securely to Amazon Redshift with MicroStrategy Driver
MicroStrategy recommends using encrypted data connections

Download and install the public key available on Amazon website

Configure the following parameters in the Security tab of ODBC Administrator

- Set Encryption Method = “1-SSL”
- Check the Validate Server Certificate checkbox
- Set Trust Store= <certificate location> \<certificate filename>
Summary

MicroStrategy and Amazon Redshift continue to have a strong partnership

Multi-faceted technical integration of products

Continued optimization provides a seamless reporting experience

Redshift integration paper: TN47453
Agenda

- Redshift
- Google Big Query
- S3
What is Google BigQuery?

Data Warehouse service on Google Cloud Platform

- Petabyte scale
- Low cost
- Columnar storage
- Familiar SQL
- Analytics AS-A-Service
- IO of 1TB in 1 second
## What is Google BigQuery?

### Cost

<table>
<thead>
<tr>
<th>Action</th>
<th>Cost</th>
<th>Notes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Storage</td>
<td>$0.02 per GB, per month</td>
<td>See Storage pricing.</td>
</tr>
<tr>
<td>Long Term Storage</td>
<td>$0.01 per GB, per month</td>
<td>See Long term storage pricing.</td>
</tr>
<tr>
<td>Streaming Inserts</td>
<td>$0.05 per GB</td>
<td>See Storage pricing.</td>
</tr>
<tr>
<td>Queries</td>
<td>$5 per TB</td>
<td>First 1 TB per month is free, see On-demand pricing for details. Flat-rate pricing is also available for high-volume customers.</td>
</tr>
<tr>
<td>Loading data</td>
<td>Free</td>
<td>See Loading data into BigQuery.</td>
</tr>
<tr>
<td>Copying data</td>
<td>Free</td>
<td>See Copying an existing table.</td>
</tr>
<tr>
<td>Exporting data</td>
<td>Free</td>
<td>See Exporting data from BigQuery.</td>
</tr>
<tr>
<td>Metadata operations</td>
<td>Free</td>
<td>List, get, patch, update and delete calls.</td>
</tr>
</tbody>
</table>
Google Big Query Demo

There are numerous ways for MicroStrategy to interact with:

- Ad-hoc Schema
- Live Connect
- In-Memory Cube
- Modeled Schema
- JDBC/ODBC Drivers
Simba

- Direct BI connectivity to data without extraction
- Supports query with BigQuery functions (UDFs)
- Maps SQL-92 to BigQuery’s SQL subset

How to connect to data stored in Google BigQuery via MicroStrategy's ODBC connectors
• Windows Google BigQuery DSN:
Benefits of Using MicroStrategy with Google BigQuery

• Support for 100% of Google BigQuery’s data types
• MicroStrategy supports UDFs defined
• Web Analytics
• Social Analytics: Tweets and Facebook posts about my company
Amazon S3

- Complete storage platform
- Simplicity
- Durability
- Scalability
- Security
- Integration with other Services
- Stores data as objects within resources called “buckets”

Ref: https://aws.amazon.com/s3/
Connecting MicroStrategy with Amazon S3

- JDBC Athena driver
- MicroStrategy S3 Connector
- File URL
Connecting MicroStrategy with Amazon S3: Athena
MicroStrategy S3 Connector

- Based on MicroStrategy Data Connector SDK
- Allows for file browsing.
- Supports JSON, CSV, text files
- Concurrent connections

S3 Connector Installation instructions:
https://community.microstrategy.com/s/article/Amazon-S3
MicroStrategy S3 Connector

Prerequisites

- Node.js with Node Package Manager installed on the MicroStrategy Web Server
- Valid certificate in PFX format on the MicroStrategy Web Server
- Registered Amazon application
- Amazon S3 bucket is configured with proper settings
- For detailed information regarding these pre-requisites, click here

Installation

1. Download the zip file attached to this listing and extract its contents to a server for hosting the connector
2. Open the file server.js located within the AmazonWDC-GA directory
   2.1. Place the PFX certificate file under the root folder of the plugin
   2.2. Update line 63 to the port made available for this connector

   ```javascript
   var secureServer = https.createServer(server_options,app).listen(7070);
   ```

   2.3. Update line 58 and replace myPfx.p12 with the name of your PFX certificate file

   ```javascript
   var server_options = {
     pfx: fs.readFileSync(__dirname + '/myPfx.p12'),
     passphrase: 'mstr123'
   };
   ```

3. Within the registered Amazon S3 application, set the Allowed Return URLs URL as http://<your-host-name>:<port>/redirect
4. Update the config.js file within the AmazonWDC-GA directory with the information obtained in step 3

   "callbackURL": "https://10.22.41.10:7070/auth/box/callback",
   "AWS_CLIENT_ID": "xxxxxxxxxxxxx",
   "AWS_CLIENT_SECRET": "xxxxxxxxxxxxxxxxxx"

S3 Connector Installation instructions:
https://community.microstrategy.com/s/article/Amazon-S3
MicroStrategy S3 Connector
MicroStrategy File URL import

Needs to edit the bucket policy in order for Web-Server I-Server to read the file

<table>
<thead>
<tr>
<th>Name</th>
<th>Last modified</th>
<th>Size</th>
<th>Storage class</th>
</tr>
</thead>
<tbody>
<tr>
<td>FlightPosition.csv</td>
<td>May 16, 2017 2:10:29 PM GMT-0400</td>
<td>92.6 KB</td>
<td>Standard</td>
</tr>
<tr>
<td>data.sql</td>
<td>May 17, 2017 10:12:22 AM GMT-0400</td>
<td>5.5 GB</td>
<td>Standard</td>
</tr>
<tr>
<td>structure.sql</td>
<td>May 17, 2017 10:21:57 AM GMT-0400</td>
<td>504.8 KB</td>
<td>Standard</td>
</tr>
</tbody>
</table>
Thank you
Sergio Sainz   ssainz@microstrategy.com