Query anything, anywhere

Query many data sources as one

Data Virtualization

Dave Williamson
Business Development, Emerging Analytics Technology
dwilliamson@ca.ibm.com
Data is Everywhere and increasingly heterogeneous
Performing Analytics Today

Costly and Complex

High Latency from source to use

Does not meet Business need

Compute resources at source not utilized

Error prone, data integrity challenges

Applications expect homogeneity

Not scalable

Not all data needs to be moved or copied
Resulting Data Architectures

Numerous ETLs

Unnecessary duplication, replication

Data governance issues accelerating
What is Data Virtualization?

The ability to view, access, manipulate and analyze data without the need to know or understand its physical format or location.
A new approach to Data Virtualization

Now in beta trial. Coming to ICP for Data in November

Query anything, anywhere.
Query many heterogenous data sources as one across cloud, on-premise and mobile with advanced analytics using the most popular languages and tools

Simplicity and scalability.
Automatically discover, and connect few to many devices and data stores into a single self balancing constellation. Avoid the complexity of centralized copies. Data only persists at the source.

Execution speedup.
Many times acceleration using the power of every device to compute and aggregate results.

Security.
Fully secure and encrypted communication and preservation of data access rights at source.
What is fundamentally different?

**Classic Federation & Edge Computing**

- **Query issued against the system**
- **A coordinator receives the request and fans the work out to edge nodes**
- **Edge nodes individually perform as much work as they can based on their own data. Individual results are sent back to the coordinator for final merging and remaining analytics.**
- **Coordinator receives intermediary results from all edge nodes, merges results, and performs remaining analytics.**

**New Computational Mesh**

- **Query issued against the system**
- **A coordinator receives the request and fans the work out to edge nodes**
- **Edge nodes self organize into a constellation where they can communicate with a small number of peers. Nodes collaborate to perform almost all analytics, not only analytics on their own data.**
- **Coordinator receives mostly finalized results from just a fraction of nodes. Completes the final work for the query result.**
## Supported Languages & Data Sources

### Query Languages

<table>
<thead>
<tr>
<th>Query Languages</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>SQL (ANSI)</td>
<td>✓</td>
</tr>
<tr>
<td>SQL (Oracle)</td>
<td>✓</td>
</tr>
<tr>
<td>SQL (DB2)</td>
<td>✓</td>
</tr>
<tr>
<td>SQL (PostgreSQL, Netezza)</td>
<td>✓</td>
</tr>
<tr>
<td>Spark SQL</td>
<td>✓</td>
</tr>
<tr>
<td>SQL Python</td>
<td>✓</td>
</tr>
<tr>
<td>SQL in R &amp; SparkR</td>
<td>✓</td>
</tr>
</tbody>
</table>

### Mix Any Combination of Data Sources

<table>
<thead>
<tr>
<th>Data Sources</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Oracle</td>
<td>✓ Excel</td>
</tr>
<tr>
<td>Db2 HDM family</td>
<td>✓ CSV (delimited text)</td>
</tr>
<tr>
<td>IIAS</td>
<td>✓ Cloudera</td>
</tr>
<tr>
<td>PDA/Netezza</td>
<td>✓ Db2/Z</td>
</tr>
<tr>
<td>PostgreSQL</td>
<td>✓ Db2 Event Store</td>
</tr>
<tr>
<td>Informix</td>
<td>✓ Apache/HDP Hive</td>
</tr>
<tr>
<td>MySQL</td>
<td>✓ Teradata</td>
</tr>
<tr>
<td>SQLServer</td>
<td>✓ Db2 iSeries</td>
</tr>
<tr>
<td>Apache Derby</td>
<td>✓ Db2 zSeries</td>
</tr>
<tr>
<td>Big SQL</td>
<td>✓ Manual driver install</td>
</tr>
</tbody>
</table>

Using Common SQL Engine
IBM Data Virtualization Manager for z/OS

The industry’s only Z-resident data virtualization solution!

Virtualize z/OS data with other enterprise data sources in real-time without data movement to provide comprehensive information that is readily consumable by analytics, cloud and cognitive applications

- ensures data is secure and in-place with real-time data virtualization
- supports Hybrid architectures (on-premises, Cloud, Hadoop and MF)
- broad API support via SQL, NoSQL, SOAP, and REST via z/OS Connect EE
- z/OS resident optimization for improved performance and TCO
- abstraction layer for improved productivity and business agility

Take action now!

✓ accelerate mainframe modernization initiatives involving Big Data
✓ gain real-time business insights across z/OS and enterprise data
✓ eliminate coding of complex z/OS apps via built-in APIs and interfaces

Get started

YouTube channel – IBM Data Virtualization Manager for z/OS
Contact mink@us.ibm.com or douglasd@us.ibm.com for more information
ICP for Data
Use Case: **Manage All Your Data** – regardless of where it lives

**Governance of Consumers and Producers**
With lineage, context & metrics

**Your Data Anywhere**
- On-Premises
- Private Cloud
- Public Cloud
- Structured & Unstructured Data

**End Users**
- Data Engineer
- Data Scientists
- Business Analysts
- App Developers

**Enterprise Search & Catalog**

**IBM Data Virtualization**

**Your Data Anywhere**

**Governance of Consumers and Producers**

**Structured & Unstructured Data**
- Public Cloud
- Private Cloud
- On-Premises

**End Users**
- Data Engineer
- Data Scientists
- Business Analysts
- App Developers