Frictionless Hybrid Data Access for Clouds

Greg Stasko, Principal Sales Engineer for Alliances - Progress DataDirect
Rik Tamm Daniels, VP Technology and Partnerships - Paxata
Audio Bridge Options & Question Submission
Frictionless Hybrid Data Access for Clouds

Greg Stasko, Principal Sales Engineer for Alliances - Progress DataDirect

Rik Tamm Daniels, VP Technology and Partnerships - Paxata
During this webinar, you’ll learn:

- Hybrid data access patterns for cloud applications
- Common approaches to accessing business data behind the firewall
- Overview of Hybrid Data Pipeline, a cloud-independent data gateway for cloud ISVs
- How customers can access on-premises data using the cloud version of Paxata's self-service data preparation platform.
Hybrid data access patterns for cloud applications
Evolution of deployment from O/S to Cloud Infrastructure (1200 survey respondents)

Amazon Web Services

CenturyLink

Digital Ocean

Google Cloud

IBM

Microsoft Azure

Oracle

Rackspace

Salesforce

VMware

Other popular responses

Linode

Pironet

Redhat OpenShift

OpenStack

Cloud Share

Thomson Reuters Elektron

SAP HANA

Claro Cloud
Popular Relational/Analytics Data Sources (1200 survey respondents)

Adoption

- SQL Server: 18.70%
- Oracle: 12.89%
- MySQL: 12.77%
- Progress OpenEdge: 7.93%
- PostgreSQL: 5.65%
- Microsoft SQL Azure: 5.27%
- IBM DB2: 4.76%
- SQLite: 3.68%
- Teradata: 2.61%
- SAP HANA: 2.30%
- MariaDB: 2.25%
- Sybase ASE: 1.92%
- Amazon Redshift: 1.79%
- Informix: 1.64%
- Sybase IQ: 1.30%
- Netezza: 1.25%
- Other (please specify): 1.13%
- Amazon Aurora: 1.00%
- Not sure: 0.97%
- Pivotal Greenplum: 0.87%
- Google BigQuery: 0.77%
- Vertica: 0.61%
Popular Big Data Sources (1200 survey respondents)

<table>
<thead>
<tr>
<th>Source</th>
<th>Adoption</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hadoop Hive</td>
<td>18.53%</td>
</tr>
<tr>
<td>Spark SQL</td>
<td>8.17%</td>
</tr>
<tr>
<td>Hortonworks</td>
<td>7.97%</td>
</tr>
<tr>
<td>Cloudera CDH</td>
<td>7.87%</td>
</tr>
<tr>
<td>Cloudera Impala</td>
<td>7.47%</td>
</tr>
<tr>
<td>Apache Solr</td>
<td>7.37%</td>
</tr>
<tr>
<td>Oracle BDA</td>
<td>6.67%</td>
</tr>
<tr>
<td>Amazon EMR</td>
<td>5.98%</td>
</tr>
<tr>
<td>Apache Sqoop</td>
<td>5.48%</td>
</tr>
<tr>
<td>MapR</td>
<td>5.38%</td>
</tr>
<tr>
<td>IBM BigInsights</td>
<td>4.68%</td>
</tr>
<tr>
<td>Apache Storm</td>
<td>4.08%</td>
</tr>
<tr>
<td>Apache Drill</td>
<td>2.39%</td>
</tr>
<tr>
<td>Apache Phoenix</td>
<td>2.39%</td>
</tr>
<tr>
<td>SAP Altiscale</td>
<td>2.19%</td>
</tr>
<tr>
<td>Pivotal HD</td>
<td>1.89%</td>
</tr>
<tr>
<td>Presto</td>
<td>0.80%</td>
</tr>
<tr>
<td>GemFireXD</td>
<td>0.70%</td>
</tr>
</tbody>
</table>
Popular NoSQL Sources (1200 survey respondents)

- **MongoDB**: 35.60%
- **Cassandra**: 14.57%
- **HBase**: 10.34%
- **Oracle NoSQL**: 9.01%
- **Redis**: 8.45%
- **Other (please specify)**: 6.01%
- **Couchbase**: 5.78%
- **DynamoDB**: 2.78%
- **DataStax Enterprise**: 2.22%
- **SimpleDB**: 2.22%
- **MarkLogic**: 1.67%
- **Aerospike**: 0.78%
- **Riak**: 0.56%
POLL #1a

- What On-premises databases do you see value in accessing from Cloud [multiple selections]?
  - Oracle
  - SQL Server
  - IBM DB2
  - PostgreSQL
  - MySQL
POLL #1b

- What On-premises databases do you see value in accessing from Cloud [multiple selections]?
  - Netezza
  - Teradata
  - Vertica
  - Greenplum
  - Hadoop Hive *(all distributions)*
POLL #1c

- What On-premises databases do you see value in accessing from Cloud [multiple selections]?
  - MongoDB
  - Cassandra
  - HBase
  - Spark SQL
  - Other [Please enter]
Scenario 1: Cloud analytics with on-premises data marts

Business Intelligence (BI) and Corporate Performance Management (CPM) software vendor known for its BOARD toolkit.

“BOARD’s first customer to leverage the connector was a fashion retailer in Belgium, with more than 125 shops in Belgium and Luxembourg. To date, six BOARD customers are leveraging the DataDirect connector, but Ferrari expects 75-80 percent of BOARD’s customers to purchase them over the next couple of years.”

Pietro Ferrari
Chief Technology Officer
Scenario 2: SaaS CRM Integration with legacy systems

Intuit develops financial and tax preparation software and related services for small businesses, accountants and individuals.

“I wanted to know about implementation complexity. Lekhter told me that each rollout to a new user group took only a week of development. They push new business capabilities into production on a monthly basis. Lekhter credits the ease of development to the heavy lifting provided by Lightning Connect and Progress connectivity.”

Jerry Lekhter
Director of Engineering

intuit
Scenario 3: Hybrid Data Management with SaaS sources

Result: Chatham Financial today has a data warehouse that meets stringent security standards, did not require internal resources to develop and deployed months faster when compared to using multiple vendors or building a custom solution.

With the financial services industry being highly competitive, Chatham Financial takes marketing seriously, continuously seeking ways to leverage its data to advance the business—but security always comes first. “Our clients expect top notch security from us, so we have to expect that from our vendors as well.”

Matt Roberts
Marketing Analytics Manager

“Salesforce, Oracle Service Cloud, Google Analytics, Oracle Eloqua, Marketo”
Scenario 4: ERP moving to cloud requiring open analytics

A leading global provider of enterprise software and information solutions migrated their business applications to the cloud, but end customers lost the ability to access data for reporting and analytics making the cloud unattractive.

Connectivity was restored for the cloud-hosted application with DataDirect Hybrid Connectivity that enabled secure and firewall-friendly access to the core ERP database. End customers were able to adopt the cloud ERP solution with standards based data access across:

1) SQL (ODBC and JDBC) connectivity for BI tools
2) REST (OData) interfaces for clicks not code integration other clouds such as Salesforce
Common approaches to accessing business data behind the firewall
Moving workloads to the cloud makes it much more difficult to provide users with **secure yet simple** access to data.
There is no standard for accessing data behind a firewall!

1. Network-Based VPN (Cloud-specific)
2. SSH Tunneling
3. Reverse Proxy Servers
Firewall Becoming Barrier for Hybrid Data Tech Adoption

Source: The 2017 State of the Firewall" produced by Firemon
Hybrid Data Pipeline – a cloud-agnostic data gateway for cloud ISVs
Progress® DataDirect® Hybrid Data Pipeline™

Completely transform the way cloud apps access data
Progress® DataDirect® Hybrid Data Pipeline is the industry’s first hybrid data pipeline that can run independently and integrate with any single or multi-vendor technology stack, enabling data access via open standards such as SQL and Odata/REST.

"Project Mustang is very exciting because it takes the ability to expose a common OData protocol over a variety of stores and allows customers to host that themselves."

Michael Pizzo
PRINCIPAL SOFTWARE ARCHITECT, MICROSOFT
What are the pieces of the Hybrid Data Pipeline?

**Clients**
- ODBC driver
- JDBC driver
- OData endpoint
  (nothing to embed... just a url)

**Hybrid Data Pipeline Server**
- One or more servers, Tomcat-based deployment, on 64-bit Intel Linux

**On-Premises Connector (optional)**
- Installs in or near the on-premises data sources to provide secure access from the cloud

**Built-In Connectors**
- All DataDirect JDBC drivers shipped in the box*
Engineered from the Award-Winning DataDirect Cloud Service

- Hybrid Data Pipeline is an evolution of the DataDirect Cloud service
- Released in 2014
- 10,000+ users in production
How Hybrid Data Pipeline works
Install Hybrid Data Pipeline
Configure On-Premises Connector
(Only required when accessing data behind a firewall)

“Firewall-friendly”, yet still secure connection without network reconfiguration or a requirement to open any ports
Configure the Data Source(s)

Data Stores
Available Data Stores

Beta Data Stores

HubSpot
Configure your data source(s)
Cloud-resident applications access data directly using SQL (ODBC or JDBC) or REST (Odata)

All data is encrypted in transit and data is not persisted in the pipeline
Supported Data Sources for Hybrid Data Pipeline 4.1

Apache Hadoop Hive 0.8.0 and higher*
Amazon EMR 2.1.4 and higher
Amazon Redshift
Apache Spark SQL 1.2, 1.3, 1.4, 1.5
Cloudera CDH update 4 and higher
Cloudera Impala 1.0, 1.1, 1.2, 1.3, 1.4
Cloudera Impala 2.0, 2.1, 2.2
Hortonworks Data Platform 1.3 and higher*
IBM BigInsights 3.0 and higher*
MapR 1.2 and higher
Pivotal HD 2.0.1 and higher
DB2 V9.1, V9.5, V9.7, 9.8 for Linux, UNIX, Windows DB2 V8.x for LUW
DB2 11 for z/OS* DB2 V10 for z/OS DB2 V9.1 for z/OS
DB2 UDB V8.1 for z/OS
DB2 i 7.1, 7.2* (DB2 UDB V7R1, V7R2 for iSeries)
DB2 i 6.1 (DB2 UDB V6R1 for iSeries)
DB2 for i5/OS (DB2 UDB V5R4 for iSeries)
Eloqua (Oracle Marketing Cloud)
Financial Force
Google Analytics
Greenplum 4, 4.1, 4.2, 4.3
Greenplum 3.3
Hubspot
Informix Dynamic Server 12.1*
Informix Dynamic Server 11.0, 11.5, 11.7
Informix Dynamic Server 10.0
Informix Dynamic Server 9.2, 9.3, 9.4
Informix Dynamic Server 11.0, 11.5, 11.7
Informix Dynamic Server 10.0
Informix Dynamic Server 9.2, 9.3, 9.4
Marketo

Microsoft SQL Server 2014*
Microsoft SQL Server 2012
Microsoft SQL Server 2008 R1, R2
Microsoft SQL Server 2005
Microsoft SQL Server 2000 Desktop Engine (MSDE 2000)
Microsoft SQL Server 2000
MySQL Community Edition
MySQL Enterprise Edition 5.0, 5.1, 5.5, 5.6*
Oracle 12c R1 (12.1)*
Oracle 11g R1, R2 (11.1, 11.2)
Oracle 10g R1, R2 (10.1, 10.2)
Oracle 9i R1, R2 (9.0.1, 9.2)
Oracle 8i R3 (8.1.7)
Oracle Service Cloud
Oracle Sales Cloud
Pivotal HAWQ 1.1*, 1.2*
PostgreSQL 9.0, 9.1, 9.2, 9.3, 9.4*
PostgreSQL 8.2, 8.3, 8.4
Progress OpenEdge 11.0, 11.1*, 11.2*, 11.3*, 11.4*
Progress OpenEdge 10.1.x, 10.2.x
Progress Rollbase 2.0 and higher*
SAP Adaptive Server Enterprise 16.0*
ServiceMax
SugarCRM 7.1.6 and higher*
Sybase Adaptive Server Enterprise 15.0, 15.5, 15.7
Sybase Adaptive Server Enterprise 12.0, 12.5, 12.5.x
Sybase Adaptive Server Enterprise 11.9
Veeva CRM

Blue text indicates cloud hosted
Blue text* indicates cloud hosted with on-premises option
Black text* indicates covered by Day One Support Policy
Planned Data Source Changes for 4.2 (Safe Harbor – subject to change)

**BRING YOUR OWN JDBC DRIVER!**

<table>
<thead>
<tr>
<th>HPE Vertica</th>
<th>IBM Netezza</th>
<th>Denodo</th>
</tr>
</thead>
<tbody>
<tr>
<td>HPE Nonstop</td>
<td>IBM BigSQL</td>
<td>Rocket DVS</td>
</tr>
<tr>
<td>Microsoft Parallel Data Warehouse</td>
<td>Microsoft Azure Data Warehouse</td>
<td>Google BigQuery</td>
</tr>
<tr>
<td>Teradata</td>
<td>SAP Sybase IQ</td>
<td>SAP Hana</td>
</tr>
<tr>
<td>SAS</td>
<td>EnterpriseDB</td>
<td>Others…</td>
</tr>
</tbody>
</table>
How customers can access on-premises data using the cloud version of Paxata's self-service data preparation platform.
Make Every Business an Information Inspired Business™
Today, IT wants to empower a data-driven organization
And there is no shortage of data
Data Analysts want data freedom

They have tools for every analytic use case
Enterprises spend up to 80% of time preparing data...
The Modern Architecture: Balancing Freedom with Responsibility

Collect and manage data

IT

The Business

Analyst teams get access to the data they need when they need it

... and IT maintains governance and enables access to raw data
• Multi-cloud, hybrid architectures are becoming the norm
• Connectivity to sources of data is a critical capability
# Top Cloud-Based Use Cases and Data Needs

<table>
<thead>
<tr>
<th>Loyalty Analytics/Customer 360</th>
<th>Supply Chain Analytics</th>
<th>Life Sciences Research</th>
</tr>
</thead>
<tbody>
<tr>
<td>• CRM</td>
<td>• IoT data</td>
<td>• Clinical Trial site data</td>
</tr>
<tr>
<td>• Customer Surveys</td>
<td>• 3&lt;sup&gt;rd&lt;/sup&gt;-party Supplier Data</td>
<td>• Public Data Services (PubMed, ClinicalTrials.gov)</td>
</tr>
<tr>
<td>• Transactional Data</td>
<td>• ERP data</td>
<td>• Genomic Data</td>
</tr>
<tr>
<td>• Loyalty Management Systems</td>
<td></td>
<td>• Proprietary Research Data</td>
</tr>
<tr>
<td>• Service Desk</td>
<td></td>
<td></td>
</tr>
<tr>
<td>• Marketo</td>
<td></td>
<td></td>
</tr>
<tr>
<td>• Google Analytics</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Self-service Connectivity and Data Preparation
Learn More about Progress DataDirect Hybrid Connectivity

https://prgress.co/hybrid

Firewall Friendly Connectivity

Around the web:
- Salesforce Developers guide to hybrid connectivity
- Progress joins Oracle for OData reveal on stage at OpenWorld
- Building Mobile applications with hybrid connectivity

Related Webinars:
- OData External Data Integration Strategies for SaaS [Guests: Microsoft and Oracle]
- Firewall Friendly Pipeline for Secure Data Access
- Webinar: Integrate Oracle Data Sources with Salesforce Connect [Guest: Appirio]

Single Interface to Cloud Data Sources

Around the web:
- First ever OData NLP Hackathon with Silicon Valley startups
- Future of SAP Business Objects Universe with hybrid connectivity
- How Oracle Data Integration (ODI) leverages hybrid connectivity
- Recommended integration solution for ODI and Oracle Marketing Cloud
- Microstrategy community introduction to hybrid connectivity
- SAP Lumira extension for hybrid connectivity

Related Webinars:
- How Big Data ISVs Get Marketing Data into Lakes [Guest: Ernst Young]
- Webinar: Binding Kendo UI to Any Data with OData [Guest: Chris Woodruff]
Questions?