

TIBCO Data Virtualization

BENEFITS

ECONOMICAL

Integrate data reliably at a fraction of physical warehousing and ETL time, cost and rigidity. Evolve rapidly when requirements change.

IMMEDIATE

Deliver up-to-the-minute data as needed, using advanced performance optimization algorithms and fine-grained security.

BUSINESS-FRIENDLY

Transform native IT structures and syntax into easy-to-understand, IT-curated data sets sharable via a self-service business directory.

WIDE-RANGING

Access data from distributed data sources including traditional enterprise, big data, cloud, and IoT. Use it across myriad analytics, self-service, business intelligence, and transactional applications.

ENTERPRISE GRADE

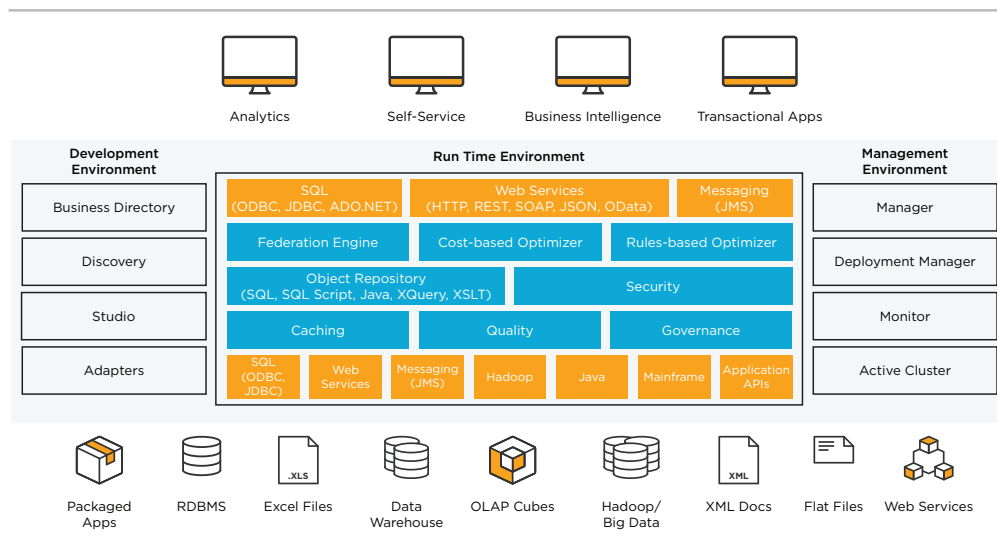
Support multiple lines of business, hundreds of projects, and thousands of users.

With data the new competitive battleground, businesses that take advantage of their data will be the leaders; those that do not will fall behind.

But gaining this advantage is a more difficult technical challenge than ever as your business requirements are ever-changing and your data is more-widely distributed across on-premises, big data, the Internet of Things (IoT) and the Cloud. Traditional physical data integration via data warehousing and ETL is often too time consuming, too rigid, and too costly to support your dynamic business needs. There has to be a better way. There is, and it's called data virtualization.

TIBCO® Data Virtualization software lets you integrate data with breakthrough speed and cost effectiveness. With it, you can build and manage virtualized views and data services that access, transform, and deliver the data your business requires to accelerate revenue, reduce costs and risk, improve compliance, and more.

TIBCO Data Virtualization is Java enterprise-grade middleware with a modular structure that supports all phases of data virtualization development, run-time, and management.



TIBCO DATA VIRTUALIZATION MODULES

Business Directory is a self-service data directory you can use to easily search, categorize, and consume IT-curated data sets developed using TIBCO Data Virtualization. Easy data set sharing and reuse helps accelerate good business outcomes while reducing IT workloads.

Discovery lets you go beyond simple data profiling to examine data, locate important entities, and reveal hidden relationships across distinct data sources. You can quickly build and display comprehensive entity-relationship (E-R) diagrams and data models so you can meet new business requirements faster and more easily.

COMPLETE DATA VIRTUALIZATION SOLUTION

Beyond TIBCO Data Virtualization software, TIBCO also provides a broad array of complementary advanced services, training, technical support, customer advisory program, knowledgebase, and partner offerings, so you get the complete solution needed to ensure your data virtualization success.

Studio is the agile modeling, development, and resource management tool that data-oriented developers use to model data, design and view data services, build transformations, optimize queries, manage resources, and more. Easy to learn and use, Studio's graphical modeling environment provides a flexible workspace where queries are created and tested, as well as a data services repository where completed objects are published. Studio also offers a rich set of transformations in addition to an easy-to-use transformation editor. Five languages complement Studio's graphical modeling capabilities: SQL, SQL Script, Java, XQuery, and XSLT.

Adapters provide a wide range of data source connectivity for databases, files, big data, cloud sources, packaged applications, and more. Beyond schema-to-schema only connectivity, TIBCO Data Virtualization adapters integrate with data source optimizers to ensure more accurate queries and higher performance. The Data Source Tool Kit allows you to build additional custom adapters.

Cost-based and Rules-based Optimizers used by the **Federation Engine** (both a massively parallel processing [MPP] execution engine or the classic, pushdown query execution engine) work directly with data source optimizers to maximize query performance. The **Objects Repository** lets you manage your data services throughout their lifecycle. Myriad fine-grained **Security** capabilities, including authentication, authorization, and encryption, safeguard that you deliver sanctioned data only, no more, no less.

Granular workload management allows you to provide more intelligent allocation of resources for important workloads. Control memory usage, request length, row counts, and more, as well as avoid potentially problematic requests. These controls can be implemented at the object, user group, server, and/or cluster group level.

Flexible **Caching** options enable higher performance and greater uptime.

Quality helps ensure you deliver the best data possible. And built-in **Governance** features provide complete visibility, traceability and control.

Manager is the administrative console used to set up user IDs, passwords, and security profiles, as well as view logs, check server activity, and more.

Deployment Manager lets you quickly and easily migrate entire projects in a single step, including their resources, cache settings, server configurations, security profiles, and more across instances to simplify and automate your development lifecycle.

Monitor provides a comprehensive, real-time view of your TIBCO Data Virtualization cluster. Monitor displays all the pertinent system health indicators required to help your IT operations staff guide corrective actions.

Active Cluster works in conjunction with load balancers to provide high availability and greater scale to meet your challenging service level agreements. Active Cluster simplifies complex operations management by automatically sharing resources, adjusting capacity on demand, and more.

DEPLOYMENT OPTIONS

TIBCO provides multiple options for deploying TIBCO Data Virtualization. You can install and run it on-premises, in your private cloud environment, or at a public cloud provider such as Amazon AWS, Google Cloud Platform, and Microsoft Azure. For AWS, TIBCO also provides TIBCO Data Virtualization on the AWS Marketplace to simplify and accelerate deployment.

KEY FEATURES

DEVELOPMENT ENVIRONMENT

Tables 1 through 4 summarize key TIBCO Data Virtualization development capabilities.

Table 1. Business Directory: Access and use data sets via an intuitive, self-service browser

FEATURE	DESCRIPTION
Search options	Find data sets using simple Google-like searches as well as advanced parameterized searches.
Browse	Find data sets using attributes such as data type, format, category, and more.
Data preview	Preview specific data included within data sets.
Security	See only the data sets you're supposed to see based on your TIBCO Data Virtualization enforced security profile.
Consumption	Use your favorite analytic/BI tools to consume data sets using Business Directory access information.
Categorization	Organize large and/or diverse data sets into user-defined or system categories.

FEATURE	DESCRIPTION
Collaboration	Allow business users and IT to interact via comments to improve data quality and utilization.
Business metadata extensions	Add custom definitions, properties, links, and status codes to enrich IT metadata with business metadata that your business users understand.
Business metadata display options	Present business metadata in a shared area, an existing properties tab, or a new custom tab.
Personalization	Receive email updates of changes and comments for data you care about.
Provisioning	Register additional TIBCO Data Virtualization instances to expand your data sets.
REST API	Access Business Directory programmatically to expose data sets to additional consuming applications.

Table 2. Modeling and Transformation: Design reusable views and data services within an agile, high-productivity development environment

FEATURE	DESCRIPTION
Graphical development environment	Graphically model data, design view and data services, build transformations, optimize queries, manage resources, and more.
Agile development methodology	Develop and refine views and data services iteratively.
Introspection	Automatically probe physical data sources and select desired resources. Optionally inspect data sources interactively.
Data discovery	Reveal data relationships across disparate entities using formal keys and fuzzy matching.
Bottom-up modeling	Design views and data services by combining data from disparate systems without worrying about underlying source access and format complexity.
Flexible modeling and transformation methods	Create views and data services graphically or via scripting languages as appropriate. SQL, SQL script, XQuery, XSLT, and Java functions.
SQL	Use familiar SQL standards.
SQL script	Implement stored procedures using a familiar scripting language.
XQuery	Create complex XML structures using a graphical XQuery editor.
XPath transformation	Establish arbitrary complex mapping of XML schema elements to XML output.
JSON querying and transformation	Query and transform JSON data from Web services into a relational format.
Analytic functions	Use a full set of analytic functions such as CORR, COUNT, NTILE, STDDEV, and VARIANCE.
Contract first design	Build data services using preexisting WSDLs and schemas.
Contract last design	Define the Java wrapper first and then develop WSDL.
Query plan	View query execution plan steps and details.
Views dependency graph	Graphically display dependencies between data sources and views and data services.
Physical table creation	Create and drop physical tables within a designated data source.
Localized UI	Chinese and Japanese language support.

Table 3. Metadata Repository: Store and manage relevant metadata

FEATURE	DESCRIPTION
Complete repository	Manage resources such as data sources, views and data services, and procedures throughout their lifecycles.
Public metadata API	Deploy a web services metadata API for easy access and sharing.
Schema change notification	Receive notice when data source schemas change.
Source metadata	Access the metadata of the physical data source.
Open API	Open access to TIBCO Data Virtualization system libraries to enable custom scripting and orchestration.

Table 4. Version Control: Control the view and data service development lifecycle

FEATURE	DESCRIPTION
Migration GUI	Add project folders directly to version control systems. Check-in and check-out folders or individual resources to track changes. Rollback revisions of folders or resources to prior versions.
Resource locking	Protect against inadvertent modifications and overwrites.
Change history	Track changes made by users with annotations.
Source control and transport	Manage artifacts and transports from development through production.
Version control systems	Integrate directly with Apache® Subversion® and Git.

RUN-TIME ENVIRONMENT**Table 5. Federated Query Engine: Run optimized queries across data sources**

FEATURE	DESCRIPTION
Federation engine	Join and aggregate data that is vertically and horizontally partitioned.
Massively Parallel Processing (MPP) Engine	Enable multi-source, big data scale virtualization workloads that dynamically distribute the work across multiple processors
Data source	Leverage data source optimizers to ensure query accuracy and maximize query performance.
Cost-based optimizer	Use statistics to create an optimal query plan that reduces unnecessary data flow across the network.
Rule-based optimizer	Allow users to specify exactly how they want to run a particular query.

FEATURE	DESCRIPTION
Granular workload management	Control memory usage, request length, row counts, and more, as well as avoid potentially problematic requests. These controls can be implemented at the object, user group, server, and/or cluster group level.
Scheduling	Run queries based on set times.
Alert triggers	Implement resource, event, and user-defined triggers. Use a published API to handle custom Java alerts.

Table 6. Performance Optimization Algorithms and Techniques: Optimize query performance

FEATURE	DESCRIPTION
Complete set of join algorithms	Automatically rewrite the query to use the most efficient join strategy (for example, hash join, sort-merge join, distributed semi-join, data-ship join, union-join flip, nested-loop join, and others).
Single-source join grouping	Run data-reducing joins at the data source rather than bringing the data across the network.
Predicate push-down	Push WHERE clause predicates down into the underlying data source to reduce data at the source.
Full and partial aggregate push down	Push aggregate functions down to source when applicable.
Serialization or parallelization of join operators	Determine the proper join order and join algorithms based on estimated cardinality and join results derived from data distribution histograms.
Projection pruning	Eliminate all unnecessary columns from fetch nodes in a query tree.
Constraint propagation	Distribute filters to multiple branches of the query plan, allowing data reduction by a single filter to potentially occur in multiple data sources.
Scan multiplexing	Reuse data sets that appear in multiple places in a single query plan.
Empty scan detection	Detect logical conditions that would produce empty data sets, and then eliminate those parts of the query plan prior to processing.
Redundant operator cropping	Eliminate redundant or extraneous operators within a complex multiple-operator query.
Blocking operator prefetching	Proactively run parts of the query plan that must finish before other parts of the query plan can continue, thereby increasing the overall responsiveness of the query.

FEATURE	DESCRIPTION
Connection pool sharing	Share access to data sources to avoid bottlenecks.
Results streaming	Stream data to consuming applications as results are processed at the underlying sources.
Hybrid memory and disk use	Balance memory and disk use for optimal performance.
Native XML support	Support XML internally for fast parsing and joins.
API	Expose the query execution plan via JDBC/ODBC.

Table 7. Caching: Move data to predesignated caches to boost availability and performance

FEATURE	DESCRIPTION
Event-based refresh	Refresh caches based on defined business rules.
Scheduled refresh	Refresh caches at set times.
Manual refresh	Refresh caches on demand as needed.
Incremental refresh	Refresh part of a cache.
Full refresh	Refresh entire cache.
Native bulk extract	Use native bulk EXTRACT functions in the source to extract data more efficiently than using SELECTs.
Native bulk load	Use native LOAD functions in destination-to-load and refresh caches more efficiently than using INSERTs.
Parallel load	Use multiple threads to load caches in parallel.
Centralized cache tracking	Centrally track caching tables distributed across multiple data sources.
Multi-table caching	Avoid contention on cache refreshes, accelerate refresh speeds, and maintain non-stop availability using multiple tables per cache view.

Table 8. Data Access: Connect and use data from distributed sources

FEATURE	DESCRIPTION
Databases	Access popular databases via Open Database Connectivity (ODBC) and Java Database Connectivity (JDBC).
Big data	Access Apache Hadoop® through Apache Hive™, Apache Impala®, and Apache HBase®.
Multidimensional data	Access multidimensional data in SAP BW.
NoSQL and cloud databases	Access NoSQL and cloud database sources such as Amazon DynamoDB, Amazon Redshift, Cassandra, and MongoDB.
Web services	Access SOAP over HTTP, XML over HTTP, and Java Message Service (JMS) services. A message pipeline allows interjection of custom logic during the web service request and response.
Packaged applications	Access SAP, Oracle E-Business Suite, Salesforce.com, and other applications using standard objects such as invoices, shipments, orders, customers, opportunities, and more.
SaaS applications	Access SaaS applications including Google Analytics, Microsoft Dynamics, and NetSuite.
Collaboration	Access collaboration apps such as email, Google Sheets, and Microsoft SharePoint.
Social media	Access social media sources such as Facebook, LinkedIn, RSS, and Twitter.
Java API	Access non-relational sources using custom procedures.
Data source tool kit	Access a set of libraries of services that can be imported into your preferred integrated development kit to facilitate and accelerate data adapter creation. Services include database mapping, data type mapping, syntax mapping, and function mapping. Together these services minimize custom code development.

Table 9. Data Delivery: Deliver requested data to myriad consuming applications

FEATURE	DESCRIPTION
Database objects	Publish views for consumption through ODBC, JDBC, and ADO.NET.
Web services	Publish data services in the form of WSDL for consumption using SOAP or SOAP over JMS. A message pipeline allows interjection of custom logic during the web service request and response.
Representational State Transfer (REST)	Publish data services in the REST format. REST CREATE, READ, UPDATE, and DELETE functions are supported.
Open Data (OData) protocol	Publish data services in the OData format.

FEATURE	DESCRIPTION
JSON	Publish JSON, including more formatting capabilities for XML-to-JSON translation, for both procedures and table outputs.
BI tool integration	Create temporary tables in BI tools to store filters for visualizations or reports.
DDL support for client applications	Support client applications that create tables using DDL statements via JDBC/ODBC/ADO. CREATE TABLE, CREATE TABLE AS SELECT, and CREATE TEMP TABLE syntax. Multiple physical data sources mappings are supported for each published database to avoid conflicts between users when creating tables.

Table 10. Security: Protect sensitive data with fine-grained, multi-level security

FEATURE	DESCRIPTION
Policy-based security	Apply authentication, authorization, and encryption rules via policies.
Single sign-on	Sign on once to access all integrated data sources and consuming applications.
Row-level authentication	Control access to specific rows via granular permissions.
Column-level authentication	Control access to specific columns via granular permissions.
Authentication types	Support basic, pass-through, Kerberos, SAML, and NTLM.
Column masking	Implement column masking rules to hide, replace, or obfuscate portions of a column's value depending on a user's level of access.
SSL over HTTP with support for mutual authentication	Mutually authenticate published services, web services data sources, and Oracle databases. Certificate-based authentication and Web Services Security (WSS) authentication are supported.
Pass-through	Use an existing user ID and password and pass through to TIBCO Data Virtualization for authentication.
Lightweight Directory Access Protocol (LDAP)	Use security profiles from LDAP to authenticate user access to protected data sources.
Pluggable authentication module	Use third-party systems for authentication.
Encryption	Encrypt passwords and data in motion via TLS.
Access management	Use TIBCO Data Virtualization as the system of record for security roles and profiles.

Table 11. Governance: Provide complete visibility, traceability and control

FEATURE	DESCRIPTION
Data lineage	Trace lineage from multiple data sources to a single data consumer.
Column lineage API	Access column lineage via an API.
Where-used	Trace where-used from a single data source to multiple data consumers.
Logging	Track system and/or user activity.
Standards enforcement	Implement internal and industry data standards.
Open API	Open access to TIBCO Data Virtualization system libraries to enable custom scripting and orchestration.

Table 12. Data Quality: Help ensure correct and complete data

FEATURE	DESCRIPTION
Standardization and conformation	Create views and data services that conform to agreed standards.
Enrichment and augmentation	Extend views and data services with additional data.
Validation	Validate data sets with users prior to publishing views and data services.
Object reuse	Share views and data services to ensure consistent data definitions.

MANAGEMENT ENVIRONMENT

Tables 13 through 15 summarize key TIBCO Data Virtualization management capabilities.

Table 13. Management: Administer and manage for efficient operations

FEATURE	DESCRIPTION
Management console options	Access the management console through TIBCO Data Virtualization Studio or a web browser.
User setup	Set up user and group profiles.
Security	Enable multiple forms of security to increase data protection.
Scheduling	Schedule TIBCO Data Virtualization activities flexibly.

FEATURE	DESCRIPTION
Deployments	Manage tasks related to development, management, configuration, and versioning.
Real-time system indicators	Monitor critical system metrics and tune for optimal performance including memory usage, query plans for currently running and past requests, data sources, and caches.
Usage metrics	Deliver usage activity detail to your reporting tool of choice via an open API.
Simple Network Management Protocol	Allow monitoring by third-party systems via SNMP API.

Table 14. Active Cluster: Cluster servers to meet reliability, availability and scalability SLAs

FEATURE	DESCRIPTION
Flexible cluster deployment	Expand capacity and improve availability simply by adding new nodes to an existing cluster or adding new clusters.
Active-active clustering	Distribute TIBCO Data Virtualization workloads across cluster nodes in conjunction with load balancers.
Shared cluster cache	Improve overall cluster performance by coalescing redundant data source hits and reducing data latency.
Replicated metadata repository	Replicate metadata across clusters to simplify management.
Restore cluster nodes	Restore cluster nodes using Web Manager, cluster_util script or API.

Table 15. Deployment Manager: Automate migration or promotion of artifacts, configurations, and settings

FEATURE	DESCRIPTION
Resource migration	Migrate or promote (create/update/delete) artifacts from one TIBCO Data Virtualization instance to another.
Cache setting migration	Migrate or promote cache table names, caching methods, refresh method, and cache policies and schedules.
Server configuration migration	Replicate server configurations, for example, enabling and disabling triggers.
User/group migration	Migrate or promote user and group IDs, security profiles, and other user and group information.

TECHNICAL SPECIFICATIONS

Tables 16 through 18 summarize TIBCO Data Virtualization technical specifications and standards support.

Table 16. Supported Platforms: Leverage standard operating systems and high performance servers

TIBCO DATA VIRTUALIZATION STUDIO CLIENTS	SERVER	JVM
<ul style="list-style-type: none"> • Microsoft Windows 	<ul style="list-style-type: none"> • CentOS • Cisco Unified Computing System • Oracle Linux and Red Hat compatibility mode • Red Hat Enterprise Linux AS • SUSE Enterprise Linux • Microsoft Windows • Solaris 	<ul style="list-style-type: none"> • 64-bit

Table 17. Data Source Adapters: Simplify data access while ensuring query accuracy and performance

STANDARD DATA SOURCE ADAPTERS	
<ul style="list-style-type: none"> • Apache Drill • Apache Hive • Apache Impala • Apache Spark • Custom Java procedure • Cloudera CDH • CDH Hive • CDH Impala • Data direct mainframe • Elasticsearch • Files (cache, delimited, and XML) • Greenplum • Hbase • Hortonworks HDP • HSQLDB • IBM DB2 • IBM DB2 z/OS • Informix • LDAP • Microsoft Access • Microsoft Excel 	<ul style="list-style-type: none"> • Microsoft SQL Server • MySQL • Netezza NPS • Oracle • Odata • Parstream • PostgreSQL • SAP HANA • Sybase • Sybase IQ • Teradata • TIBCO Data Virtualization • Vertica • WSDL • XML (flat files over HTTP)

GENERAL PURPOSE ODBC AND JDBC DATA SOURCE ADAPTERS

- Custom JDBC adapter
- JDBC-ODBC Bridge

CRM AND ERP ADAPTERS

- Microsoft Dynamics CRM (on-premises and online)
- Microsoft Dynamics GP
- Microsoft Dynamics NAV
- NetSuite CRM
- NetSuite ERP
- Oracle EBS
- Salesforce.com
- Siebel

DATA SOURCE TOOL KIT

- Software Development Kit for Data Source Adapter Development

MARKETING AUTOMATION ADAPTERS

- Google Adwords
- Google Analytics
- HubSpot
- Marketo
- Oracle Eloqua

COLLABORATION ADAPTERS

- Email
- Google Apps
- Google Sheets
- Microsoft Active Directory
- Microsoft SharePoint (on-premises and online)
- Microsoft SharePoint Excel services

SAP ADAPTERS

- SAP Netweaver
- mySAP Business Suite
- SAP BW
- SAP Business Explorer (BEx)

NOSQL AND CLOUD DATABASE ADAPTERS

- Amazon DynamoDB
- Amazon RedShift
- Cassandra
- Couchbase
- Google BigQuery
- HBase
- MongoDB

SOCIAL MEDIA ADAPTERS

- Facebook
- LinkedIn
- RSS
- Twitter

Table 18. Additional Technical Specifications: Take advantage of industry standards

DELIVERY INTERFACES	SECURITY PROTOCOLS	CACHE REPOSITORIES
<ul style="list-style-type: none"> • ADO.NET • ODBC • Hadoop • JDBC • REST • SOAP • SOAP and JMS TIBCO Enterprise Message Service™ • SOAP and JMS Sonic MQ 	<ul style="list-style-type: none"> • Base64 • Kerberos • NTLM • SAML • SSL • WS-Security 	<ul style="list-style-type: none"> • File • Greenplum • HSQLDB • IBM DB2 • Microsoft SQL Server • MySQL • Netezza • Oracle • PostgreSQL • SAP HANA • Sybase ASE • Sybase IQ • Teradata • Vertica
ENTERPRISE SERVICE BUSES	DIRECTORY SERVICES	DATA SHIP SOURCES AND TARGETS
<ul style="list-style-type: none"> • OpenMQ • Sonic MQ • Enterprise Message Service™ 	<ul style="list-style-type: none"> • Active Directory • Oracle Directory Server Enterprise Edition • Novell eDirectory • TIBCO Data Virtualization 	<ul style="list-style-type: none"> • IBM DB2 • Microsoft SQL Server • Netezza • Oracle • PostgreSQL • Sybase IQ • Teradata • Vertica
WEB SERVICES PROTOCOLS	SOFTWARE DEVELOPMENT STANDARDS	
<ul style="list-style-type: none"> • .NET • OData • REST and JSON • SOAP • WSDL • WSI • XPath • XQuery • XSLT • XML (flat files or over HTTP) 	<ul style="list-style-type: none"> • SQL 92 and 99 • Unicode • JDK • J2EE • JNDI 	



Global Headquarters
3307 Hillview Avenue
Palo Alto, CA 94304
+1 650-846-1000 TEL
+1 800-420-8450
+1 650-846-1005 FAX
www.tibco.com

TIBCO fuels digital business by enabling better decisions and faster, smarter actions through the TIBCO Connected Intelligence Cloud. From APIs and systems to devices and people, we interconnect everything, capture data in real time wherever it is, and augment the intelligence of your business through analytical insights. Thousands of customers around the globe rely on us to build compelling experiences, energize operations, and propel innovation. Learn how TIBCO makes digital smarter at www.tibco.com.

©2017-2018, TIBCO Software Inc. All rights reserved. TIBCO, the TIBCO logo, and Enterprise Message Service are trademarks or registered trademarks of TIBCO Software Inc. or its subsidiaries in the United States and/or other countries. Apache Drill, Hadoop, HBase, Hive, Impala, and Subversion are trademarks of The Apache Software Foundation in the United States and/or other countries. All other product and company names and marks in this document are the property of their respective owners and mentioned for identification purposes only.

11/14/18