Tuesday May 16, 2017 04:40 PM - 05:20 PM

What’s New in PowerCenter, Modern Data Integration & Big Data Solutions

Awez Syed  
VP Product Management

John Haddad  
Sr. Director Product & Technical Marketing
Safe Harbor

The information being provided today is for informational purposes only. The development, release and timing of any Informatica product or functionality described today remain at the sole discretion of Informatica and should not be relied upon in making a purchasing decision. Statements made today are based on currently available information, which is subject to change. Such statements should not be relied upon as a representation, warranty or commitment to deliver specific products or functionality in the future.
300,000+ Informatica Developers

ETL Developer → Data Engineer
Data Engineers

Self-service and agile development
- Get started quickly and rapidly deploy Informatica
- Self-service integration of multiple systems without the hairball
- Manage and monitor resources, capacity, and execution on-premises & Cloud

Big data (volume, variety, velocity) processing
- Integrate and process all types of data at scale
- Process Big Data in the Cloud
- Seamlessly move from batch processing to real-time streaming

Hybrid deployments (Cloud & on-premises)
- Integrate all types of data in a hybrid world
- Deploy all types of integration patterns with Cloud services
- Integrate industry standard & non-standard partner data faster
Self-service and agile development
1-Click Deployments on AWS and Azure

Get started quickly and rapidly deploy Informatica

- **Amazon AWS**
  - Certification for PowerCenter and Big Data Management and EIC
  - Connectivity to Amazon Redshift, Amazon RDS, S3

- **Microsoft Azure**
  - Certification for PowerCenter, Big Data Management and EIC
  - PC Azure SQL Data Warehouse connector, PC Azure Blob connector
PowerCenter Deployment Automation
Get started quickly and rapidly deploy Informatica

- Deploy PowerCenter using Docker
- Create your own Docker Containers
- Comply with corporate deployment standards
- Spin up new PowerCenter instances in a fraction of the time

Deployment Automation
- pmrep Create Query helps automate code deployment via command line
- pmrep Create Connection helps automate environment setup steps
- Audit PowerCenter metadata XML import during code deployment
1-Step BDM Hadoop Installation w/Zero Footprint
Get started quickly and rapidly deploy Informatica

- 1-step install
  - Cluster Configuration Object is 1st class repository object
  - Integration with Cloudera Manager, Ambari
  - Easy to override configuration and 1-click connection creation & refresh for HDFS, Hive, HBase & Hadoop Engine

- Zero footprint
  - No design-time footprint, No more RPMS / Parcels / Stacks, self managed, self contained application execution
  - Relies on Hadoop distros cache for dependency distribution
  - Minimal job overhead for the 1st job execution for dependency resolution
Big Data Readiness Assessment for PowerCenter
Get started quickly and rapidly deploy Informatica

• Interactive report to assess reuse of PowerCenter logic for Big Data
• Drill down by folder, engine, mapping, etc
• Import mappings with multiple pipelines
• Import PowerCenter Workflows
• Import Sessions: Pre-SQL, Post-SQL, Source qualifier overrides
Modern Integration Hub Pub/Sub Architecture
Self-service integration of multiple systems without the hairball

IT creates publications for all data sources

Centralized data processing for consistency

Governed Data Management + Self-service = Agility, Consistency and Efficiency
Data Integration Hub Self-Service and Collaboration
Self-service integration of multiple systems without the hairball

- Easily manage integration flow
- Visual view of Integration Hub entities and their relationship
- Show important indicators: error events, invalid/incomplete, pubs/subs
- Drill-down into entities from overview diagram
- Wizards for auto-generation of pubs/subs by less technical users
Operational Insight
Manage and monitor resources, capacity, and execution on-premises & Cloud

- Monitoring & Asset Exploration
  - Bird’s eye view of all PowerCenter deployments from one UI
- Insightful and predictive analytics
- Recommendations based on logs, system performance and statistics
Big data (volume, variety, velocity) processing
Enhanced Support for Blaze and Spark
Integrate and process all types of data at scale

All mapping logic can run in Blaze

Hierarchical & stateful variables in Spark

New transforms & SQOOP in Spark
Big Data Management in the Cloud
Process Big Data in the Cloud

Certified for Cloud eco-system

- Amazon and Azure Connectivity
  - Support for S3 and RedShift in Blaze & Spark modes
  - S3: Ability to use distcp (bulk load), Ability to read and write entire directories
  - RedShift: Ability to save the staging files for reprocessing, audits, etc
  - Support for Azure DW and Blob
  - Support for HDFS and Hive on ADLS and WASB
Informatica Intelligent Streaming
Seamlessly move from batch processing to real-time streaming

- First to support AWS Kinesis Firehose streaming
- First Change Data Capture (CDC) integration
- Modern Data Type Support; XML, JSON, Avro
Hybrid deployments (on-premises and cloud)
PowerExchange for Cloud Applications
Integrate all types of data in a hybrid world

Connect to Informatica Cloud from PowerCenter

Access connections defined under Informatica Cloud Account from PowerCenter

Import any object and create mappings in PowerCenter
Cloud Data Integration Services
Deploy all types of integration patterns with Cloud services

- Cloud Data Warehouse: Data Ingestion
- Replicate bulk data to Cloud
- Synchronize SaaS and On-Premises Application Data

- Mapping Designer for developers
- Wizards for citizens integrators
- Mapping template and parameterization

- PowerCenter interoperability and reuse
- Large library of expressions, transformation and lookup
- Agent Group for HA and Scaling
- Blue-Green high availability upgrades
Cloud Application Integration Services

Deploy all types of integration patterns with Cloud services

- Integrate your apps, your data, and your processes in real time
- Expose your APIs to your applications and partners
- Automate your business processes and workflows

Cloud and On-Premises Processing
- REST, JSON Services, SOAP and OData API Providers
- API Gateway Services
- API and Service Orchestration

Data, API and Service Connectivity
- Cloud and On-Premises Messaging
- Built-in Data Services
- File Content Listeners
- Process Automation and Guided Workflow
Hierarchical Enhancements for Broader Connectivity

Deploy all types of integration patterns with Cloud services

Easily map complex hierarchical data structures using:

• Hierarchy Builder & Parser to Read/Write XML/JSON and convert to relational outputs.

• Support for REST/SOAP web-services with simplified field mapping
B2B Industry Accelerators and Partner Portal
Integrate industry standard & non-standard partner data faster

- Accelerators for EDI, HIPAA, GDSN
- Prebuilt workflows, mappings and dashboards
- Validation and error reporting
- All relevant document types for business processes
- Ability to customize or extend
- Easily manage file exchange using secure self-service Partner Portal
  - Self Provision
  - Monitoring
  - Ad-Hoc Upload/Download
Using AI with B2B Gateway
Integrate industry standard & non-standard partner data faster

- Expedite and simplify on-boarding partners that use custom files
- Run structure discovery on sample files
  - Structure is auto-discovered
  - Tune and adjust discovered structures: Remove sections, rename fields, combine fields and more
- Use the identified structure to process incoming files

CLD 218: Extend cloud integration to trading partners with Cloud B2B Gateway for EDI
Data analysts spend 80% of their time finding & preparing the data”
Data Analysts and Data Scientists

Access the best data for analytics
• Google for enterprise data, know what you can trust
• Find and discover entities in both structured and unstructured data

Quickly mash up data sets for analysis without waiting for IT
• Collaborative self-service, machine learning guided data prep with governance
• Quickly prepare data with on-the-fly data wrangling and cleansing

Validate data is ready for analysis
• Leverage unified tools to quickly test, validate, & provision analytic models
DI&BD303: How to use Enterprise Information Catalog for Analytics and data governance

Unified Metadata

- Databases
- Data Warehouses
- Applications
- BI Tools
- Cloud
- Big Data

Google for Enterprise Data
- Discovery
- Semantic Search
- Relationships
- Proﬁling
- Data Similarity
- Collaboration
- Recommendations
- Annotations
- Smart Tags
- Alerts
- Lineage
- Impact Analysis
- Classifications
- Smart Tags
- Impact Analysis

Enterprise Information Catalog
Domain Discovery from Unstructured Files
Find and discover entities in both structured and unstructured data

- Identify and Classify entities from both structured and unstructured data
- New Unstructured Sources Supported: Excel, Word, PDF, Text Files
- Support for both rule based and learning based data domains
Automatically Combine Data Domains to Detect Entities
Find and discover entities in both structured and unstructured data
Intelligent Data Lake
Collaborative self-service, machine learning guided data prep with governance

DI&BD302: Self-service Big Data Discovery and Preparation made easy with Intelligent Data Lake.
Intelligent Data Lake DEMO
Data Quality Rules as Part of Data Prep

Quickly prepare data with on-the-fly data wrangling and cleansing

- Promote reusability, improve productivity and consistency by using pre-defined rules
- Use 1000s of pre-defined DQ rules as part of interactive data preparation
- Execute data quality rules at Big Data scale
- Rules can be defined in Informatica Developer Tool and Analyst Tool
Integration with Data Visualization Tools
Leverage unified tools to quickly test, validate, & provision analytic models

- Create Apache Zeppelin Notebook with charts and visualization to validate data sets for analysis
- Share interesting charts to help assess data for other users
- Automatically generates smart visualization recommendations
Data Integration & Big Data Sessions

**Monday**

Modern Data Integration and Big Data Industry Perspective
3:30-5:00 Room 2002-2006

**Tuesday**

Future Ready Data Integration - Optimized, Scalable, Agile and Smart
03:40-04:20 Room 2008

Delivering Data Fast – Asurion’s Journey with Data Lake Mgmt on AWS
04:40-05:20 Room 2001

What’s New With PowerCenter and Big Data Solutions
04:40-05:20 Room 2002

**Wednesday**

Nissan’s “Data Alchemy” and What’s New in Hubs
10:40-11:20 Room 2002

Applying Artificial Intelligence in the Very Real World
10:40-11:20 Room 2008

Journey to Highly Scalable B2B with Axalta and Dentaquest
11:40-12:20 Room 2002

Big Data, Bigger Performance: Tune BDM for better performance
11:40-12:20 Room 2006

ConocoPhilip’s Data Agile Story with Enterprise Information Catalog
02:20-03:00 Room 2002

Anthem’s Journey Driving Digital and Consumer Centricity with PowerCenter
03:20-04:00 Room 2002

**Thursday**

Architecture Day
09:00-04:00 Room 2003

Visit our large selection of Roundtables and Deep Dives happening on the 3rd floor throughout the conference!
## Data Integration & Big Data Deep Dives

<table>
<thead>
<tr>
<th>Course Code</th>
<th>Course Title</th>
</tr>
</thead>
<tbody>
<tr>
<td>DI&amp;BD302</td>
<td>Self-service Big Data Discovery and Preparation made easy with Intelligent Data Lake.</td>
</tr>
<tr>
<td>DI&amp;BD303</td>
<td>How to use Enterprise Information Catalog for Analytics and data governance</td>
</tr>
<tr>
<td>DI&amp;BD304</td>
<td>Big Data Management on cloud</td>
</tr>
<tr>
<td>DI&amp;BD305</td>
<td>How to use Spark 2.x for Big data batch and real time processing</td>
</tr>
</tbody>
</table>
26 Data Integration & Big Data Roundtables

- DI&BD201: PowerCenter Version 10 Upgrade
- DI&BD205: Big Data Management version 10- Upgrade & Performance Best Practices
- DI&BD208: Troubleshooting Tools, Informatica Support network and Support services
- DI&BD209: Next Generation Analytics: Streaming, Self-Service and Big Data
- DI&BD210: Deciding Which Informatica Product To Develop/Deploy with
- DI&BD229: Modern hybrid Data Integration Hub to boost agility with self-service
- DI&BD230: Handling Complex Data Types with Informatica Data Transformation
- DI&BD231: Informatica PowerCenter and Cloud - Performance, Scalability, Tips & Tricks
- DI&BD232: PowerCenter on Cloud (Amazon Web Services and Microsoft Azure)
- DI&BD233: Reusing PowerCenter applications for Big Data
- DI&BD234: Building scalable solutions using Informatica Blaze and Spark
- DI&BD235: Rapid deployment of Big Data Management in the Cloud
- DI&BD236: What's new in Big Data Management
- DI&BD237: "Making a Data Lake Intelligent: The Journey to Analytical Insight"
- DI&BD238: Exploring from Metadata Manager towards Gen-Next of Metadata solutions
- DI&BD239: Informatica PowerCenter 10.2 - New Features
- DI&BD240: Introduction to Intelligent Operational Insights for PowerCenter and beyond
- DI&BD241: Manage your partner community with B2B Data Exchange
- DI&BD242: Up-to-the-Minute Data Integration with Change Data Capture
- DI&BD243: Auto-scale PowerCenter domain and Pay-As-You-Go Subscription Licensing
- DI&BD244: Moving from Reference Architectures to Solution Architectures
- DI&BD245: Data Catalog for Self Service Analytics & Data Governance
- DI&BD246: Getting the Most out of Catalog and Data Lake - Best Practices
- DI&BD247: Self-service big data preparation with Intelligent Data Lake
- DI&BD248: Processing real-time data with Informatica Intelligent Streaming
- DI&BD249: Performance Best practices for Big Data Management & Intelligent Streaming
Review us on Peer Insights – Gartner’s Yelp for Enterprise Technology

- Peer reviews by managers for managers
- End-user IT or business users can leave reviews
- Reviewers are anonymous on publicly displayed reviews
- Validated by Gartner
- Non-Gartner clients *can* leave a review
- Takes ~15 minutes to complete
- Can start a review and come back to finish

- Kiosk on Moscone West 3rd Level Hallway to rooms 3014 – 3024 (see photo)
- Online [https://www.gartner.com/review/home](https://www.gartner.com/review/home)
- Search for *Informatica* to see products available to review – almost our entire portfolio is listed
Thank You