

Oracle Big Data Discovery (BDD) Hadoop Visualization

ORACLE Big Data Discovery

Kshitij Kumar
Wilfrido Solano



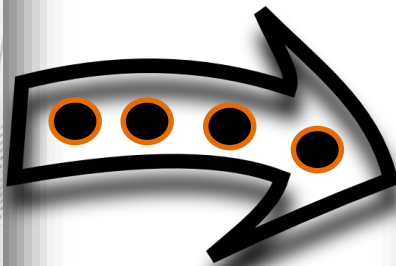
Agenda

- **Evolution of Data**
 - Product Position
 - Data Lakes
- **Big Data Discovery**
 - Technical Overview
 - Product Walkthrough
- **Closing**
 - Resources

This presentation contains confidential and proprietary information about Apps Associates' services, products, offerings and sales plays. Distribution of this content in any digital or printed manner is not allowed without Apps Associates' written approval.



We are back where we started (sort of)





Evolution of Computing

1980 - 1990



1. Departmental applications
2. Departmental reporting
3. Minimal integration
4. Evolution of data center
5. Structured data
6. No Big Data

1990 - 2005



1. Evolution of ERP (monolithic)
2. Collect lot of data
3. Integrated cross departmental business applications
4. Mostly departmental reporting
5. Some cross departmental
6. Data (center) grew
7. Purchase for peak capacity – Capital lock
8. Start of unstructured data
9. Start of mobile devices
10. Big Data only for large organizations

2005 – 2015



1. Explosion of data
2. Explosion of devices
3. Social media
4. Just ERP – no value
5. Just reports – no value
6. DW / BI critical for competition
7. Cloud prominent
8. Start of IaaS, SaaS, PaaS
9. Beginning of decoupling of ERP
10. Back to purpose built business application on cloud – SFDC, Workday
11. Part ERP on Prem, Part ERP on cloud – Hybrid
12. Big Data – seen as valuable by most organizations (cost)
12. Too much data

2015 – 2020



1. Era of Hybrid computing (on-prem and cloud compute)
2. ERP and BI
3. More and more applications on cloud
4. Building custom integration between apps acceptable
5. More and more BI on cloud
6. Big Data – necessity for most
7. Integrated cloud applications by mega vendors
8. Integration of traditional DW with Big Data
9. Data lakes – dump everything
10. Cannot define what analysis upfront



Performance. Growth. Excellence.

› Global Reach for Global Customers

- Founded in 2002 in Boston - 650+ employees
- US (East, Central, West),
- Europe, India, Middle East

› Most Recognized & Credentialed Oracle BI & EPM Partner

- P3 Partner Group – Certified, Specialized and Recognized by Oracle for BI/EPM.
- Beta program for OBI 12c, ODI 12c
- Onsite, Remote and Offshore Delivery Models
- BICS Accelerators including BICS and PBCS integration
- AWS Certified for Oracle on AWS (only 2 partners worldwide)
- First Oracle Exalytics Certified Delivery Partner
- One of the First partner to be trained on BDD



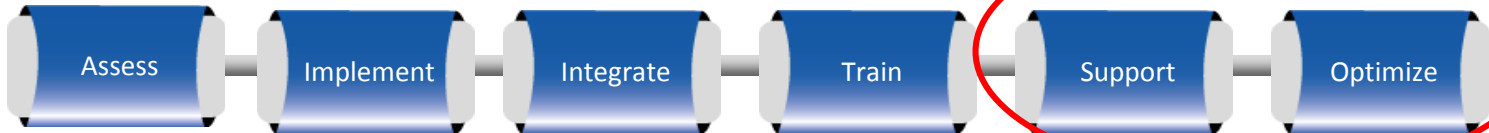
Exalytics Certified



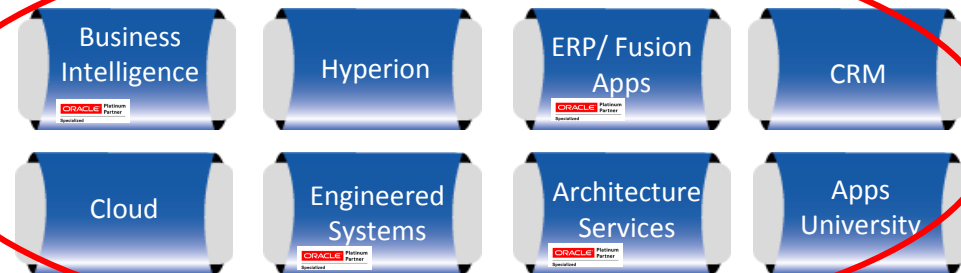


Multi-Pillar Partner Engaging Across the IT Lifecycle

Lifecycle of Services



Oracle Centers of Excellence



Strong Process Background – Engage Business
Architecture/Security/Performance
Multiple Distributed Sources
Integration. Integration. Integration
Training & Adoption



Exalytics Certified Delivery Partner



Oracle Specialized/Validated





Complete BI and EPM Service Offerings (all platforms)

Advise

- Big Data Discovery
- BI / EPM Assessment and Roadmaps
- OBI Architecture Assessment
- Data Governance Process Design and Implement
- MDM for BI / EPM
- OBI / Hyperion and R12 Planning
- Integration of structured and unstructured data

Implement

- OBI Foundation Suite
- BI Apps (ODI, Informatica)
- EPM
- BDD
- User Adoption & Training
- BI / EPM on the Cloud
- Endeca Information Discovery (EID)
- EID for EBS (connectors)
- BIFS
- Exalytics

Optimize & Support

- Upgrades
- Support Services for OBI and EPM
- User Training
- Functional Design
- Tool Adoption Methods
- Report Rationalization
- New Module Development
- Managed Services

Integrate

(Representative sources)

- EBS, PSFT, JDE, SAP, LAWSON
- Salesforce (Bi Directional)
- Siebel
- Clarity
- Kronos
- PPM
- ADP
- Demantra
- Oracle Governance (GRC)
- Oracle Learning Management



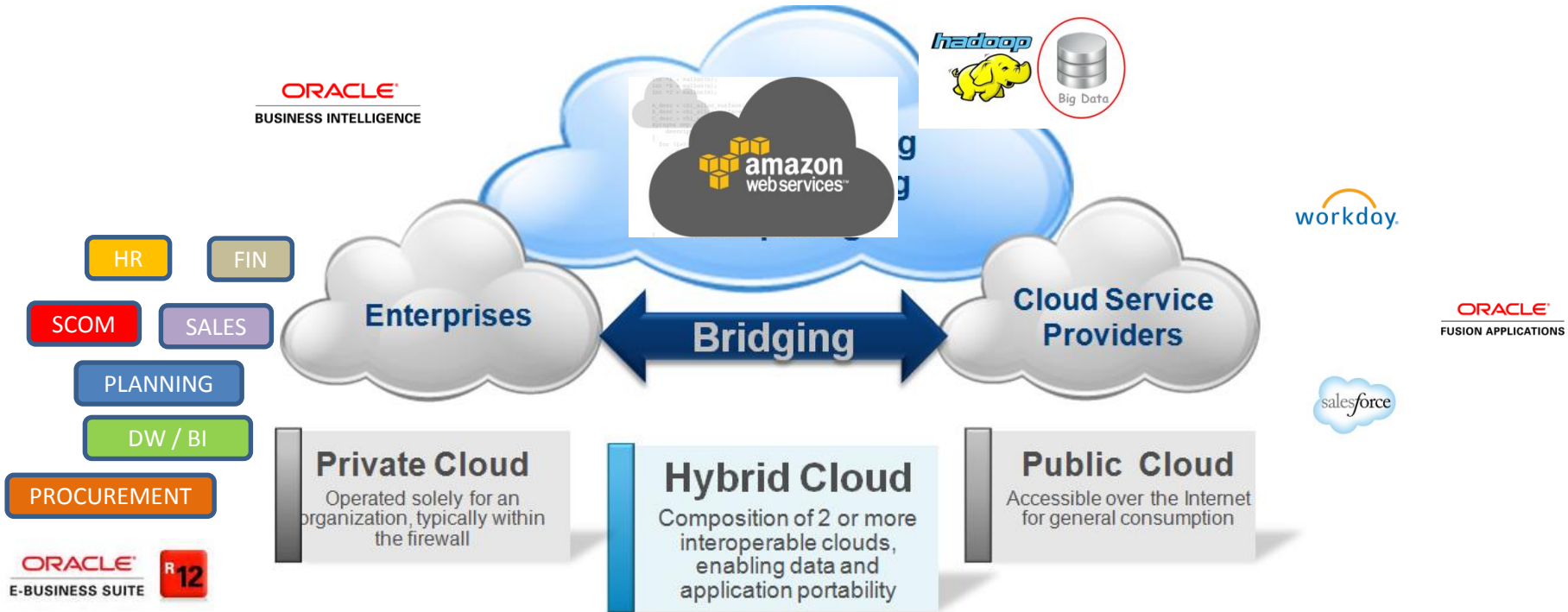
Oracle Database Appliance



On premise



Hybrid Cloud Framework





Analytics In Hybrid Cloud : Problem with this

Private Cloud / On Prem

Public Cloud (SaaS)

Public Cloud (IaaS) / Oracle Big Data Appliance

Transactional Data

Transactional Data

Big Data

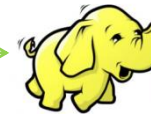
Transactional Data

File Data

Streaming Data

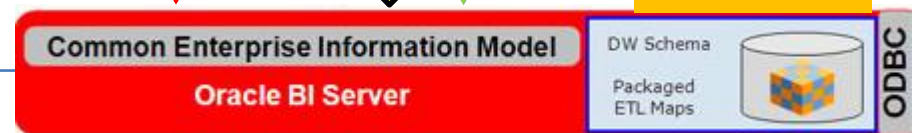


On Prem Traditional DW



Cloud Hadoop DW

On Prem or IaaS OBI-EE





What is a Data Lake



Structured

Unstructured

Store all data
No need to model
Define relationship when needed
Allow access thru tools
Store results back or into D/W



Build on low cost storage
Cloud best suited
Scalability and Performance
Can be just Structured
Achieve or Staging for D/W

Process

Visualize

ORACLE® Big Data Discovery

The Need





Key Challenges in Managing Big Data

Key Big Data Challenges



Addressed by Big Data Discovery

“Through 2015, more than 85 percent of Fortune 500 organizations will fail to effectively exploit big data for competitive advantage.”

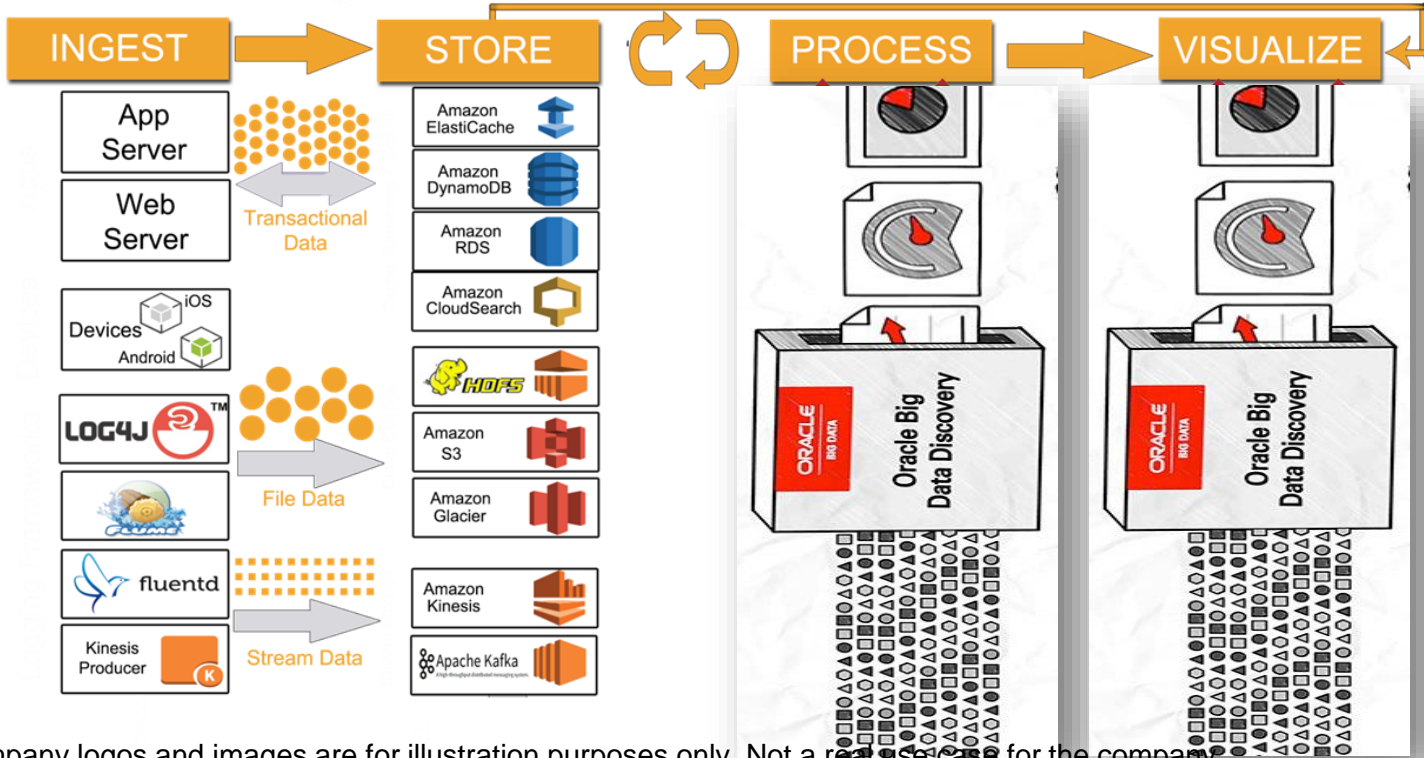
– Gartner

Figure 1: What key challenges do you face or foresee in managing big data?

Informatica Study May 2013



BDD Value Proposition



Note: company logos and images are for illustration purposes only. Not a real use case for the company.

Source : AWS



The Visual Face of Hadoop

The screenshot displays the Oracle Big Data Discovery interface, which is a data analytics platform. It features a sidebar for navigation, a main workspace with various data sets and visualizations, and a bottom section for data exploration. The interface is overlaid with six red circular icons representing key data processing stages: find, explore, transform, discover, and share. The 'find' icon is a magnifying glass, 'explore' is an eye, 'transform' is a wrench, 'discover' is a lightbulb, and 'share' is a share symbol. The background shows a complex dashboard with charts, tables, and filters, illustrating the platform's capabilities in handling large-scale data.

Source : Oracle



Behind the Scenes Components

- Hadoop
- MapReduce
- HDFS
- Hive
- HCatalog
- Spark
- Impala
- Pig
- YARN
- Oozie

Oracle Big Data Discovery. Technical Innovation on Hadoop





Supported Hadoop Distributions

1. CDH 5.3.x - 5.4.x Cloudera Distribution for Hadoop
2. HDP 2.2.4 - 2.3.x Hortonworks Data Platform



Required Hadoop Components

Component	Description
Cloudera Manager (CDH)/Ambari (HDP)	<p>The BDD installer uses a RESTful API to query Cloudera Manager (if you're using CDH) or Ambari (If you're using HDP) for information about specific Hadoop nodes, such as their hostnames and port numbers.</p> <p>Cloudera Manager/Ambari must be installed on at least one node in your cluster, although it doesn't have to be on any that will host BDD.</p>
ZooKeeper	<p>BDD uses ZooKeeper services to manage the Dgraph instances and ensure high availability of Dgraph query processing. ZooKeeper must be installed on at least one node in your cluster, although it doesn't have to be on any that will host BDD. For more information on ZooKeeper and how it affects the cluster deployment's high availability, see the Administrator's Guide.</p> <p>All Managed Servers must be able to connect to a node running ZooKeeper.</p>
HDFS	<p>BDD stores the Hive tables that contain your source data in HDFS. HDFS must be installed on at least one node in your cluster, although it doesn't need to be on any that will host BDD. HDFS must be installed on all nodes that will run Data Processing.</p>
HCatalog	<p>The Data Processing Hive Table Detector monitors HCatalog for new and deleted tables that require processing. HCatalog must be installed on at least one node in your cluster, although it doesn't have to be one that will host BDD.</p>



Required Hadoop Components

Component	Description
Hive	<p>All of your data is stored as Hive tables on HDFS. When BDD discovers a new or modified Hive table, it launches a Data Processing workflow for that table.</p> <p>Hive must be installed on all nodes that will run Data Processing.</p>
Spark on YARN	<p>BDD uses Spark on YARN to run all Data Processing jobs. Spark on YARN must be installed on all nodes that will run Data Processing.</p>
Hue	<p>You can use Hue to load your source data into Hive and to view data exported from Studio.</p> <p>Note: HDP doesn't include Hue. If you have an HDP cluster, you must install it separately and set the HUE_URI property in BDD's configuration file. You can also use the bdd-admin script to update this property after installation, if necessary. For more information, see the Administrator's Guide.</p>
YARN	<p>YARN worker nodes run all Data Processing jobs. YARN must be installed on all nodes that will run Data Processing.</p>
Sentry	<p>BDD doesn't require Sentry. However, if your Hadoop cluster uses it, you must configure it to allow BDD access to the tables it requires in Hive.</p>



Find (Catalog) – Data Sets Tab

- Searchable
- Navigable
- Recently Viewed, Most Popular, Newly Added
- Preview

ORACLE Big Data Discovery

Search Everything

Ingested Data Sets

30 Projects [View all](#)

258 Data Sets [View all](#) [+ Data Set](#)

Personal File Upload

Refine By

- USAGE
 - Created By Me
- CONTENT
 - Contains Dates
 - Contains Locations
 - Language
- METADATA
 - Project Tags
 - Data Set Tags
 - Last Modified
 - Number of Records
 - Number of Attributes

Navigate by Metadata

Recently Viewed Data Sets

- Customer Snapshot**
Snapshot of Customer Demographics
(304797 records) [Preview](#)
- Purchase Logs (Enriched...)**
Web logs enriched with purchases and pr...
Tags: Demo Tag
(2147483647 records) [Preview](#)
- Census Population**
Population Demographics
Data Source: Census Population.xlsx (33120 records)
[Preview](#)
- Purchase Logs (Enriched)**
Web logs enriched with purchases and pr...
(2147483647 records) [Preview](#)

Most Popular Data Sets [View More](#)

- Youtube Crawl**
Google Adwords Youtube Campaign Anal...
Data Source: YoutubeCrawl.csv (9077 records) [Preview](#)
- Customer Snapshot**
Snapshot of Customer Demographics
(304797 records) [Preview](#)
- Purchase Logs (Enriched...)**
Web logs enriched with purchases and pr...
Tags: Demo Tag
(2147483647 records) [Preview](#)
- Brand Loyalty Surveys**
Brand Loyalty Surveys
(20981715 records) [Preview](#)
- Cart Web Logs**
Web logs for shopping cart
(2147483647 records) [Preview](#)
- Census Population**
Population Demographics
Data Source: Census Population.xlsx (33120 records)
[Preview](#)
- CensusEstimates Civilia...**
CensusEstimates CivilianPopulation2009
(22360385 records) [Preview](#)
- CensusEstimates Overs...**
CensusEstimates OverseasPopulation2009
(38674660 records) [Preview](#)

Newly Added Data Sets [View More](#)

- Purchase Logs (Enriched)**
Web logs enriched with purchases and pr...
(2147483647 records) [Preview](#)
- Cart Web Logs**
Web logs for shopping cart
(2147483647 records) [Preview](#)
- Purchase Logs (Enriched...)**
Web logs enriched with purchases and pr...
Tags: Demo Tag
(2147483647 records) [Preview](#)
- Census Household Inco...**
Household income by demographic from A...
Data Source: Census Household Income... (32936 records) [Preview](#)

Preview

Source : Oracle



Find (Catalog) – Data Set Quick Look

- Used in Projects
 - Project name
 - Data Sets used
 - Created by

Customer Snapshot
Snapshot of Customer Demographics
(304797 records)
[Preview](#)

Purchase Logs (Enriched, Platform)
Web logs enriched with purchases and pr...
Tags: Demo Tag
(2147483647 records)
[Preview](#)

Census Population
Population Demographics
Data Source: Census Population.xlsx
(33120 records)
[Preview](#)

Purchase Logs (Enriched, Platform) (2147483647 records)

[Data Set Info](#) [Used in Projects \(8\)](#) [Related Data sets \(2\)](#)

Projects using this data set

Customer Segmentation (Last Updated: 9/26/14 10:21:42 PM EDT)
Data Sets: Customer Snapshot, Purchase Logs (Enriched, Platform), Youtube Crawl
Created By: Admin Admin

Customer Segmentation2 (Last Updated: 11/11/14 3:28:59 PM EST)
Data Sets: Customer Snapshot, Purchase Logs (Enriched, Platform)
Created By: Admin Admin

Demo (Last Updated: 10/1/14 9:29:42 PM EDT)
Data Sets: Purchase Logs (Enriched, Platform)

Actions
[Explore](#)
[Add to project](#)
[Edit Tags](#)
[Delete](#)

Summary
64 Views
Last Updated
9/11/14 3:05:04 AM EDT

Data Set Used in Project Quick Look

Source : Oracle



Explore

- Search:
 - Attributes
 - Refinements
 - Keyword
- Sort order
 - Name
 - Information Potential
 - Relationship to an attribute
- Navigable

ORACLE Big Data Discovery

Search

Explore Transform Discover

ChicagoCrime_20141111

ADD TO PROJECT

VISUALIZE THE DATA

Add attributes to visualize relationships

96K of 96K Records Sampled 96K Records Viewed 27 Attributes

DATA TYPE: Boolean ◯ DateTime ◯ Geocode ◯ Numeric ◯ String ◯

Sort: By name

Navigation menu

Search menu

Tiles vs. Tabular view

Sort Order

Visualizations shown:

- arrest: Horizontal bar chart showing 'false' and 'true' counts.
- beat: Histogram showing distribution of beat numbers from 0 to 2600.
- block: Text list of addresses including 008XX N MICHIGAN AVE, 000XX W TERMINAL ST, 001XX N STATE ST, 008XX N STATE ST, 000XX N STATE ST, 076XX S CICERO AVE, and 22354 others.
- case_number: Horizontal bar chart showing counts for HT341462, HT534028, HT590573, HT332050, HT446294, and 96157 others.
- community_area: Histogram showing distribution of community area numbers from 0 to 78.
- date: Line chart showing data from January 2011 to December 2011.
- description: Text list of descriptions including SIMPLE DOMESTIC BATTERY SIMPLE, \$500 AND UNDER, POSS: CANNABIS 30GMS OR LE, and TO VEHICLE ENTRY, with 263 others.
- district: Histogram showing distribution of district numbers from 0 to 33.
- domestic: Horizontal bar chart showing 'false' and 'true' counts.
- fbi_code: Horizontal bar chart showing counts for 6, 08B, 14, 18, 26, 5, and 19 others.

Source : Oracle



Explore – Sort Attributes

- Sort order
 - Name (alpha)
 - Information Potential
 - Based on Entropy
 - Relationship to an attribute
 - Based on Information Gain

The screenshot displays the Oracle Data Explorer interface. At the top, there is a search bar and an 'Add to project' button. Below this, a list of attributes is shown, including 'arrest', 'beat', 'block', 'case_number', 'community_area', 'date', 'description', 'district', 'domestic', 'fbi_code', 'id', 'iucr', and 'latitude'. A bar chart visualization is overlaid on the list, showing the distribution of values for the 'community_area' attribute. The 'Sort' dropdown menu is open, showing three options: 'By name', 'By information potential', and 'Relationship to an attribute'. A blue arrow points from a 'Sort by' button to the 'Sort: By name' option in the dropdown.

Source : Oracle



Explore – Quick Look – Geocode

- Overview
- Details
- Summary stats
- Refine-able

The screenshot displays a data exploration interface for a field named 'location'. At the top, a navigation bar shows several fields: 'id', 'iucr', 'latitude', 'location', and 'location_description'. The 'location' field is highlighted with a yellow box and a mouse cursor. Below this, the 'location' field is selected, showing a globe icon and the text 'From ChicagoCrime_20141111'. To the right of the field name are buttons for 'Add to scratchpad', 'FAVORITE', 'SORT', and 'HIDE'. Below the field name are two tabs: 'Overview' and 'Details'. The 'Overview' tab is active, showing a map of the United States. The map has a legend titled 'Record Count' with two categories: '48082 or less' (light blue) and 'More than 48082' (teal). The map shows the state of Illinois highlighted in teal. To the right of the map is a 'Summary statistics' section with a table:

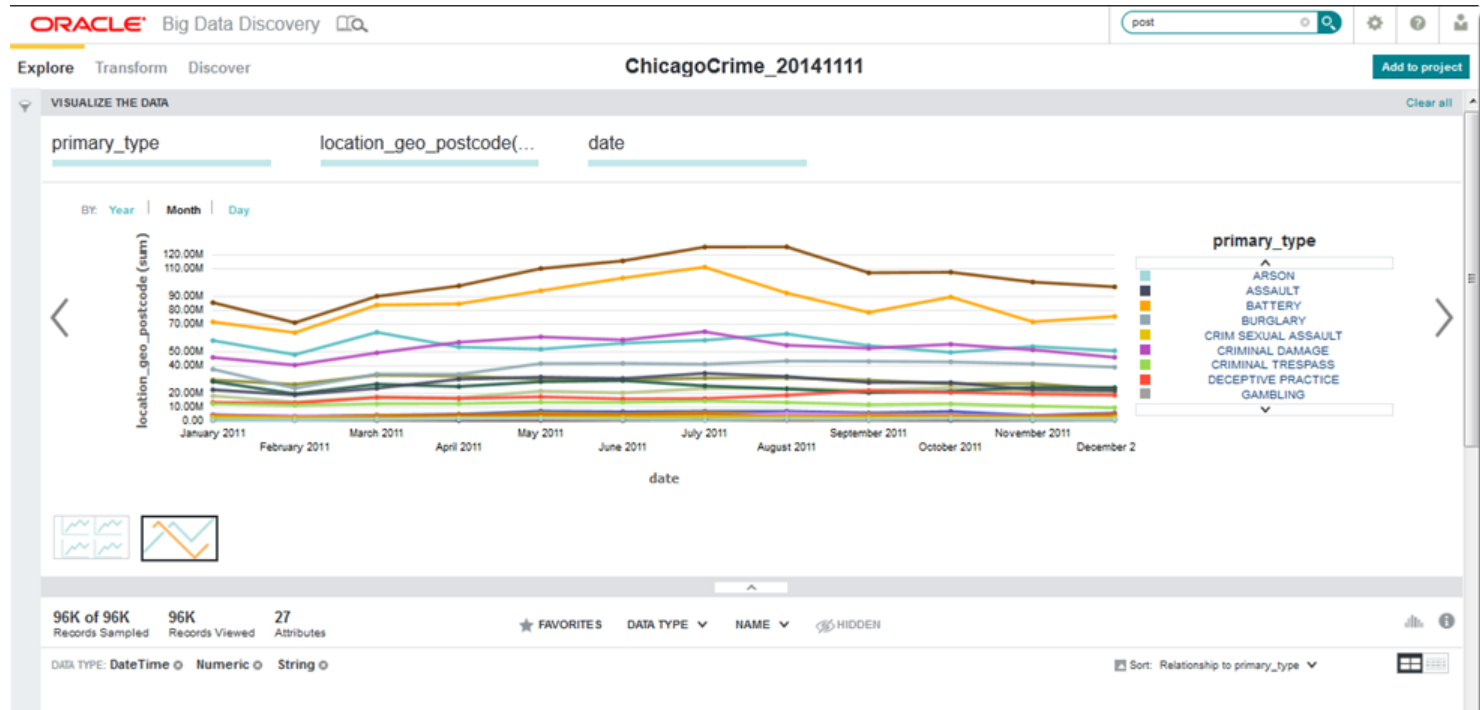
Summary statistics	
Valid	96,165
Missing	0
Anomalies	0
Unique values	71,864

Source : Oracle



Explore – Scratchpad

- Graphic type changes as additional attributes are added
- Autoselects best visualization
- Offers next best graphics option(s)



Source : Oracle



Transform - Overview

The screenshot displays the Oracle Big Data Discovery interface for a project named "Chicago Sample Project: ChicagoCrime_20141111". The interface includes a navigation menu on the left with options like "Explore", "Transform", and "Discover". A search bar is located at the top right. The main area shows a data table with columns: arrest, # beat, block, case_number, and # community_area. A context menu is open over the "block" column, listing options such as "Quick Transformation", "Absolute Value", "Convert to...", "Advanced Transformation", "Data Cleansing", "Custom Transform", "Other Actions", "Sort", "Favorite", and "Hide".

Annotations on the screenshot include:

- Smart Attribute Filtering**: Points to the "Attributes" section at the top of the table.
- Faceted and Record Level Views**: Points to the "Sort: By name" dropdown and the table's header area.
- Attribute Transformation**: Points to the context menu over the "block" column.
- Full Guided Navigation**: Points to the left-hand navigation pane.
- Visual Data Quality Summaries**: Points to the "Visual Data Quality Summaries" icon in the top right.
- Interactive Transform History**: Points to the "Interactive Transform History" icon in the top right.

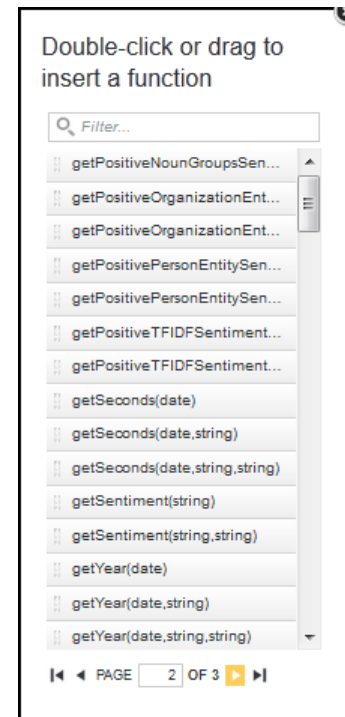
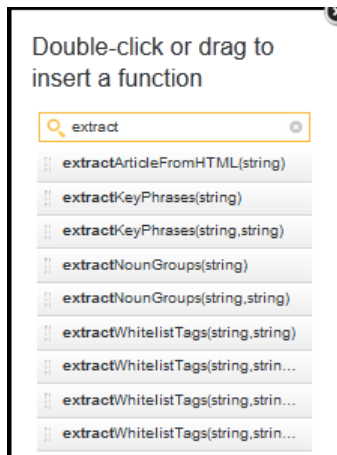
arrest	# beat	block	case_number	# community_area
false	1,512		HT185911	63
true	1,313		HT371874	30
false	1,121		HT405341	25
true	933		HT234255	
false	921		HT215173	
true	824		HT431858	
false	1,032		HT321800	61
true	2,322		HT642765	63
false	632	095XX S COTTAGE GROV...	HT564387	44
true	823	035XX W 62ND PL	HT476019	66
false	813	040XX W 59TH ST	HT222062	62
false	824	000V W 50TH ST	HT504983	46

Source : Oracle



Transform – Function Families

- Text Enrichment functions (partial)
 - Key Phrase
 - Whitelist
 - Entity
 - Sentiment
 - TFIDF sentiment
 - Language detection
 - Noun Groups

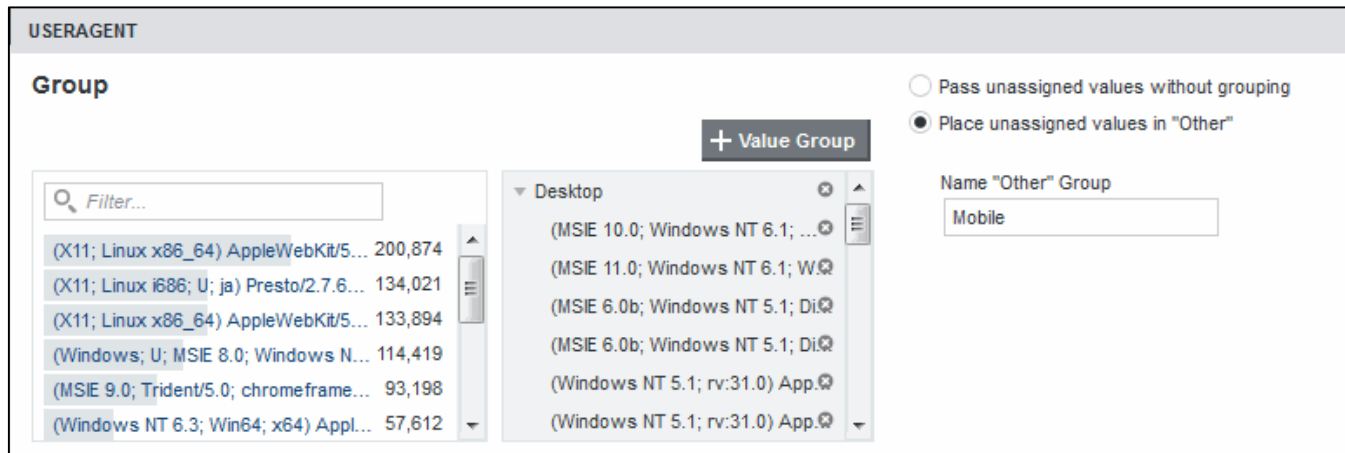
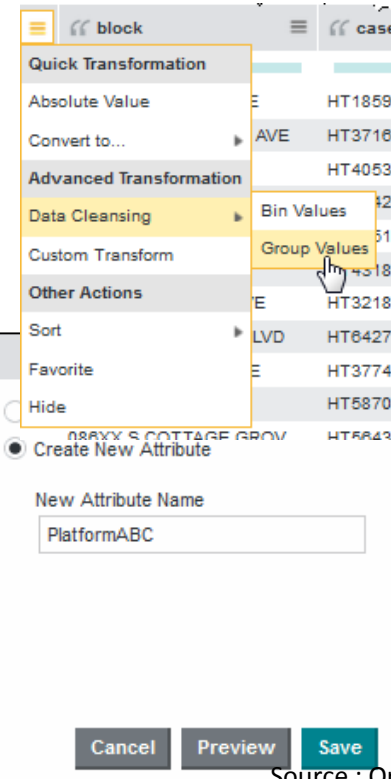


Source : Oracle



Transform – Function Families

- Data Cleansing
 - Grouping
 - Applies to string type attributes
 - Can be applied to existing attribute or create a new one
 - Accessed from Quick Transformation menu



Source : Oracle



Transform – Function Families

- Data Cleansing
 - Splitting
 - Used to separate multi-value attributes

ORACLE Big Data Discovery

Explore Transform Discover Warranty_Test: warranty_10_vehicles

Split

modelyear_modelname

Select the delimiter to use to identify how to split the current value into multiple values

Split method: Delimiter

Values delimited by: Custom

Custom delimiter: -

New Attribute 1: modelyear_split1

New Attribute 2: modelyear_split2

Cancel Preview Save

999K of 5M Records Sampled 999K Records Viewed 135 Attributes

FAVORITES DATA TYPE NAME HIDDEN

All Attributes

fleetindicator	modelyear_modelname	modelyear_split1	modelyear_split2	# nnsalesmodelcode	p_companyvehicle...	p_customername	p_dealerordernumber	p_distrib
Y	2007 - MOSTA	2007	MOSTA	8,217	Neither Affiliate or Company ...	PERSON 675634	R510773	307
N	2007 - TUFFIN	2007	TUFFIN	79,617	Neither Affiliate or Company ...	PERSON 2353464	T880468	307
N	2007 - MYSTA	2007	MYSTA	7,817	Neither Affiliate or Company ...	PERSON 2115767	S893430	307
Y	2007 - MOSTA	2007	MOSTA	8,217	Neither Affiliate or Company ...	PERSON 675634	X811298	307
Y	2007 - MYSTA	2007	MYSTA	7,617	Neither Affiliate or Company ...	PERSON 675634	X819693	307
N	2007 - MYSTA	2007	MYSTA	7,817	Neither Affiliate or Company ...	PERSON 1278004	W735350	307
N	2006 - EXCELLA	2006	EXCELLA	5,216	Neither Affiliate or Company ...	PERSON 1326995	V757166	307
N	2007 - TUFFIN	2007	TUFFIN	79,417	Neither Affiliate or Company ...	PERSON 579465	Z861877	307
N	2007 - MYSTA	2007	MYSTA	7,817	Neither Affiliate or Company ...	PERSON 1493973	U956621	307
N	2007 - EXCELLA GCC	2007	EXCELLA GCC	55,717			U974384	307
N	2009 - NUMBA-3 SEDAN	2009	NUMBA-3 SEDAN	90,619	Neither Affiliate or Company ...	PERSON 1608673	E884365	307
N	2007 - MYSTA	2007	MYSTA	7,817	Neither Affiliate or Company ...	PERSON 230984	V920544	307
N	2009 - TUFFIN	2009	TUFFIN	59,219			GC44182	307
N	2006 - EXCELLA	2006	EXCELLA	5,216	Neither Affiliate or Company ...	PERSON 1386633	T718008	307
N	2007 - EXCELLA	2007	EXCELLA	5,717	Neither Affiliate or Company ...	PERSON 2127999	Y870540	307

Source: Oracle

WARRANTY CLAIMS ▬ Clear All

vehicledealergeocode_geo_countr
US

REFINE BY ✕

claimamount

claimdate

Between ▾

5/27/2009 - 8/23/2012

8620 results Submit

claimmonth

claimperiod

claimyear

commodity

Filter...

- BATTERY
- SERVICE BRAKES, HYDRAULIC...
- ENGINE AND ENGINE COOLING...
- ELECTRICAL SYSTEM
- VEHICLE SPEED CONTROL
- AIR BAGS
- POWER TRAIN-AUTOMATIC TRA...
- ELECTRICAL SYSTEM-ALTERNA...
- SERVICE BRAKES, HYDRAULIC...
- SERVICE BRAKES, HYDRAULIC...

Select All

complaint_key_phrases

complaint_lang

laboramount

labordescription

make

manufacturer

mis

model

modelyear

partamount

COMPONENT CONTAINER

THEMATIC MAP

claimamount (sum)

- 9,275 or less
- 9,275 - 29,248
- 29,248 - 53,502
- 53,502 - 74,712
- 74,712 - 101,779
- 101,779 - 145,181
- 145,181 - 275,387
- More than 275,387

VIEW GEO ATTRIBUTE GEOGRAPHIC GRAIN

Warranty Claims vehicledealergeocode State

SUMMARIZATION BAR

8,820 397.96

Claims Claim Amt (Median)

Bleed Lines

Top Repair

TAG CLOUD

AND BRAKE **engine**

generator HAD HAS HAVE IS JB

lights ON

satisfaction THAT THE

THE CAR THE VEHICLE TO VEHICLE

warning WAS

Explore complaint_key_phrases by Number of records

RESULTS TABLE

General 0 RECORDS SELECTED VIEW OPTIONS ACTIONS

	saletedate (Year-Month...	partnumber	model	laboramount	claimyear	productionyear	compla
<input type="checkbox"/>	11/21/08	p256CALWVVMB-3203	CAMRY	117.12	2,010	2,008	en
<input type="checkbox"/>	11/28/08	p796JBB6NEQ2B-1050	ML500	350.1	2,010	2,008	en
<input type="checkbox"/>	4/25/09	p87304MAFS0IA-90238	MUSTANG	146.64	2,009	2,008	en
<input type="checkbox"/>	11/4/09	ba260ANFG45GHXMN2...	PACIFICA	188.91	2,009	2,009	en
<input type="checkbox"/>	1/24/09	p1468Z160GWVVA-1444	SILVERADO 1500	363.55	2,010	2,009	en
<input type="checkbox"/>	4/26/09	p145ABJKE10LO4-23108	GRAND CHEROKEE	108.49	2,009	2,009	en
<input type="checkbox"/>	11/4/09	ba260ANFG45GHXMN2...	MUSTANG	237.22	2,009	2,008	en



Discover

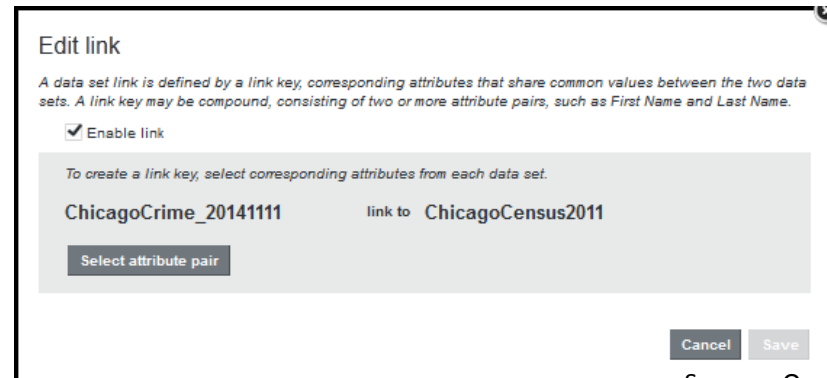
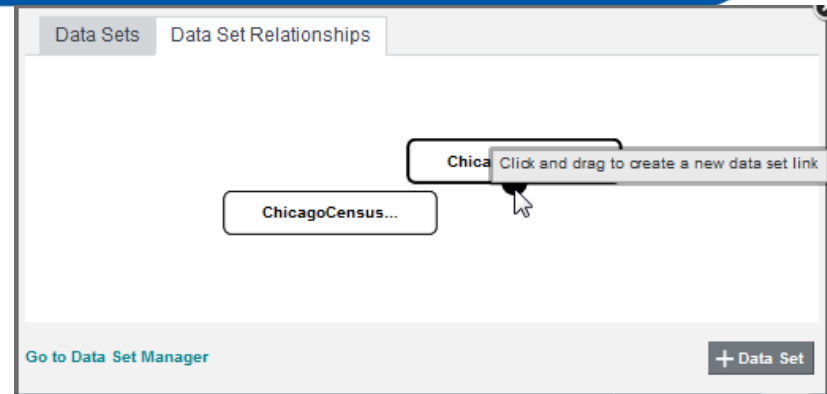
- Join and blend data for deeper perspectives
- Compose project pages via drag and drop
- Use powerful search and guided navigation to ask questions
- See new patterns in rich, interactive data visualizations

Source : Oracle



Discovery

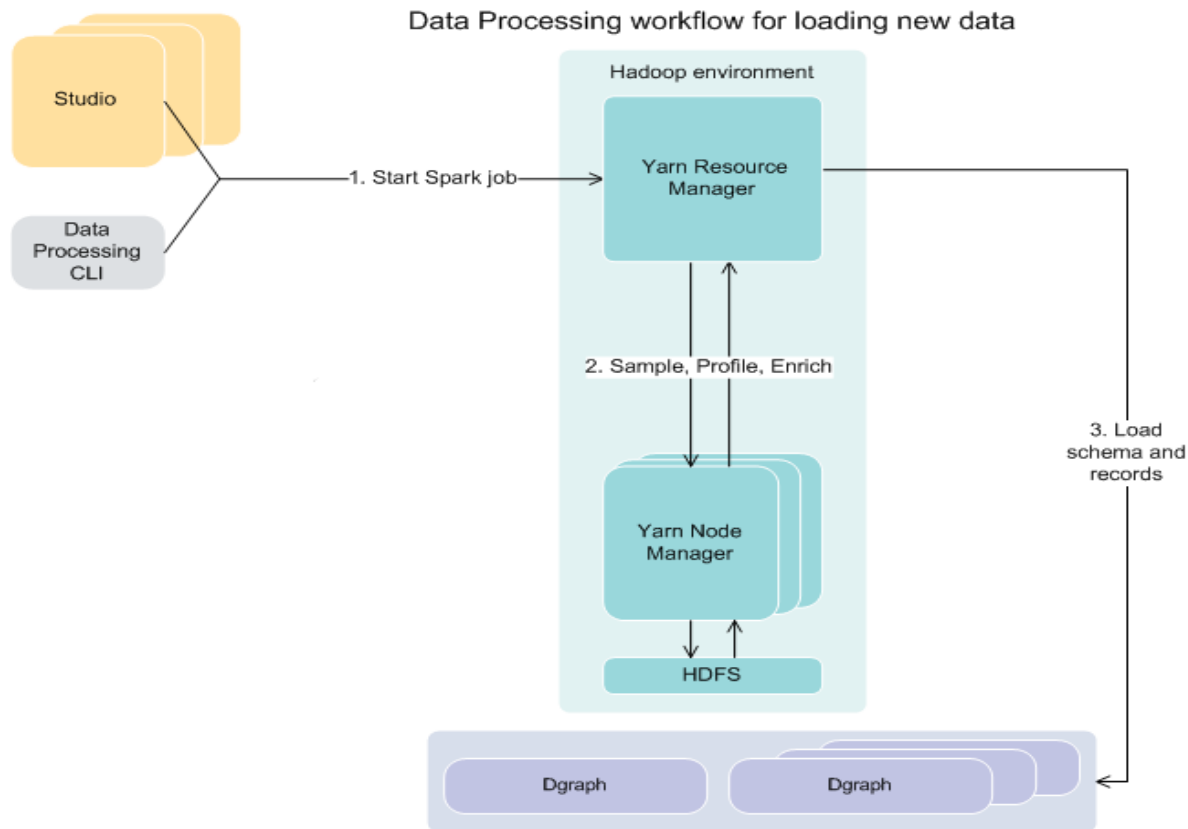
- Data Set Linking
 - Links 2 or more data sets within a project
- Highlights
 - Visual attribute linking
 - Automatic view creation
 - Auto-widening of datasets
 - Linked Navigation



Source : Oracle



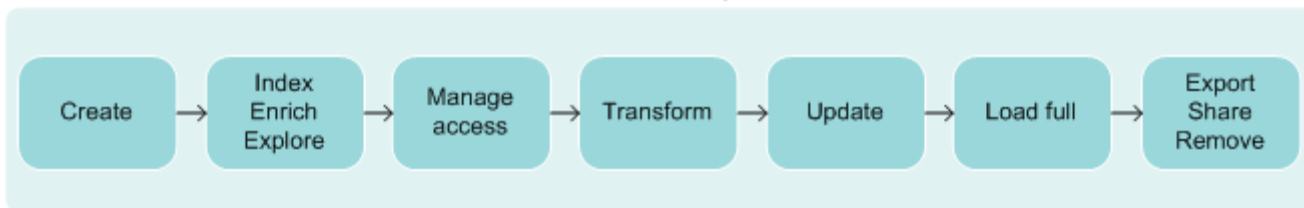
Data Processing Workflow





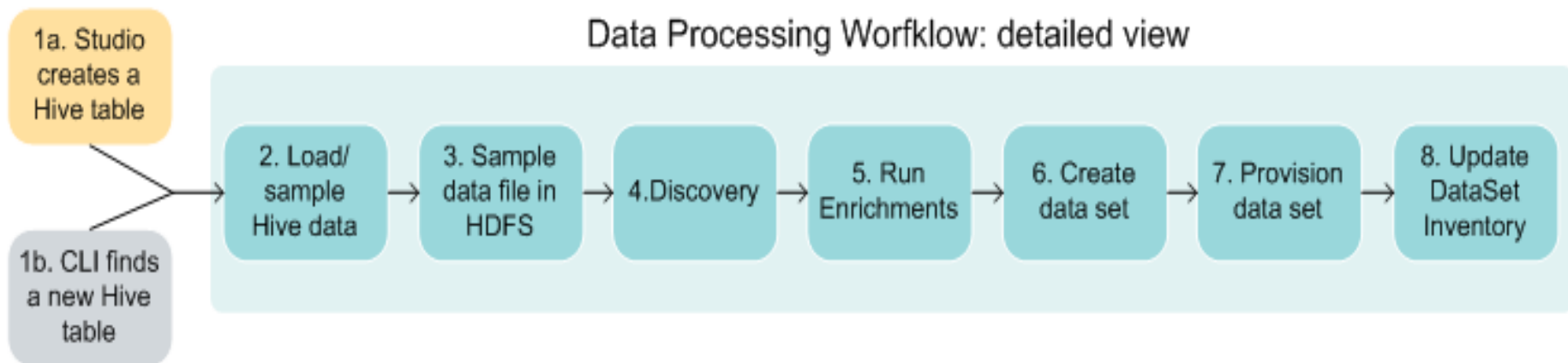
Data Set Lifecycle

Data set lifecycle





New Hive Table Workflow

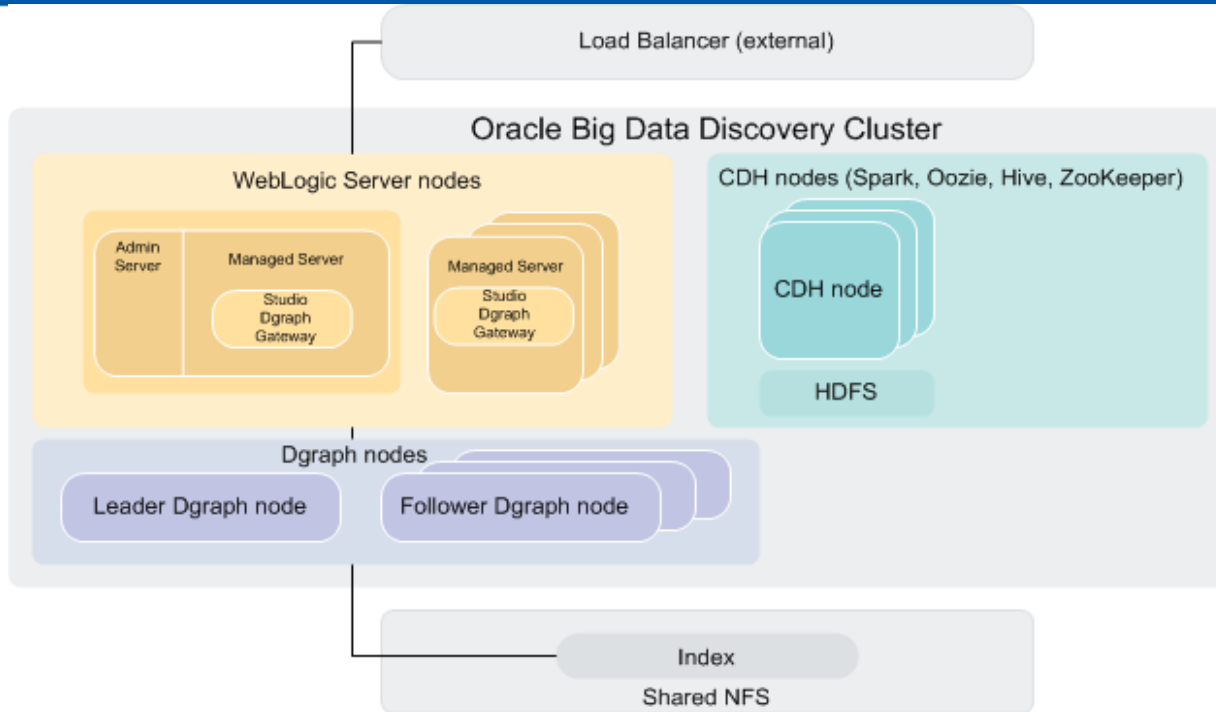




Exporting from Studio



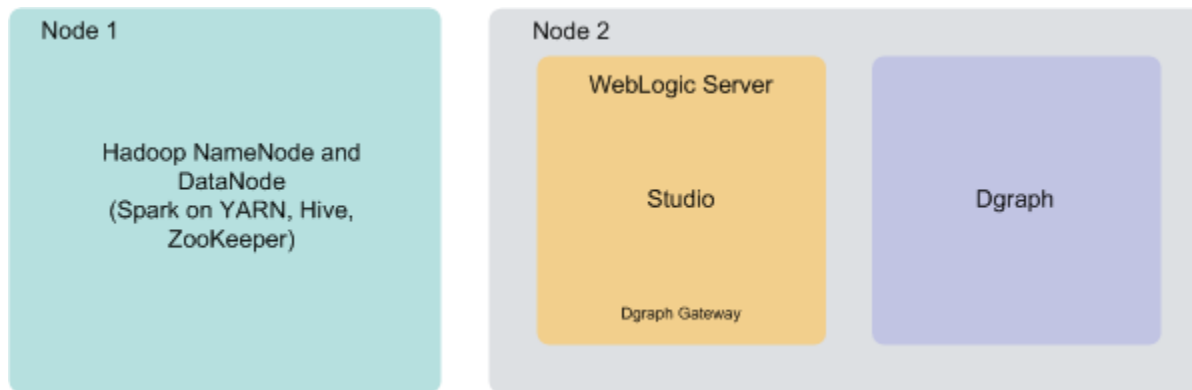
1. From Transform in Studio, you can select to export the data into HDFS. This sends an internal request to export the data to the Dgraph.
2. The Dgraph communicates with the Dgraph HDFS Agent, which launches the data exporting process and writes the file to HDFS.
3. Optionally, you can choose to create a Hive table from the data. If you do so, the Hive table is created in HDFS.



1. Nodes that host WebLogic Server with Studio and Dgraph Gateway.
2. CDH only nodes. These nodes do not host WebLogic Server or Dgraph instances. They run Data Processing jobs, within a BDD deployment.
3. Dgraph nodes. These nodes are solely dedicated to hosting Dgraph instances.



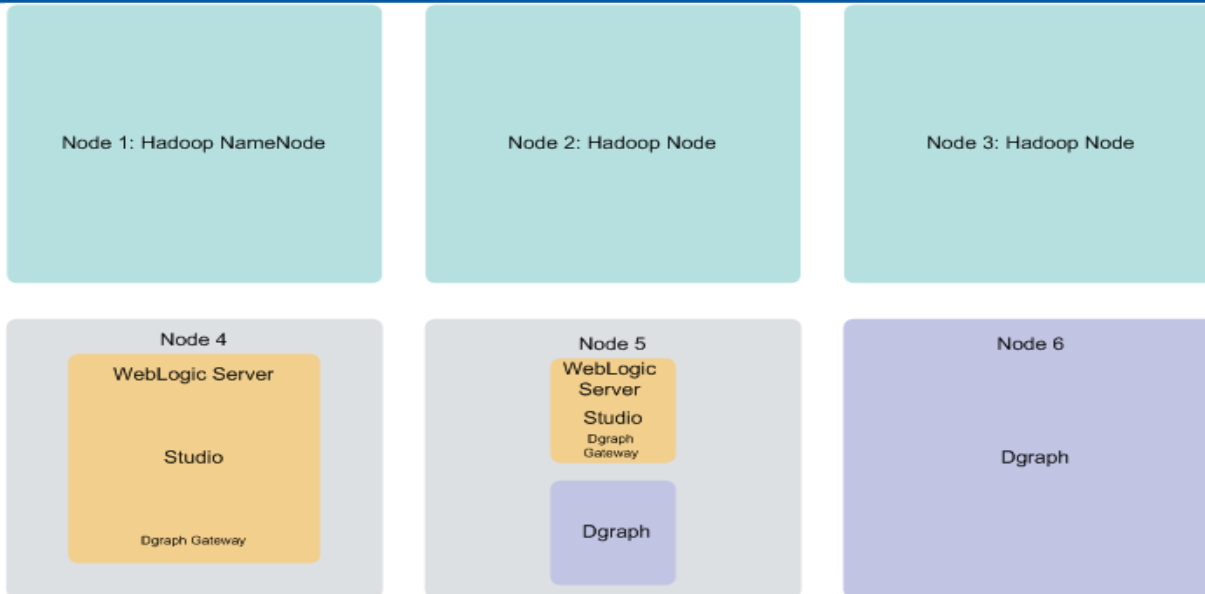
Two Node Development Deployment



1. You can deploy BDD to two nodes for a development environment. This configuration can handle a slightly larger index than a single-node configuration, but is not recommended for production as it does not provide high availability of Dgraph or Studio services and also has limited capacity for processing queries on high volumes of data.
2. In a two-node configuration, Hadoop (including the NameNode and one DataNode) is hosted on the first node. The WebLogic Server (including Studio and the Dgraph Gateway) and the Dgraph instance are hosted on the second node.



Six Node Production Deployment



1. Nodes 1, 2 and 3 are running Hadoop. Note that BDD is also deployed on these nodes. After the installation, Data Processing jobs are launched from these nodes and run on other BDD nodes. Having three Hadoop nodes ensures enhanced availability of BDD services, including query processing performed by the Dgraph.
2. Nodes 4 and 5 are running WebLogic Server with Studio. This ensures minimal redundancy of the Studio instances.
3. Nodes 5 and 6 are running the Dgraph instances. This creates a Dgraph cluster within the BDD cluster, which in turn increases the availability of query processing.



Benefits of BDD for Data Lakes



- Give power of analysis on big data to power users
- Reduced dependency on Data Scientists
- Data Scientists can focus on core analysis
- Easy to define structure on data using GUI based
- GUI will answer over 70% of scenarios

- Works on commercial hardware
- Works on Oracle Big Data Appliance
- Works as a service (coming soon)

- Allows to integrate RDBMS
- Allows to integrate 'My Data'
- Allows to save results back in Hadoop
- Allows to save results in WDW
- Allows to perform full 360 degree analysis
 - Tie back key discovery from BDD to actual transaction for the company



Five Easy Steps from Data to Insight



Find

relevant data



Explore

the data to understand its potential



Transform

and enrich the data to make it ready for analysis



Discover

powerful new insights



Share

those insights for enterprise leverage



Oracle Big Data Discovery Videos, Training and Manuals

1. <https://www.oracle.com/big-data/big-data-discovery/index.html>
2. https://apexapps.oracle.com/pls/apex/f?p=44785:141:101487530416965::NO::P141_PA GE_ID%2CP141_SECTION_ID:157%2C1816

OBDD Learning Library

3. https://www.youtube.com/channel/UC0B-dhxifP7R_1Py-2IS2iw

More videos: Linking Data

4. https://docs.oracle.com/cd/E64107_01/index.html

Version 1.1 Documentation



ORACLE® Cloud



Big Data Discovery

Coming Soon



C
T
A
tra

Key Contacts

Kshitij Kumar
CTO

Kshitij.Kumar@AppsAssociates.com

Wilfrido Solano
Practice Director

Wilfrido.Solano@AppsAssociates.com



North America

- ▶ Boston (Headquarters)
- ▶ New York
- ▶ Atlanta
- ▶ Chicago

Asia

- ▶ India
Global Development Center

Europe

- ▶ Germany
- ▶ Netherlands

Middle East

- ▶ Oman

www.appsassociates.com