BIG DATA

**BIG DATA DISCOVERY KEY FEATURES** 

all data in Hadoop

· Access a rich, interactive catalog of

 Use familiar search and guided navigation to find information quickly

· See data set summaries, user annotation and recommendations Provision personal and enterprise data to Hadoop via self-service

Visualize all attributes by type

Sort attributes by information

· Assess attribute statistics, data

· Use a scratch pad to uncover

correlations between attributes

quality and outliers

Find

Explore

potential

Transform



Disclaimer: This document is for informational purposes. It is not a commitment to deliver any material, code, or functionality, and should not be relied upon in making purchasing decisions. The development, release, and timing of any features or functionality described in this document remains at the sole discretion of Oracle.

# Oracle Big Data Discovery The Visual Face of Hadoop

Today's Big Data challenge is not how to store it, but how to make sense of it. Oracle Big Data Discovery is a fundamentally new approach to making sense of Big Data, empowering organizations to quickly see and understand the potential of raw data in Hadoop, easily transform the data to make it better, and intuitively discover and share new value-all within a single visual product. Oracle Big Data Discovery offers tremendous speed at massive scale, streamlining Big Data analytics to unlock new value for everyone.



Figure 1. Oracle Big Data Discovery

## A Game Changing Product for Big Data Discovery

Hadoop continues to gain momentum, with organizations increasingly using it to store the vast, uncurated data that exists beyond the enterprise data warehouse. While collecting and storing Big Data are necessarily the first steps in using it for analytics, they are for many organizations the only steps they can take. Existing approaches to analytics simply cannot accommodate Big Data which, by its nature, is very new, extremely diverse and of varying quality, constantly changing, and largely unfamiliar. This presents organizations with tough choices: hire highly specialized resources who can use complex and unintegrated tools to make sense of Big Data; or stitch together a series of emerging point solutions to try to get a picture of what's in the data and what value it might offer.

Clearly what's needed is a holistic approach to Big Data that offers organizations an easy, intuitive way to see what they've collected in Hadoop and quickly understand its potential; work with the data visually and dynamically without switching tools;



Get the data ready for analytics via Intuitive, user driven data wrangling

- Leverage an extensive library of data transformations and enrichments
- · Preview results, undo, commit and replay transforms
- Test on sample data in memory then apply to full data set in Hadoop



### 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

## Share

- Share projects, bookmarks and snapshots with others
- Build galleries and tell Big Data stories
- · Collaborate and iterate as a team
- Publish blended data to HDFS for leverage in other tools

### **KEY BENEFITS**

- Quickly understand Big Data potential and know where to start
- Be able to forecast and justify investment
- Immediately know when not to proceed
- Transform and enrich data on a massive scale, to easily improve data for everyone
- Spend 80% of the effort on analytics and insights, instead of data prep
- Expand Big Data teams to include business analysts and drive collective discovery
- Process data in place in Hadoop, substantially reducing data movement and management costs
- Deploy quickly, in existing or new environments, to get started right away
- Remove technical barriers by integrating with your existing infrastructure and other Big Data tools
- Implement as an Engineered System or on commodity hardware, and easily expand to accommodate growth

and engage with their data via interactive discovery, moving rapidly to create visualizations and share insights with colleagues, allowing companies to leverage more of their analytic talent and driving innovation.

Oracle Big Data Discovery is expressly designed to address these needs, making Big Data more accessible to everyone across the organization, decreasing the risks currently associated with Big Data projects, and speeding time to value.

## See the Potential in Big Data

Justifying investment is difficult without some sense of the return—but without knowing what's in the data, it's hard to estimate potential value. Oracle Big Data Discovery solves this problem by providing rapid visual access to all the data in Hadoop, so organizations can:

- Find relevant data quickly, through a rich interactive catalog of the raw data in Hadoop
- · Load local data from Excel and CSV files through self-service wizards
- View data set summaries, annotations from other users, and recommendations for related data sets
- · Explore the data through familiar search and guided navigation

ORACLE Big Data Disc	overy					(Search Q	0	0
	6 Projects	158 Data Se	ts + Data Set					
fine By	Recently Viewed Data Sets							
USAGE	and an extension of the second se		Concernence of					
reated By Me	twitterTest	hiten test ds2	Bill Cities		wine	aw-classic		
ONTENT		Teger test						
antains Dates	Data Source: TW/TTERPOST (1).xts	Data Bource: ats: delauit, small sisk	Data Source: cities15000.cev		arrent according to the	and a second second second		
Contains Locations	Provider A Ameri	Peries A Long	Preview	4 100	Preview	Panlara (0.10	-	
Plan disabati 👻		Contrast Contrast Contrast		Re av				
lso								
nue	Moet Popular Data Sets View More							
IETADATA	an classic	wine	Bill Cities		Claime	biton tost de2		
toject Author	un ciudaic	HILL	Din Onioa		Chairna	THEORY CORE		
Data Set Author						Tagsi tugi		
Project Tags	(80855 records) (5 @	(57078 recorda) (5)	(40254 mccrch)		(2000200 raconta)	(10 moords) (1) (0		
aata ber raga	Prevlew	Papiew	Previou	A	Populaw 🎄 🛄	Patrian 4		
ast Modified								
survivor of Attributure	twitterTest	UK Road Accidents 2013	Warranty Claims		11attr	1attrzhon		
27								
	Data Source: TW/TTERPOST (1) xts	Data Source: accidents/2013.cev	Data Source: Claims.xisx		Dela Source: Tattorhonadax	Data Source: Latrationalise		
and the second s	(10001 records) (3)	(2433 records) (1)	(9913 seconds) () ()		(timeset)	(1 mcords)	-	
	PTDADA SE Broad	Piloven II Band	PTOVDE	M. Barrelly	Proven in Based	Printer St.		
ween V								
- 231	Newly Added Data Sets View More							
8 resulta Submit	twitterTest	hiten test ds2	good_values		UK Road Accidents 2013	Warranty Claims		
	Data Source: TWITTEPPOST (1).x8	Tagar: Lag1 Data Source: att_delaut_preal.xtox	(22 records) db		Data Source: accidente2013.cov	Deta Source: Cainsulas		
	Pressee	Prevan A	Preview	1.00	Preview	Preview	-	

Figure 2. Big Data Discovery's Interactive Catalog

Together with statistics about each individual attribute in any data set, these capabilities expose the shape of the data, empowering users to rapidly understand data quality, detect anomalies, uncover outliers, and ultimately determine potential. This enables organizations to:

- Visualize attributes by data type, to see at a glance which are the most relevant
- · Sort attributes by potential, so the most meaningful information is displayed first
- Use the scratch pad to uncover potential patterns and correlations between attributes



Figure 3. Exploring Data with Oracle Big Data Discovery

## Quickly Make Big Data Better

One of the challenges with analyzing data, at any scale, is that it's rarely ready to use as-is, and typically requires varying degrees of cleansing and transformation. Getting the greatest analytic value can also mean extending the data even further, extracting themes and sentiment, and blending data sets to expose new patterns. In today's Big Data landscape, transformation and enrichment are frequently handled upstream and in different tools, which can alter context and increase the time it takes to achieve meaningful results.

	Big Data Discovery						Search	۰ ۹	0	
re v Tr	ransform v Discover v	CI	aims Anal	lysis: Wa	rranty Claims				0 1	'n
ARRANTY C	LAIMS	hicledealergeocode_geo_country								
	Clear All US	5								
ransformati	ion Editor									
Jse refineme	ent state as a conditional statement					Functions	Attribute New Attribute N	ame		
1 ext	tractKeyPhrases(toLowerCase(comp	laint))					complaint_phro	3505		
		0					Data Type			
	Double click or drag to in	oorta Ŭ					Sound		v	
	function	JUL 1					Single Assig	n		
	Inneuon			-						
	O <sub>v</sub> Filter			9						
	abs(double)	Double-click or drag to inser	rt an							
	ii abs(float)	attribute								
	abs(int)	Show all V O, Filter					Cancel P	veview Add	I to Scri	iş,
_	ii abs(long)	c laimamount	# ^							
K of 9I	ii acos	claimdate	m	YPE V	NAME Y IN HIDDEN				-AL	
Jecoros a	I addTime/date integer string)								-	
	and introduced and for a new fit	c laimmonth	6.6					-	-	
\II Attrib	asin(double)	claimperiod	66				Sort. By name	· E	F	
( cl	asin(double)	claimperiod claimperiod	11 11 #		ii complaint	ii complaint_lang	Sot: By name	v E	mount	
ii d	esin(double)   etim(double)   etim(double)   etim(double, double)	clammonth claimperiod claimyear commotity	22 22 # 24		ii complaint	ii complaint_lang	Sot: By name	# labora	mount	1
All Attrib	asin(double)    atan(double)    atan(double)    atan2(double,double)    cardinality(object)	c barmon/th c barmon/th c barmon/th c barmon/th c barmon/th c commos/ty c complement	22 22 # 22	UTOMATI	II complaint AT 40 MPH THE CAR VIBRA	ii complaint_lang	Sot: By name	¥ iabora	mount	1
All Attrib (( cl 9 9	addr (double)     atar(double)     atar(double)     atar(double)     cardinata(object)     cbr(double)	clammonth claimperiod claimperiod claimpear commotity complaint complaint key phrases	23 24 25 25 25	UTOMATI	II complaint AT 40 MPH THE CAR VIBRA WHILE DRIVING, ML500 BE	ii complaint_lang en en	Sot: By name ii complaint_phrases car abs, service center, vehicle,	<ul> <li># labora</li> <li>117.12</li> <li>350.1034</li> </ul>	mount	1
All Attrib (( c) 9 9 10	Bain(double)   atim(double)   atim(double)   atim(double,double)   atim(double,double)   cardinality(object)   cardinality(object)   cardinality(object)	ctammonth ctampentod ctampentod complaint complaint complaint_lang_phrases	22 22 22 22 22 22 22	UTOMATI ITEM SINE COO	II complaint AT 40 MPH THE CAR VIBRA WHILE DRIVING, ML500 BE MISC CUSTOMER SATISFA	ff complaint_lang en en	Sot: By name if complaint_phrases car aba, service center, vehicle, satisfaction	<ul> <li># labora</li> <li>117.12</li> <li>350.1034</li> <li>146.63561</li> </ul>	mount	
9 9 10	sain(double)     stan(2(double)     stan(2(double))     cardinality(object)     cardinality(object)     cardinality(object)     cardinality(object)     cardinality(object)     cardinality(object)     cardinality(object)     cardinality(object)	clammonh clamperiod clamperiod clamperiod commoBly complaint_ley_phrases complaint_leng biblocremount	22 22 22 23 23 24 24 24 24 24 24 24	UTOMATI STEM SINE COO	II complaint At 40 MPH THE CAR VIBRA WHILE DRIVING, MLSOO BE MISC.CUSTOMER SATISFA MISC.CUSTOMER SATISFA MISC.CUSTOMER SATISFA	fi complaint_lang en en en en	Soc. By name ii complaint_phrases car abs, service center, vehicle, satisfaction satisfaction	<ul> <li># labora</li> <li>117.12</li> <li>350.1034</li> <li>146.63561</li> <li>188.90971</li> <li>188.90974</li> </ul>	mount 7	1
9 9 10 11 6	serie(double)   stan(double)   stan(double)   stan(double)   certic(double)   certic(double)   certic(double)   concat(string())   concat(string())	c stammonn c latampear c commotity c compant c compant_levg_thrases c compant_levg_thrases t compant_leng laboramount biotorates colon	22 22 42 22 22 22 22 22 22 22 22 22 22 2	UTOMATI ITEM DINE COO S. HYDRA S. HYDRA	II complaint AT 40 MPH THE CAR VIERA WHILE DRIVING, MLSOD BE MISC CUSTOMER SATISFA 2002 CHEVROLET SILVERA MISC CUSTOMER SATISFA.	ii complaint_lang en en en en	Car ado, service center, vehicle, satisfiction satisfiction consumer, brake catherity for	<ul> <li># labora</li> <li># labora</li> <li>117.12</li> <li>350.1034</li> <li>146.63561</li> <li>168.90971</li> <li>363.5478</li> <li>108.4003</li> </ul>	7 6	=
411 Attrib (( c) 9 9 10 11 6 8 11	i exert double i exert double i extra (double) i exert double) i exert	c atampend c tampend c tamped c company ar c company, ary physics c company, ary physics c company, and c compa	22 23 24 25 25 25 25 25 25 25 25 25 25 25 25 25	UTOMATI ITEM SINE COO S. HYDRA S. HYDRA	II complaint AT 40 MPH THE CAR VIBRA VH-LLG DRIVING ML500 BE MISC CUSTOMER SATISFA 2022 CHEVROLET SILVERA MISC CUSTOMER SATISFA MISC CUSTOMER SATISFA	ii complaint_lang en en en en en en	La Socie By none ii complaint_phrases car abs_service center, vehicle, satisfaction consume, brake satisfaction satisfaction	<ul> <li># labora</li> <li>117.12</li> <li>120.1034</li> <li>146.63561</li> <li>168.90971</li> <li>363.5478</li> <li>108.49303</li> <li>237.21550</li> </ul>	7 6 4	=
All Attrib (( c) 9 9 10 11 6 8 11 3	anividoolei     anividoolei     anividoolei     anividoolei     anividoolei     anividoolei     anividoolei     anividoolei     anividoolei     cantoliaduo     cantoliaduo     cantoliaduo     concatististing()     concatisting()	Lamonom Elamporo Catanyear compart compart_larg_transes compart_larg_transes compart_larg_transes tomporout biorotechplon maile maile	22 23 23 23 23 23 23 23 23 23 23 23 23 2	UTOMATI ITEM DINE COO S, HYDRA S, HYDRA B, HYDRA	II complaint AT 40 MPH THE CAR MBRA MISC CUSTOMER SATISFA MISC CUSTOMER SATISFA MISC CUSTOMER SATISFA MISC CUSTOMER SATISFA MISC CUSTOMER SATISFA MISC CUSTOMER SATISFA MISC CUSTOMER SATISFA	ii complaint_lang en en en en en en en en	ESot: By name if complaint_phrases car abs, service center, vehicle, astifaction astifaction consume; bale astifaction astifac	<ul> <li># labora</li> <li># labora</li> <li>117.12</li> <li>150.1034</li> <li>146.63561</li> <li>168.90971</li> <li>363.5478</li> <li>108.4303</li> <li>237.21550</li> <li>140.0316</li> </ul>	7 6 4 6	-
All Attrib (( c) 9 9 9 10 11 6 8 11 3 8	environment	Lammonth Lampered Lampered compact compact compact compact compact compact compact compact compact compact compact bock bock compact compact compact bock compact com	22 23 23 23 23 23 23 23 23 23 23 23 23 2	UTOMATI ITEM DINE COO S. HYDRA S. HYDRA S. HYDRA S. HYDRA	11 complete AT 40 MPH THE CAR VIBRA WHELE DRYNNE, NLEOD BL. MISC CUSTOMER SATISFA MISC CUSTOMER SATISFA MISC CUSTOMER SATISFA MISC CUSTOMER SATISFA DIT'THE CONTACT STATE DIT'THE CONTACT STATE	ii complaint_lang en en en en en en en en en	Esc: By nave If complaint; phrases and sample of the second second settle context while settle context on consume, the settle context settle	<ul> <li># labora</li> <li>117.12</li> <li>350.1034</li> <li>146.63561</li> <li>188.90971</li> <li>363.5478</li> <li>108.45030</li> <li>237.21550</li> <li>140.0316</li> <li>142.45598</li> </ul>	7 6 4 6	=
SI Attrib (i o 9 9 10 11 6 8 11 3 8 11	anardychala         anardychala           constantisting         anardychala           anardychala         anardychala           anardychala         anardychala	Lamonom Lamproof Lamproof Lamproof Lamproof Compact Compact Compact Compact Compact Lamproof Modescription mela mela mode Info Compact Com Com Com Com Com Com Com Com	22 23 44 23 23 23 23 23 23 23 23 24 24 24 25 25 25 25 25 25 25 25 25 25 25 25 25	UTOMATL TEM GINE COO S. HYDRA S. HYDRA S. HYDRA S. HYDRA	If complete At a MPH THE CAR VBRA. WHELE DRIVED, BLOD BE. USC CUSTOMER SATEFA. MISC CUSTOMER SATEFA. MISC CUSTOMER SATEFA. MISC CUSTOMER SATEFA. MISC CUSTOMER SATEFA. MISC CUSTOMER SATEFA.	if complaint_lang en en en en en en en en en en en	Est: By neve If complaint_phrases or auto-tool subtraction subtraction subtraction subtraction subtraction subtraction subtraction subtraction subtraction subtraction subtraction subtraction	<ul> <li># labora</li> <li># labora</li> <li>117.12</li> <li>350.1034</li> <li>146.63561</li> <li>168.90971</li> <li>363.5478</li> <li>108.46003</li> <li>237.21556</li> <li>140.035646</li> <li>142.45596</li> <li>240.55546</li> </ul>	7 6 4 6 7	-
SI Attrib (i c) 9 9 10 111 6 8 111 3 8 11 11	anticodation         anticodation	Lemmonth Lempord Lempord Lempord Commonly Compart Compart Compart Compart Compart Modesee Modesee mail mature mail mature	22 23 23 23 23 23 23 23 23 23 23 23 23 2	UTOMATI ITEM 3NE COO S. HYDRA S. HYDRA S. HYDRA S. HYDRA	If complete AT 40 MPH THE CAR VERA WHELE DRIVING, ALLOD BE SIGC CUSTOMER SATESA MICC CUSTOMER SATESA MICC CUSTOMER SATESA MICC CUSTOMER SATESA MICC CUSTOMER SATESA MICC CUSTOMER SATESA WARNING LIDHTS CHIER	ii complaint_lang en en en en en en en en en en en en en	Esc: By neve If complaint_phrases cer ada, tenis c onter, velas b, satisficcion satisficcion satisficcion satisficcion satisficcion satisficcion satisficcion gista, generator, saming satisficcion	<ul> <li># Inbora</li> <li># Inbora</li> <li>177.12</li> <li>170.1024</li> <li>146.63561</li> <li>146.63561</li> <li>146.63561</li> <li>142.45595</li> <li>140.0316</li> <li>142.45595</li> <li>240.65564</li> <li>281.05456</li> </ul>	7 6 4 6 7 2	
9 9 10 11 6 8 11 3 8 11 11 5	matrix/status/	Lemonom Lemonom Lemonom Lemonom Lemonom Lemonom Compare Compare Compare Compare Compare Compare Lindonscreption mails Lindonscreption Lindonscrept	22 22 22 22 22 22 22 22 22 22 22 22 22	UTOMATL ITEM DIRE COO 5, HYDRA 5, HYDRA 5, HYDRA 5, HYDRA 5, HYDRA 1, HYDRA	If completent AT 40 MPH THE CAR MBRA. WHELE DRIVING, MLDO BE, MISC OUSTONER SATISFA. MISC OUSTONER SATISFA. MISC OUSTONER SATISFA. DIT THE CONTACT STATE. MISC OUSTONER SATISFA. DIT THE CONTACT STATE. MISC OUSTONER SATISFA. MISC OUSTONER SATISFA.	11 complaint_tang	Est: By name If compilant_shrase Cr abs, service centry, vehicle, autofaction consume, faule con	<ul> <li># International State</li> <li># International State</li> <li>117.12</li> <li>150.1034</li> <li>146.63561</li> <li>168.00071</li> <li>363.5478</li> <li>108.4500</li> <li>237.21550</li> <li>140.0316</li> <li>142.45500</li> <li>240.55546</li> <li>241.55546</li> <li>251.0548</li> <li>159.0148</li> </ul>	7 6 4 6 8 7 2	1

Figure 4. Transforming Data with Oracle Big Data Discovery

Not so with Big Data Discovery. Data transformation and enrichment are handled natively within the product, in an intuitive and interactive visual interface, which leverages the power of Apache Spark behind the scenes to transform massive amounts of data at scale, ensuring no loss of context. Users from data scientists to business analysts can:

- Wrangle data in place in Hadoop via an intuitive, spreadsheet-style interface
- Leverage an extensive library of common data transformations, such as split, merge, group, or replace values, and many more
- Enrich to infer language or generate geographic hierarchies from address fields, as well as extracting sentiment and themes automatically from raw text
- Test transformations on sample data in memory before applying them to the full data set in Hadoop
- · Preview results, and undo or replay transformations before explicitly committing them

## Unlock Big Data Discovery for Everyone

Importantly, Big Data Discovery opens up the discovery process to business analysts as well as data scientists, allowing everyone on the Big Data team to spend much less time on preparation and much more time on analysis. Big Data Discovery accelerates the analytic process by streamlining the initial phases of understanding, transforming, and enriching Big Data, leveraging best practices for interactive visualization and data discovery. This means everyone on the extended Big Data team can:

- Ask questions of the data and get answers as easily as shopping online, using familiar world-class search and guided navigation
- Drag and drop to create and combine rich, interactive visualizations and discovery dashboards
- Broaden perspectives or change the course of investigation by blending new data sets into an analysis, updating visualizations and dashboards in the moment
- Build galleries of snapshots that tell Big Data stories, and drive collective discovery by easily sharing galleries, bookmarks, and entire projects across teams
- Extend the reach and value of the data by publishing blended data back to Hadoop Distributed File System (HDFS), for use across the enterprise



Figure 5. Discover and Share New Insights with Oracle Big Data Discovery

## **BIG DATA DISCOVERY**

Big Data Discovery is a member of the Oracle Big Data Analytics product suite which, together with Oracle's other Big Data solutions, offers customers the industry's most comprehensive Big Data platform.

#### RELATED PRODUCTS

The following products complement Oracle Big Data Discovery:

- Oracle Big Data Appliance
- Oracle Big Data SQL
- Oracle R Advanced Analytics for Hadoop

## **Technical Innovation on Hadoop**

Oracle Big Data Discovery offers true technical innovation on Hadoop, natively leveraging the power of distributed storage and computing, across servers (or nodes), to process massive amounts of information without having to move it around.



Figure 6. Oracle Big Data Discovery Technical Innovation

- Studio is the unique web-based user interface that makes it easy for anyone to find, explore, transform, discover and share data.
- Dgraph server is the hybrid search-analytic database that allows users to operate on in-memory data sets for interactive performance.
- Data processing uses Apache Spark to profile, sample, transform and enrich massive amounts of information across all the data nodes in the Hadoop cluster.

## Oracle's Unified Big Data Management and Analytics Solutions

Big Data Discovery is a core component in Oracle's overall Big Data management and analytics strategy, enabling customers to:

- Use Oracle R Advanced Analytics for Hadoop, for better predictive analytics
- Leverage Oracle Big Data SQL to query the data in HDFS without moving it at all
- Implement solutions on Oracle engineered systems, enabling rapid application deployment, optimized performance benefits and lower total cost of ownership



#### CONTACT US

For more information about Oracle Big Data Discovery, visit oracle.com or call +1.800.ORACLE1 to speak to an Oracle representative.

#### Hardware and Software, Engineered to Work Together

Copyright © 2014, Oracle and/or its affiliates. All rights reserved. This document is provided for information purposes only, and the contents hereof are subject to change without notice. This document is not warranted to be error-free, nor subject to any other warranties or conditions, whether expressed orally or implied in law, including implied warranties and conditions of merchantability or fitness for a particular purpose. We specifically disclaim any liability with respect to this document, and no contractual obligations are formed either directly or indirectly by this document. This document may not be reproduced or transmitted in any form or by any means, electronic or mechanical, for any purpose, without our prior written permission.

Oracle and Java are registered trademarks of Oracle and/or its affiliates. Other names may be trademarks of their respective owners.

Intel and Intel Xeon are trademarks or registered trademarks of Intel Corporation. All SPARC trademarks are used under license and are trademarks or registered trademarks of SPARC International, Inc. AMD, Opteron, the AMD logo, and the AMD Opteron logo are trademarks or registered trademarks of Advanced Micro Devices. UNIX is a registered trademark of The Open Group. 0115



CONNECT WITH US

blogs.oracle.com/oracle D

facebook.com/oracle

twitter.com/oracle

oracle.com