ORACLE DATA SHEET

Oracle Big Data Discovery
The Visual Face of Hadoop

ORACLE BIG DATA

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.

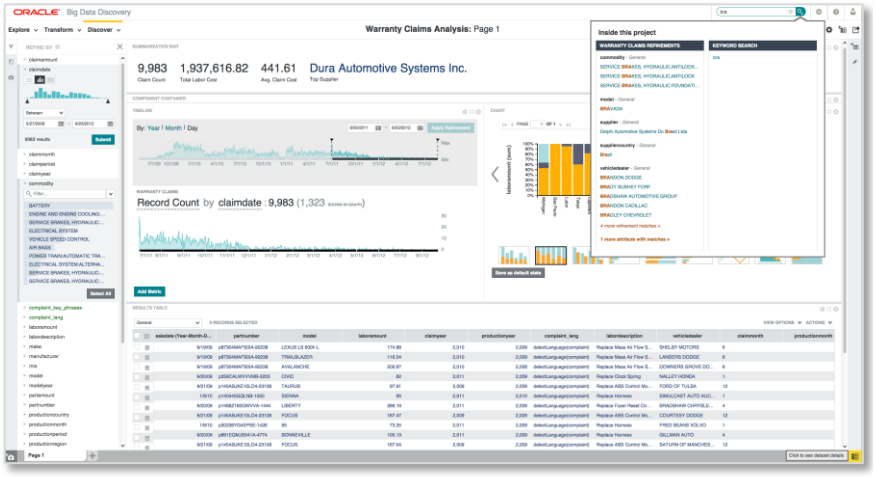

Figure 1. Oracle Big Data Discovery

BIG DATA DISCOVERY KEY FEATURES

Find

- Access a rich, interactive catalog of all data in Hadoop
- 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

Explore

- Visualize all attributes by type
- Sort attributes by information potential
- Assess attribute statistics, data quality and outliers
- Use a scratch pad to uncover correlations between attributes

Transform

- 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

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;

ORACLE

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

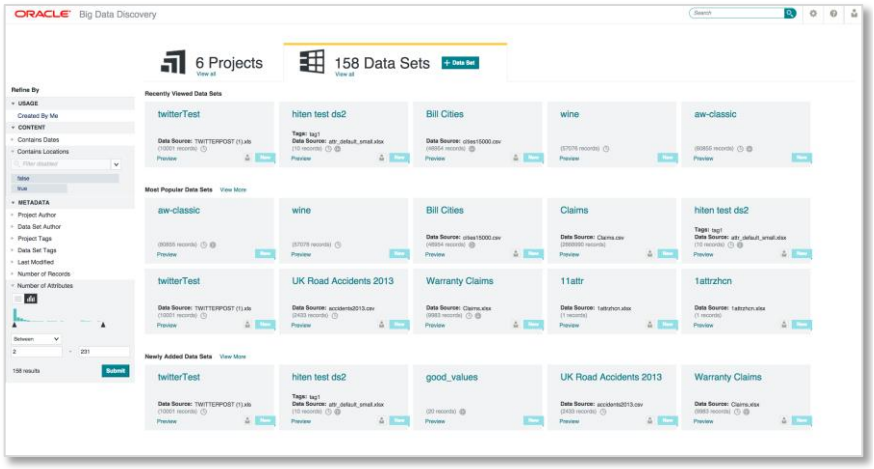

Figure 2. Big Data Discovery's Interactive Catalog

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

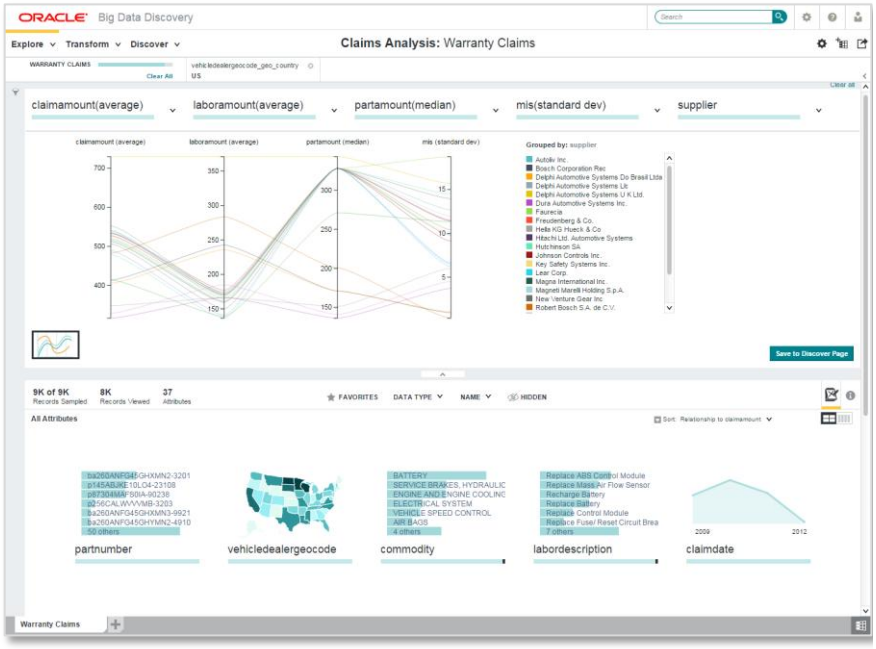

Figure 3. Exploring Data with Oracle Big Data Discovery

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.

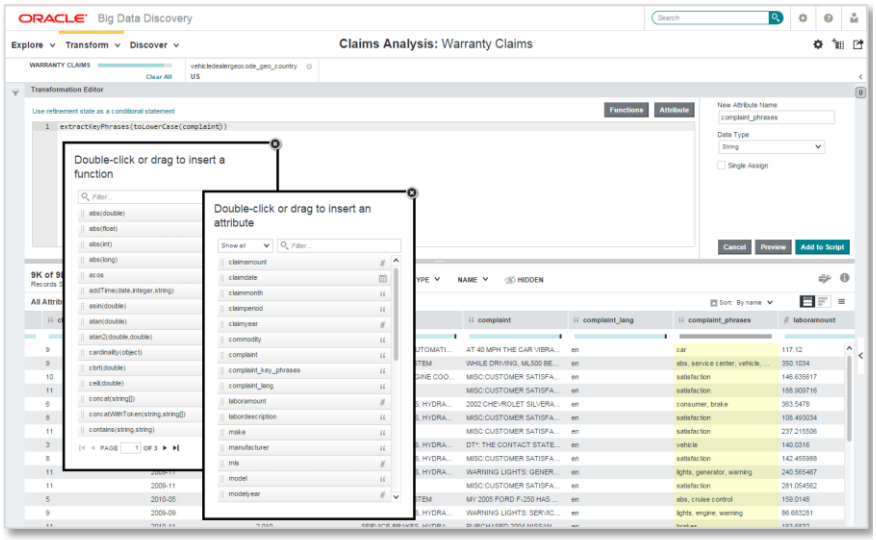

Figure 4. Transforming Data with Oracle Big Data Discovery

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

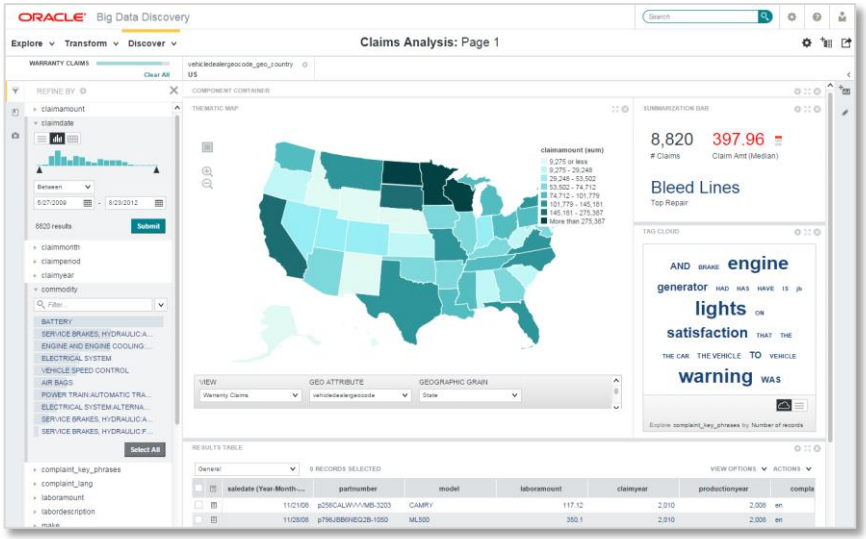

Figure 5. Discover and Share New Insights with Oracle Big Data Discovery

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.

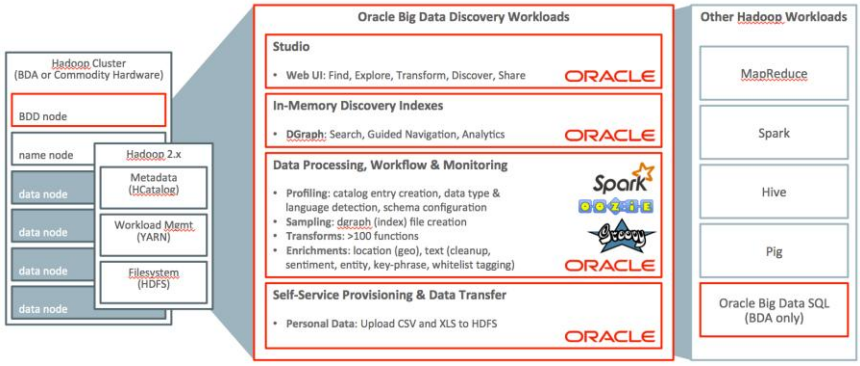

Figure 6. Oracle Big Data Discovery Technical Innovation

- 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

ORACLE

CONTACT US

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

CONNECT WITH US

blogs.oracle.com/oracle

facebook.com/oracle

twitter.com/oracle

oracle.com

Hardware and Software, Engineered to Work Together