Oracle White Paper
2017

A Unified Approach to Data Processing and Analytics

Big Data, Fast Data, All Data

National Security Industry Focus

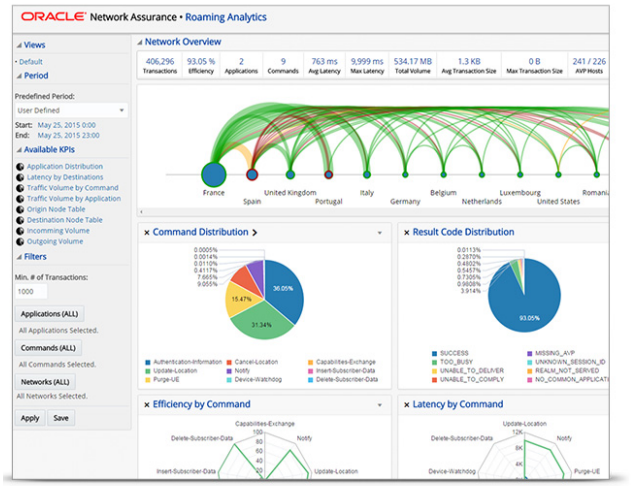
ORACLE

Disclaimer

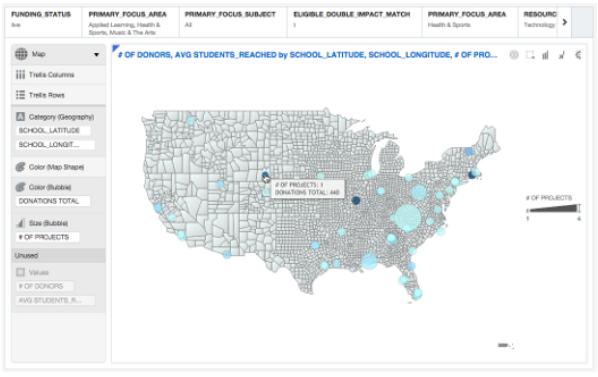
The following is intended to outline our general product direction. It is intended for information purposes only, and may not be incorporated into any contract. It is not a commitment to deliver any material, code, or functionality, and should not be relied upon in making purchasing decisions. The development, release, and timing of any features or functionality described for Oracle's products remains at the sole discretion of Oracle.

Executive Overview

Data Processing and Advanced Analytics is the foundation to producing good intelligence. However, analytics means many things to many people. Advanced analytics utilizes data of different types, from different sources and applies precise algorithmic processing. Valued intelligence results from the timely correlations and insights amongst this data, the algorithm results, and the inter-relationships that exist from different data sources.



The challenge to producing mission results in data processing and analytics comes from numerous areas. Frequently, different data types are managed in different data stores. This causes information fragmentation and hinders analysis. Additionally, there are inefficiencies in the rigorous ETL (Extract, Transform, Load) and integration required to do analysis. Many believe 80% of the time and cost spent is just preparing the data for analysis. Likewise, the pugh and pull of the developer community between NoSQL, SQL, Schema, Schema-less, Hadoop, No-Hadoop has created a pile of failed programs with limited analytical results. All these issues result in significant delays in mission execution that is unnecessary given the advancements with today's technology.



To overcome what we see our enterprise customer face on a daily basis, Oracle and our thousands of developers have created an efficient and robust standards-based platform that addresses the practical needs of the analytic enterprise. Oracle has addressed major issues like multi-data type support in the same data engine, easier data wrangling, automation, machine learning and unification of Hadoop, Relational and InMemory eco-systems. Our technology supports rapid, low cost, iterative and adaptive analytics. By empowering analysts with a self-service platform, users can perform rapid tests and evaluations on the data with current analytical methods.

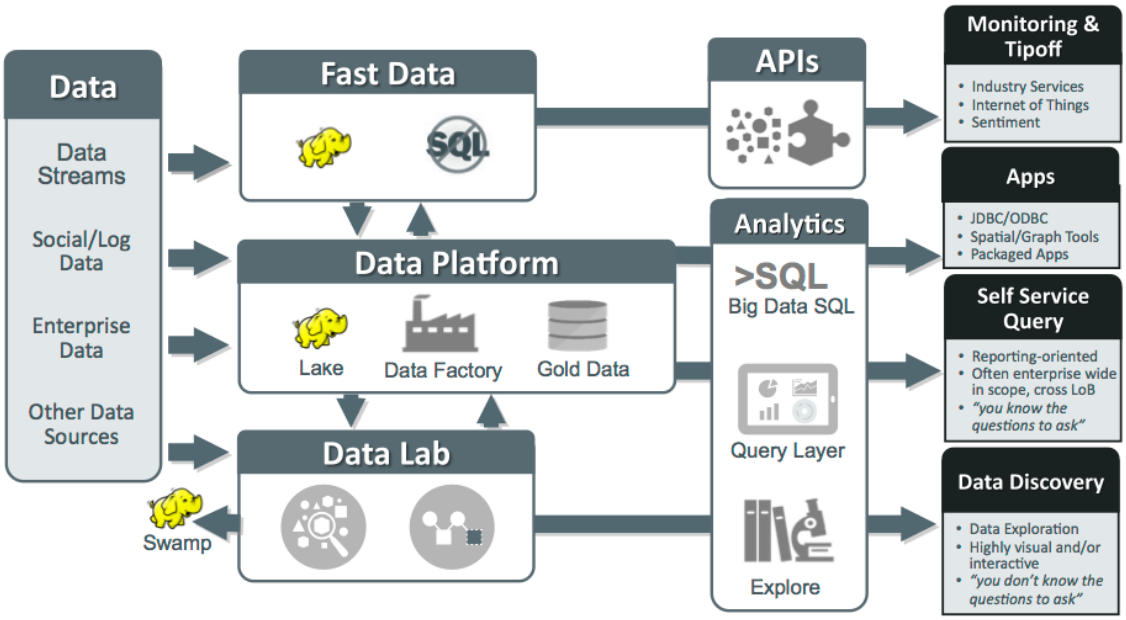
Oracle's technology provides a solution that is a fully complete analytic environment that supports full-spectrum data ingest, wrangling, data exploration & discovery through advanced and predictive analytics. It represents the combination of software, cloud computing and/or supporting hardware that has been professionally engineered, optimized, developed and deployed to support the analytic challenges faced today.

A key differentiated objective is to empower analysts to explore, test, and evaluate in a self-service fashion thus reducing the need for costly programmers and data scientists.

Unfortunately, the overloaded use of the word "analytics" creates not only ambiguity in conversation but also incomplete and inefficient solution architectures. The reason has to do with the fact that the analytical tools used to perform the above processes vary as widely as the definition of analytics itself. The reality is that the different tools and different ways we store and manage data for analytics creates impedance in doing higher level, result-oriented advanced analytics.

Oracle's Unified Data Processing and Analytic Platform

Oracle has created a holistic, standards-based and unified approach to provide integrated analysis for all data types, analytic methods and user classes.



This figure is an excellent roadmap to understand the landscape of data processing and analytics. Each area is explained in more detail.

DATA

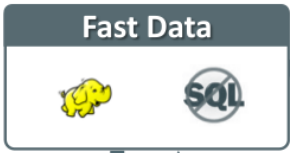
In a modern data centric enterprise you may have "all of the above" with regards to data sources. Data streams denote fast moving data (sensors, internet of things, message traffic, GPS, telemetry, ephemeris, financial transactions).

Social and log data is often high volume data with semi-structure that generally is not professionally managed or archived for long periods of time. In essence, the data has a time allotment when it decreases in value, therefore professional management and administration of the data is not an efficient investment. This data however is valuable for analytical purposes and spotting sentiment, trends, or anomalies.

Enterprise data is known valuable data. This data source is understood and the format of the data is known. It can be transactional, metadata, textual (email), message oriented, database exports, files (spreadsheets), XML, JSON, or via a middleware bus. This data in most cases is all valuable; kept as a record, secured, available, and used immediately for analytical dashboards, query tools, and processes. A few examples of this data might be: intelligence reports, intel tasking, targeting, imagery/video metadata, collection accomplishments, signals metadata, spatial data, networks of individuals, watch-lists, etc.

Other Data Sources includes data that varies in variety and is often new data sets or unknown data. This source of data is often very textual such as documents, blogs, or web pages. This data is often pre-processed in the data factory for indexing or to derive structured data.

FAST DATA

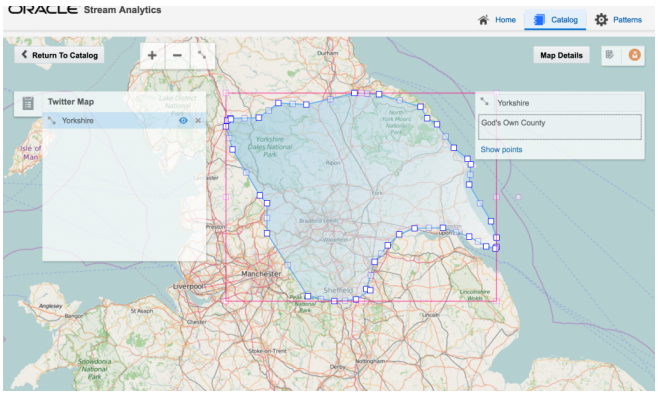


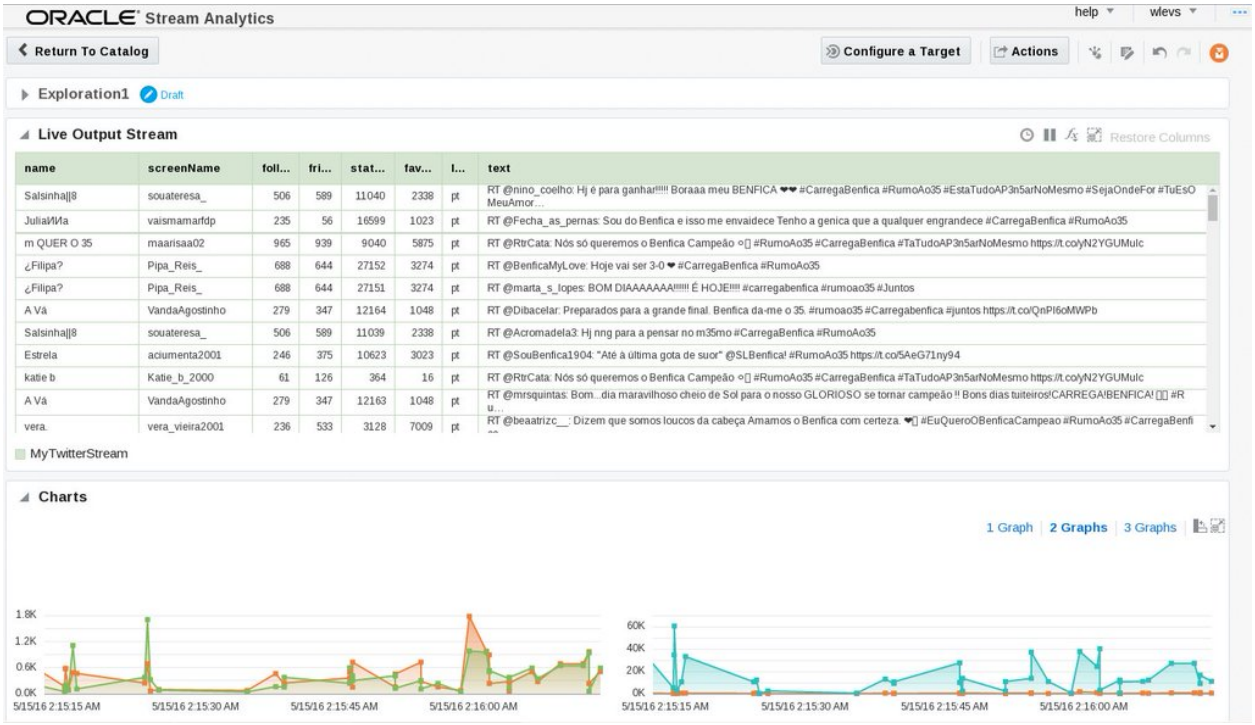
With exploding intelligence data from increased number of connected devices, cyber sensors, collection platforms and social networks, there is an increase in the volumes and speed of dynamically changing data. High-velocity data (Fast Data) brings high value, especially to national security decision-making processes. However, some of this data loses its operational value in a short time frame. Big Data allows the luxury of time in processing for actionable insight. Fast Data, on the other hand, requires extracting the maximum value from highly dynamic and strategic intelligence. It requires processing much faster and facilitates taking timely action as close to the generated data as possible. Fast Data can also be a means of getting early answers or tip-offs of what's in your big data holdings. Industry and

government needs have led to huge internal research investments at Oracle. Oracle has responded with a robust platform for Fast Data and Stream Analytics.

Oracle Stream Analytics

The Oracle Stream Analytics platform provides a compelling combination of: an easy-to-use visual façade to rapidly create and modify Event Stream Processing applications, an advanced streaming analytical capabilities, and a comprehensive, flexible and diverse runtime platform that uses Apache Spark, Kafka and SOA streaming technology.





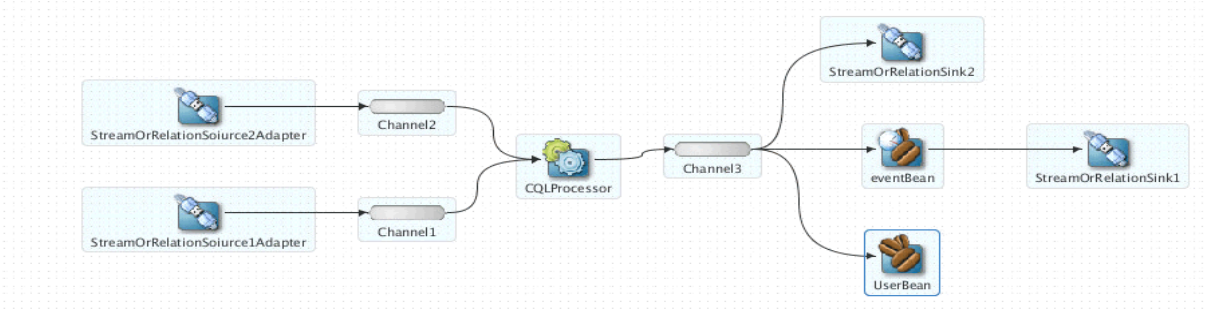
Oracle Stream Analytics provides a pre-developed library of algorithms and detection features to leverage spatial, statistical, machine learning and well-known patterns. The abstraction and

definition of streams and targets allow immediate joining of streaming data to gold copy look-up databases, and explorations that provide a stunning visual representation of real time event data.

The user has the ability to rapidly create stream exploration and analysis without writing custom code in Apache Storm. In addition to real-time event sourcing, the Oracle Stream Analytics design environment supports standards-based, **continuous query execution** across both event streams and persistent data stores. This enables the platform to act as the heart of intelligence for systems needing answers in microseconds to discern patterns and trends that would otherwise go unnoticed. Event processing use cases require the speed of in-memory processing with the mathematical accuracy and reliability of standard database SQL. This platform listens to incoming event streams and executes registered queries continuously, in-memory on each event, utilizing advanced, automated algorithms for query optimization. Examples of this type of detection:

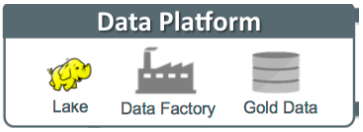
- Correlated events: If event A happens, event B almost always follows within 2 seconds of it.
- Missing or Out-of-Sequence events: Events A, B, C should occur in order. C is seen immediately after A, without B would be an alert.
- Spatial/Temporal events: Target of a particular type has passed through a geo-boundary at a particular time while an attribute about the target registers positive in a watch list database.

The Oracle Stream Analytics platform also allows for both SQL and Java code to be combined to



deliver robust event processing applications, leveraging standard industry terminology to describe event sources, processes, and event output or syncs. The platform provides a meta-data driven approach to defining and manipulating events within an application. Developers and analysts use a visual, directed-graph canvas and palette for application design to quickly outline the flow of events and processing across both streams and data sources. Developing the flow through drag and drop modeling and configuration wizards, the analyst can then enter the appropriate metadata definitions to connect design to implementation.

DATA PLATFORM



The data platform is the layer of the data management and analytic architecture where data is staged, transformed, secured and managed. The platform must provide the scalability, security and high availability that enterprises require. Oracle has focused significant R&D and innovation in this area because we realize that it is critical to all other aspects of analytics, cyber protection and mission outcomes. Oracle believes that the data platform is a unified environment, which may include an HDFS data lake, SQL and NoSQL databases, with data transformation, data cleansing, query, and movement functions that we call the factory. All of these capabilities work in unison to bring a modern data platform to the enterprise. Where data resides, where it's staged, where it is stored, where it is managed within the platform is really dependent on numerous factors including: type, security, analysis, governance, compliance, cost and processing method. Choosing one data method, like SQL or NoSQL, for every use case is not wise. Likewise, using Hadoop/HDFS for everything is not wise. All three serve specific purposes. Oracle has invested significantly making this data platform unified. Below outlines the technologies that make up the data platform eco-system.

Gold Data Using ORACLE 12c

The gold data concept is often referred to as "gold" because of the value. Gold data is actionable. It is data that has been cleansed and is as accurate as possible. The gold data environment must have several critical features:

encryption, security, high availability, survivability and transactional integrity and high performance. The platform must also be able to handle all data types such as: spatial, graph, xml, json/key value, video, imagery and relational.

All too often enterprises put in place separate databases for every data type. This causes analytical, management and security turmoil because there are separate databases for graph, spatial, relational, xml, JSON etc. Oracle 12c is a multi-model database. It is capable of providing gold data services for all of these data types in the same database. This improves analytics by creating cohesive result sets, improves security by securing all of the data together and significantly reduces the cost of all the software, infrastructure and labor compared to a multiple database gold environment.

Oracle NoSQL Database

The Oracle NoSQL Database provides network-accessible multi-terabyte distributed key/value pair storage with predictable latency. Data is stored in a very flexible key-value format, where the key consists of the combination of a major and minor key (represented as a string) and an associated value (represented as a JSON data format or opaque set of bytes). It offers full Create, Read, Update and Delete (CRUD) operations, with adjustable durability and consistency guarantees. It also provides a powerful and flexible transactional model (with ACID) that eases application development.

The Oracle NoSQL Database is designed to be a highly available and extreme scalable system, with predictable levels of throughput and latency, while requiring minimal administrative interaction.

It is also network topology and latency aware. The database driver working in conjunction with highly scalable, fault tolerant, high throughput storage engine enables a more granular distribution of resources and processing, which reduces the incidence of hot spots and provides greater performance on commodity-based hardware.

Big Data SQL

Big data's great potential has been touted for years. For many organizations, however, the challenge to making their big data vision a reality is that their ability

to collect data has gotten far ahead of their ability to use it.

One of the hurdles is that an enterprise's data is typically scattered across departments, systems, and regions, and much of what is collected is of low or unknown value. Vast quantities of this data are now stored in Apache Hadoop, which makes it possible to store and process huge amounts of data at low cost. Other data is housed in relational and NoSQL databases.

The holy grail of a big data strategy is to be able to easily leverage the information in all of these data stores to get a deeper, richer, and far more valuable view of customers, business processes, and opportunities. But Hadoop programmers are in short supply, so many organizations don't have the skills they need to leverage Hadoop data. And it's not feasible to move all of that data to a single data store; the cost and the security issues are prohibitive.

Oracle Big Data SQL enables organizations to get the utmost value from data by providing queries to big data systems integrated with existing enterprise information architectures. You can then quickly leverage big data into reports or applications using existing interfaces.

In addition to simplified access and integration, Oracle Big Data SQL uses some of the proven high-performance technology of Oracle Exadata and the industry-leading security features of Oracle Database to provide super fast speed and enterprise-class security across all of your data stores.

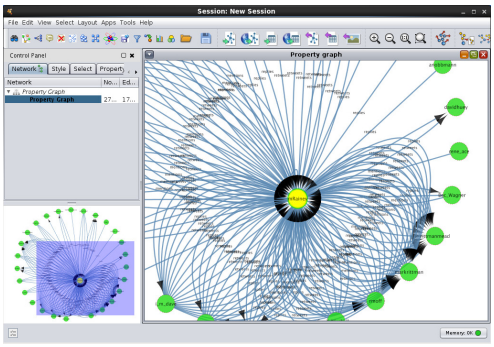
Leveraging Oracle Exadata Smart Scan technology, Oracle's Smart Scan on Hadoop processes SQL queries at the Hadoop storage level where data is located, scans the data, and brings back only the relevant data to the end user. Less data moved means faster results—and speed is critical when leveraging big data for real-time innovations in national security processes.

On the security front, Oracle Big Data SQL extends the advanced security capabilities of the Oracle database to Hadoop and NoSQL data. With Oracle Big Data SQL, you can take advantage of proven Oracle Database security solutions for data redaction and privilege analysis, with strong controls that limit privileged user access to data.

Oracle Big Data Spatial and Graph

Oracle Big Data Spatial and Graph offers a set of analytic services and data models that support Big Data workloads on Apache Hadoop, Oracle NoSQL Database, and Spark technologies. For over a decade, Oracle has offered leading spatial and graph analytic technology for the Oracle

Database. Oracle is now applying this expertise to work with social network data and to exploit Big Data architectures.



Oracle Big Data Spatial and Graph has three components: a distributed property graph with 40 high-performance, parallel, in-memory analytic functions; a wide range of spatial analysis functions and services to evaluate data based on how near or far something is to one another, whether something falls within a boundary or region, or to process and visualize geospatial map data and imagery; and, a multimedia framework for processing video and image data in Apache Hadoop, such as facial recognition.

Using the graph features, analysts can discover relationships and connections among targets, organizations, and assets. With the spatial capabilities, users can achieve insight into location-based patterns and trends across big data volumes, harnessing inherent location relationships in disparate data sources through harmonization and enrichment.

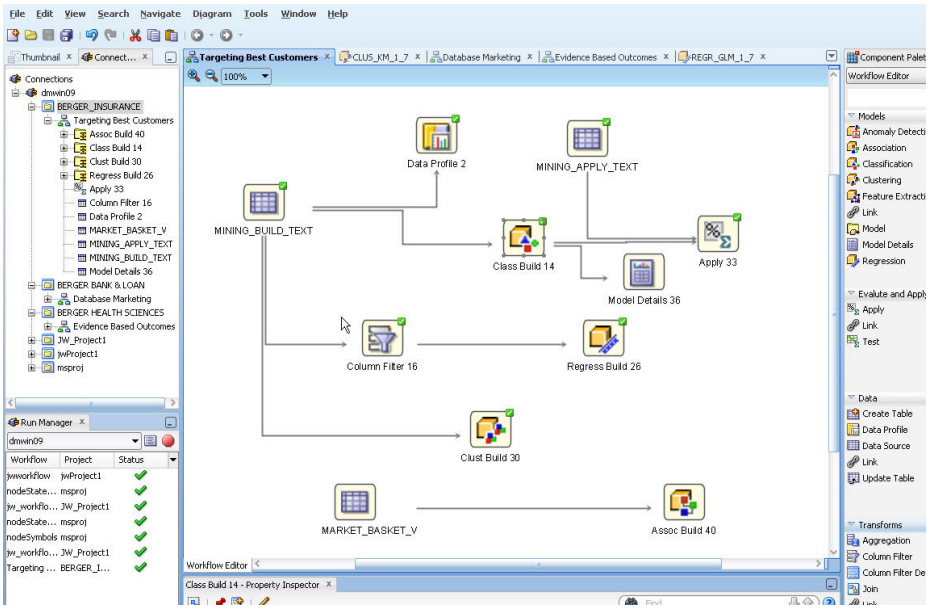
Oracle Advanced Analytics and Machine Learning

Oracle provides several enabling technologies for data analytics, statistical analysis, time-series analysis, modeling, and machine learning requirements. These technologies can actually be used in both the product data platform or the data lab. Traditional advanced analytics are inherently weak at information technology management such as:

- data extracts and data movement
- data duplication resulting in no single-source of truth
- data security exposures
- multiple analytical tools (commercial and open source) and languages (SAS, R, SQL, Python, SPSS, etc.)

Problems become particularly egregious during a deployment phase when the worlds of data analysis and information management collide.

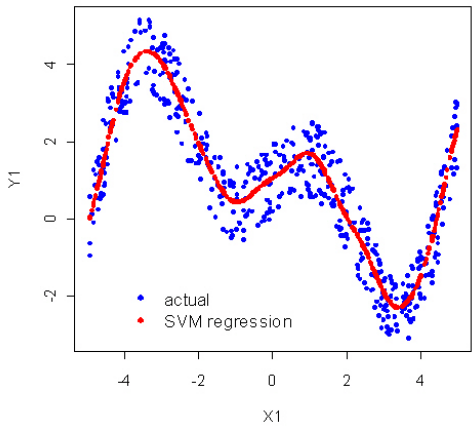
Oracle delivers an analytics platform and visualization that eliminates the traditional extract, move, load, analyze, export, paradigm when wanting to apply an advance algorithm on data. The Oracle Advanced Analytics Option in Oracle 12c and the Advanced Analytics for Big Data options perform analytic functions on data where it resides. Data remains in the database or HDFS cluster, which reduces data movement performance impacts, promotes data security control mechanisms, and provides greater performance and scalability.



Oracle Advanced Analytics Option extends the database into a comprehensive advanced analytics platform for data mining and data analytics. It delivers scalable, parallelized, in-database implementations of 20+ analytics algorithms (e.g., clustering, regression, prediction, associations, text mining, associations analysis, anomaly detection, etc.) as SQL functions within the Oracle Database 12c. Oracle Advanced Analytics exposes these predictive algorithms as SQL functions accessible via Oracle Data Miner, the Oracle Data Miner "drag and drop" workflow GUI, an extension to Oracle SQL Developer, and through tight integration with open source R (Oracle R Enterprise).

Oracle Advanced Analytics provides a mechanism not only to study the data through statistical means, and put those statistical and machine leaning modules into production. The ease of migration from the data study/analysis phase, into an enterprise data production environment using these models is a key differentiator our technology has over other competitors. This allows our customers to quickly modify or employ multiple machine learning techniques based on business or mission need.

Enterprise R



Oracle R Enterprise (ORE) integrates the open source R language for statistical computing and graphics within Oracle Database 12c to perform deep time series statistical data analysis and advanced analytics, as well as generating sophisticated graphics. The Oracle R Distribution is the first and only supported commercial distribution of R from an established company, and it extends the database's analytical capabilities by leveraging R's library of statistical functions and pushes down computations to the database instead of fetching all the data to the client and processing data there.

ORE eliminates R's typical memory constraint on client workstations by pushing the R computations to the enterprise database server and enabling R algorithms to work directly, transparently, and with a high degree of parallelism on the database server where the data lives. Only the results are transmitted to the client for visualization. Users are able to leverage the latest R algorithms, Comprehensive R Archive Network (CRAN) packages, and Data Visualization dashboards to send time series computations to the Oracle Database server. There are 100s of open source CRAN time-series algorithms relevant to DoD/IC mission. Examples include: Applied Statistical Time Series Analysis (astsa), Distance Covariance and Correlation for Time Series Analysis (dCovTS), Time Series Analysis Toolkit Based on Symbolic Aggregate Discretization, i.e., SAX (jmotif), Titration analysis for mass spectrometry data (titan), Peak Picking for High Resolution Mass Spectrometry Data (enviPick), Isotope Pattern, Profile and Centroid Calculation for Mass Spectrometry (enviPat), Classes and Methods for Seismic Data Analysis (IRISSeismic), Seismic Time Series Analysis Tools (RSEIS), Visualization for Seismic Tomography (RTOMO), and Bayesian Analysis of a Vector Autoregressive Model with Stochastic Volatility and Time-Varying Parameters (bvarsv).

Machine Learning

Machine Learning is the ability to automatically sift through large amounts of data to create models that find previously hidden patterns, discover valuable insights, and make predictions through the

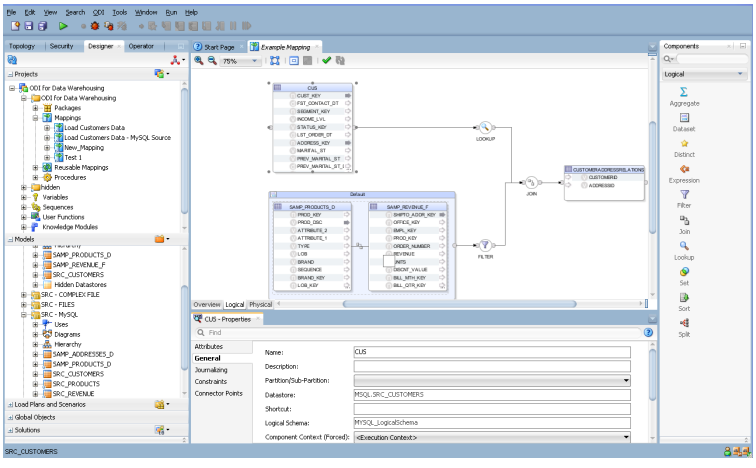
use of analytics models. These capabilities are important to multiple commercial industries including; banking, financial, retails, and the DoD. The ability to predict fraud on-line is a key element for these sectors, and these models are based on past behavior and methods to allow the model to predict and the system to take alternative measures with the transaction.

Oracle Advanced Analytics and Enterprise R technology allow the data scientist to review the data via built in statistical algorithms and extend R algorithms to determine key data patterns that define the behavior or characteristic under investigation. These models can then be tested and validated, and placed into production to run against real-time data coming into the system and provide NRT prediction on the system behavior and/or characteristic under investigation.

THE DATA FACTORY

The Data Factory in the eco-system serves the role of transformation, integration, enrichment, cleansing, movement and governance. These tasks are often referred to as data wrangling. It is very common for an enterprise to spend 80% of their time wrangling and 20% analyzing data. Oracle's focus has been to provide technology that expedites and automates the data factory in order to free up resources for analysis. Our approach includes pre-built algorithms, code generators, logical modeling, and the concept of analysts conducting their own data transformations without the need for a developer.

Oracle Data Integrator



Oracle Data Integrator delivers high-performance data movement and transformation across enterprise platforms with its open and integrated E-LT (Extract, Load, Transform) architecture and extended support for Big Data. Oracle Data Integrator is critical to leveraging data integration initiatives on-premise or in the cloud, such as Big Data management, Service Oriented Architecture and Business Intelligence. An easy-to-use user interface combined with a rich extensibility framework helps Oracle Data Integrator improve

productivity, reduce development costs and lower total cost of ownership among data-centric architectures.

Oracle Data Integrator's extract, load, transform (E-LT) architecture leverages disparate relational database management systems (RDBMS) or Big Data engines to process and transform the data. This approach optimizes performance and scalability and lowers overall solution costs. Instead of relying on a separate, conventional ETL transformation server, the architecture generates native code for disparate RDBMS or big data engines (SQL, HiveQL, PySpark, Pig Latin or bulk loader scripts, for example). The E-LT architecture extracts data from the disparate sources, loads it into a target, and executes transformations using the power of the database or Hadoop.

Oracle Data Integrator uses a declarative flow-based user interface for enhanced user experience and productivity. The interface combines the simplicity and ease-of-use of the declarative approach with the flexibility and extensibility of configurable flows. This blend simplifies common data integration design and deployment use cases, shortening implementation times. Data integration designers describe source and target data formats and data integration processes. The business user or the developer can focus on describing what to do, not how to do it. Oracle Data Integrator generates, deploys and manages the code required to implement those processes across the various source and target systems.

Oracle Enterprise Data Quality

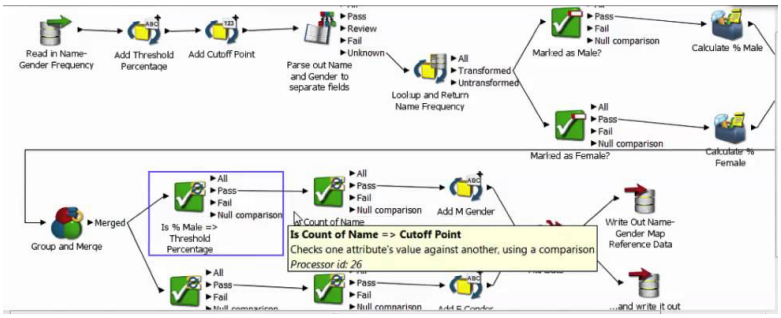
Data is very rarely available in a completely neat and ordered fashion. Typical problems include:

- Constructed fields, where a customer ID may be made up of a location code, a customer reference, and an account manager code
- Misfielded data, such as names, comments, or telephone numbers in address blocks
- Poorly structured data such as addresses, where data can flow from one field to the next.
- Notes fields that store information that the data structure doesn't support, but that contain useful semi-structured data that normally cannot be analyzed or extracted

All of these problems can be solved using Oracle Enterprise Data Quality. Using a data-driven approach to rapidly tag or describe data, it can manipulate a single record by parsing it into multiple structured elements (and, if required, records) and standardize results according to predefined rules.

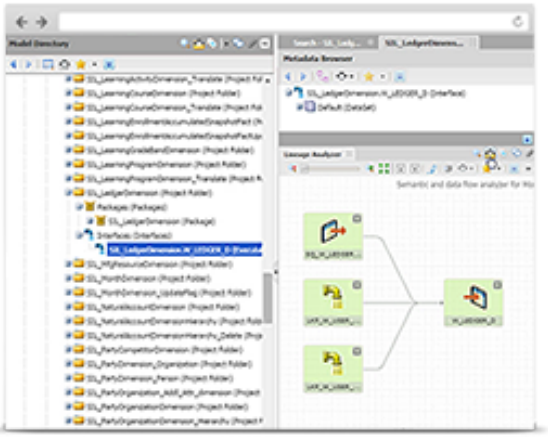
Innovative parsing and phrase analysis technology uniquely allows you to find hidden knowledge within any text field and create rules to standardize it into structured data.

Oracle Enterprise Data Quality provides a rich palette of functions to transform and standardize data using easily managed reference data and a simple graphical configuration. In addition to functions for basic numeric, string, and date fields, functions for contextual data such as names addresses and phone numbers are provided. Users can also quickly configure, package, share and deploy new functions that encapsulate rules specific to their data and industry without any coding.



Using a web browser, workers and managers can monitor and review ongoing data quality against defined metrics. Data quality dashboards allow problems to be quickly identified and dealt with before they start to cause significant business impact. Graphical views show data quality trends over time, helping your organization protect its investment in data quality, by giving visibility to the right people.

Oracle Metadata Management



Oracle Enterprise Metadata Management brings powerful business capabilities to the modern enterprise to harvest and govern metadata across its whole Data Management technologies. By being able to provide data transparency not only within Oracle but also 3rd party technology, Oracle Enterprise Metadata Management is a must have technology for any organization looking to seriously tackle Governance, Productivity Improvement and Lifecycle Management challenges.

Metadata Management is essential to solve a wide variety of critical business and technical challenges. They include but are not limited to; how report figures are calculated, understanding the

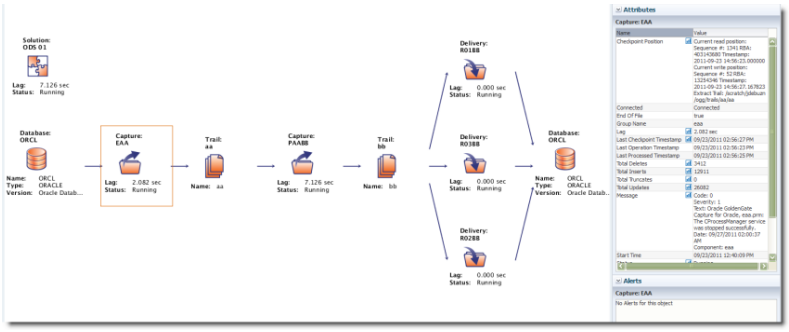
impact of changes to data upstream, and surfacing data lineage reports in a business friendly way UI and providing reporting capabilities on the entire metadata of an enterprise for analysis and improvement. The key features of Oracle Enterprise Metadata Management include:

- Report to Source Lineage
- Impact Analysis
- Model Versioning
- Annotations and Tagging
- Supports Metadata Standards
- Build and maintain Business Glossary
- Import 3rd party Business Intelligence Metadata
- Import 3rd party ETL Metadata
- Import 3rd party Database Metadata
- Big Data Enabled

Oracle Enterprise Metadata Management reduces the risk inherent in making any changes to the data in the organization, be it from the source to the final report. Where many business applications and decision systems rely on the same set of data, it is important to be able to authoritatively determine the impact that a change can affect to upstream and downstream applications.

Golden Gate

Oracle Golden Gate is a heterogeneous real-time database replication, integration, big data streaming, and high availability platform. It provides organizations real-time data integration for analytics, zero downtime migrations, cloud consolidations, real-time reporting replicas, and active-active data distribution for continuous availability.

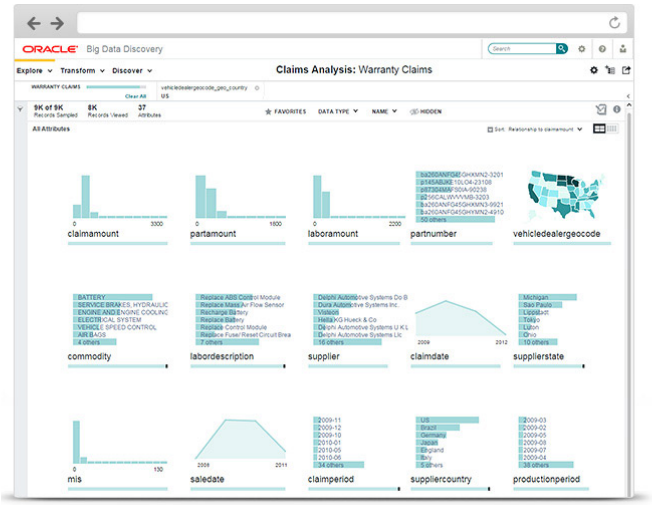


Oracle Golden Gate Studio enables you to design and deploy high-volume, real-time database replication by automatically handling table and column mappings, allowing drag and drop custom mappings, generating best practice configurations from templates.

DATA LAB

The data and analytic enterprise requires a laboratory environment. The Data Lab, often staffed by Data Scientist, is a separate data environment from production or mission critical operations. It will use similar tools as the production environment (eg. Big Data, RDMS, NOSQL) allowing for analysis and experimentation on new data sets and models to improve and validate data hypothesis prior to migration to the data platform. The lab provides creates algorithms to achieve a new level of analysis, discovery and understanding of new data sets. The lab also supports the triage and cleansing processes in the Data Factory, and quick analysis of the value or relevance of data to the enterprise. Oracle's data analytics portfolio provides key tools the make the data lab functions easier to execute and faster to perform.

Big Data Discovery

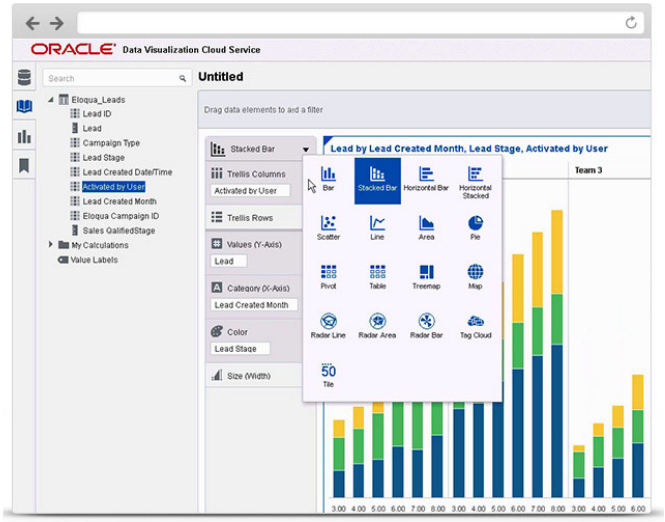


Big Data Discovery is a data discovery, exploration and data wrangling application that runs on the Hadoop eco-system. It's primary purpose is to provide the data scientist or analyst the ability to visually understand files in HDFS or data in data stores, to transform the data to fit the needs of the user, to visually query and perform analytics on the data and share those results with other scientist and analysts.

Big Data Discovery performs a triage function on the file and categorizes all of the file attributes, indexes all contents for search and displays null value readings on each attribute. The transform functions allows for very sophisticated transforms with an easy to use menu of pre-built functions including entity extraction on raw text and geo-coding. To execute the transforms on large data sets, the application auto-generates and executes Map/Reduce code for the user. The visualization feature allows for simple drag and drop design with machine learning algorithms picking the best format for the visualization of interest.

VISUALIZING ANALYTICS

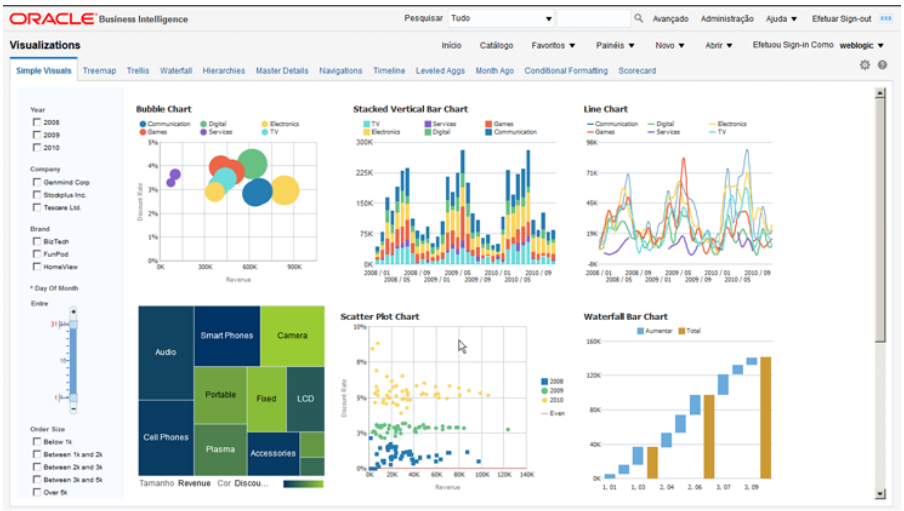
Oracle Data Visualization



Providing interactive data displays, Oracle Data Visualization Desktop is a application that enables analyst and other end-users to access, explore, and blend data so that they can quickly see patterns and share intelligent data visualizations. Visualizing data is as easy as dragging attributes onto the screen. Optimal visualizations are automatically displayed based on the type of data selected, with no upfront configuration, and are also positioned automatically without requiring precise placement—so users get started analyzing right away, rather than wasting time configuring graphs, charts, and layouts. Further, Oracle Data Visualization automatically infers connections between separate data sources, as well as provides an intuitive user experience for redefining connections and creating new ones. No modeling is required.

Oracle Data Visualization can pull data for analysis from any part of the data analytic enterprise including SQL and NoSQL databases, Hadoop databases, and even cloud data stores.

Oracle Business Intelligence Answers



Oracle Business Intelligence 12c Answers is a unique platform that enables customers to uncover new insights and make faster, more informed business decisions by offering agile visual analytics and self-service discovery together with best-in-class enterprise analytics. Instant mobile, highly interactive dashboards, powerful operational reporting, just-in-time alerts, content and metadata search, strategy management, native access to Big Data sources, sophisticated in-memory computing, and streamlined systems management combine to make Oracle BI 12c a comprehensive solution that reduces the total cost of ownership and increases return on investment for the entire organization.

Oracle BI Answers provides true end user ad hoc capabilities in a pure Web architecture. Users interact with a logical view of the information—completely hidden from data structure complexity while simultaneously preventing runaway queries—and can easily create charts, pivot tables, reports, and visually appealing dashboards, all of which are fully interactive and drillable. They can be saved, shared, modified, formatted, or embedded in the user's personalized Oracle BI Intelligence Dashboards. The results are new levels of business user self-sufficiency in an environment that is fully secure and controlled by IT.

HARDWARE ANALYTICS ADVANTAGE

With the accumulative sourcing of intelligence data from a proliferation of connected devices, sensors, collection platforms and social networks, there is an increase in the volume, variety, and velocity of data that can overwhelm standard IT architectures. Some of this data loses its operational value in a short time frame, i.e. perishable insights. The ongoing challenge is to ingest, cleanse, transform, integrate, analyze, provide predictions, and visualize this data in shorter

timeframes to allow timely decisions and actions to be taken. Hardware architecture selection is key for these mission-critical systems to meet current and future operational requirements.

In general, Oracle applications, analytics, middle-ware, and database technologies discussed in this paper can operate on a wide variety of non-oracle hardware, storage, and/or operating systems. At Oracle, we also design our own silicon, compute servers, operating systems, storage systems, engineered systems, and management tools. All of these technologies are designed to be the best in tier for what it does; however, part of our goal is to make the IT simpler, integrated, standards based, and performant so that our customers can solve increasingly complex mission problems. This is a huge differentiator for Oracle compared to other companies in the hardware business; it is our ability to take all of these different technologies and the understanding of these technologies to go build better individual components as well as co-engineering between these information technology tiers.

While our applications and database technologies are relatively high in the IT technology stack, with our co-engineering approach, we've been investing in wide-ranging R&D efforts to improve the compute, storage, network, and processor-chip "engines" that make our technology perform at the speed of mission. Thereby, running Oracle Software on Oracle hardware allows the software to recognize and take advantage of Oracle's comprehensive applications-to-disk co-engineering design improvements, and end-to-end testing to simplify our customers' most complex system requirements.

Oracle X86 Servers and Partnership with Intel

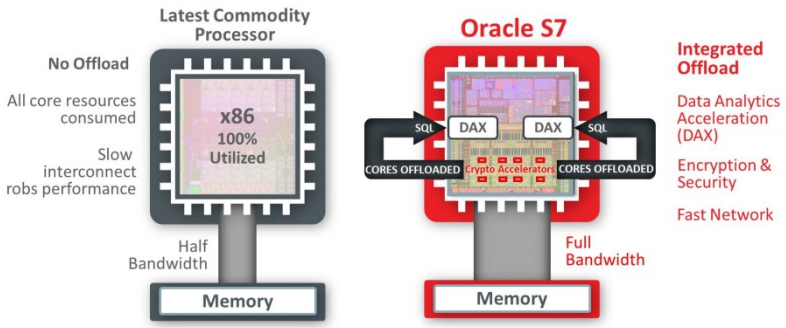
Oracle has an extensive multi-year NDA partnership with Intel to build enterprise class x86 servers that are used within Oracle's Engineered Systems, Oracle's Cloud offerings, and by our commercial customers[1]. Intel and Oracle staffs have access to each other's development centers, and support each other during the chip and server design lifecycle. Unique to Oracle compared to other hardware suppliers is that while others may use Intel's reference motherboard architecture , Oracle designs 100% of its motherboards to optimize firmware design, fault diagnostics, and increase memory throughput, and other optimization beyond the reference design patterns. Oracle has co-developed with Intel an x86 chip that allows the core vs. clockspeed ratio to be changed without

[1] https://www.oracle.com/servers/x86/x5-8/oracle-intel-x86.html

reboot to maximize the needs for either batch processing or single thread performance. By co-developing the server with Intel, and having design control over the firmware, BIOS, motherboard, and other system components, Oracle provides extreme performance and highly reliable systems for running customer specific and/or Oracle software on either Oracle Solaris, Oracle Linux, or other certified operating systems like Windows.

Oracle SPARC M7 and S7 Processor Key Technologies for SPARC Servers

Oracle has made extensive investment in SPARC processors and servers since the acquisition of Sun in 2010. Over 6 new chip designs in 5 years, to provide enhanced single thread, multi-thread, memory I/O, and encryption performance over x86 chips – and drive to the unique Oracle capability of providing "Software on Silicon" in the M7 and S7 SPARC Chips. All of these are enhancement to the compute engine for addressing the needs of mission-critical systems and advanced analytics.



The Software in Silicon capabilities available on both the 32 core M7 Processor and the 8 Core S7 Processor are:

1. Each processor has **Data Analytics Accelerator (DAX)** coprocessors that offload decompression and analytics processing from the SPARC M7/S& cores, freeing those cores to process other pipeline instructions and workloads virtually without overhead. The low-overhead interprocess communication permits DAX coprocessors on different processors to also exchange messages and access remote memory locations without CPU involvement.
    a. Up to a 10x/core query acceleration when used with Oracle In-Memory Option
2. **Silicon Secured Memory** real-time data integrity checking guards against the pointer-related software errors and vulnerabilities usually employed by malware. This fast and efficient in-silicon, on-chip monitoring allows applications to immediately identify unauthorized or erroneous memory accesses, diagnose the cause, and take appropriate recovery actions.
3. Hardware in-line decompression can fit up to three times (3x) more data in the same memory footprint without a performance penalty. Now, larger objects, such as entire database tables, can take advantage of processing with the Oracle Database In-Memory option.

Accelerated on-chip cryptography performs encryption and decryption operations at hardware rates, eliminating the performance and cost barriers typically associated with the high level of secure computing that is increasingly essential for all business applications. Oracle SPARC has 4x threads per core than an x86 processor for higher degrees of parallel processing of database analytics, and the SPARC processor has higher memory bandwidth to increase the processor's I/O performance with the server's memory – which on SPARC servers can be as much as 16TB.

Oracle SPARC Servers[2] T7-2, T7-4, M7-8, M7-16, and S7-2 provide the performance and throughput advantages of SPARC design and Software on Silicon for Oracle Database, Java, Weblogic, Apache SPARK on SPARC[3], and other software technologies compared to other hardware alternatives. Combine the SPARC chip with follow software technologies allows for these performance gains over x86[4]:

- 10x faster per core query performance for Oracle Database with In-Memory
- 1.5x faster per core performance on Oracle Database Spatial and Graph
- Up to 4.8x faster per core with Oracle Database and Oracle Advanced Analytics
- 1.9x faster per core with Oracle NOSQL Database
- 15.6x faster per core performance with Sparc SQL
- 3.6x-21.8x fsater performance of Java JDK8 Streams

The M7-16 SPARC Server provides these on-chip features described above with 512 processing cores, 4096 threads of compute, and 16TB of memory. This is the most powerful general purpose server available on the market today. The commercial industries that are interested in these advances include Financial, Retail, Telecommunication, Public Sector, and more.

Oracle provides SPARC servers as well as our x86 servers as commercial offerings, though we continue the co-engineering efforts with these compute platforms building blocks to provide extreme performance and simplify IT though our Engineered System solutions.

[2] https://www.oracle.com/servers/sparc/index.html

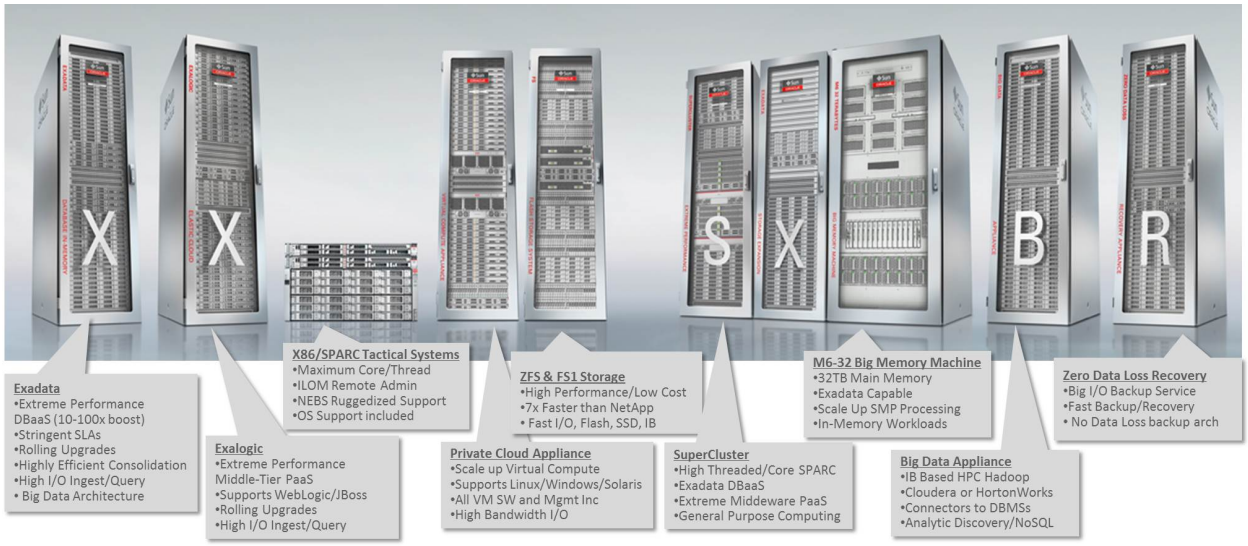
[3] https://community.oracle.com/docs/DOC-994843

[4] https://blogs.oracle.com/bestperf/, https://blogs.oracle.com/BestPerf/entry/accelerating_spark_sql_using_sparc

Oracle Engineered Systems

Oracle's family of engineered systems combines best-of-breed hardware and software components with game-changing technical innovations. These systems include Exadata Database Machine, Exalogic Elastic Cloud Machine, SPARC Super Cluster, Big Data Appliance, Oracle Private Cloud Appliance, Exalytics InMemory Machine, Oracle MiniCluster, Oracle Database Appliance, Oracle Cloud Machine (OCM), and Oracle Zero Data Loss Recovery Machine (ZDLR).

The components of Oracle's engineered systems are designed and preassembled for targeted functionality and then - as a complete system - optimized for extreme performance. Designed, engineered, and tested to work together, Oracle's engineered systems are part of the Oracle Public Cloud and over thousands of customer datacenters with the significant benefits of being extremely performant, requiring less datacenter footprint, and streamlining data center operations to make traditional deployments more efficient. Engineered Systems are available in partial rack to multi-rack configurations based on Customer need.



Oracle Exadata Database Machine

As an example of the co-engineering in Exadata Database Machine, the database management servers are connected to the storage servers via 40Gb/sec Infiniband that allows Remote-Direct-Memory-Access (RDMA) which removes the traffic congestion and latency of Ethernet, Quality of service to specific high-priority database protocols, all while providing increased bandwidth for moving data from storage to the server, and Exadata storage is loaded with Smart Scan database

software to offload queries requests from the database server in a highly paralyzed fashion for search, flash memory acceleration, compression, and other database operations. An Exadata can come in different configuration of flash to high-density disk drives, as well as several x86 or SPARC server configurations.

As an example of the low latency and high throughput of Engineered Systems: A traditional full rack Exadata Database Machine X6-2, with 8 database servers and 14 Extreme Flash storage servers, can achieve up to 350 GB per second of analytic scan bandwidth from SQL, and 0.25 ms Database I/O latency at 2.4 Million Flash IOPS when running database workloads. A slightly different full rack combination, with 10 database servers and 12 Extreme Flash storage servers, can achieve up to 5.6 Million random 8K read and 5.2 Million random 8K write I/O operations per second (IOPS) from SQL, which is an industry record for database workloads.

Oracle Database Appliance X6-2S, X6-2M, and X6-2HA

The Oracle Database Appliance is an Oracle Engineered System that saves time and money by simplifying deployment, maintenance, and support of high availability database solutions. Optimized for the world's most popular database— Oracle Database—it integrates software, compute, storage, and network resources to deliver high availability database services for a wide range of custom and packaged online transaction processing (OLTP), in-memory database, and data warehousing applications. It portfolio offers and entry level X6-2S model for development, to and X6-HA for High availability within a remote office or tactical capability.

The Oracle Database Appliance X6-2-HA hardware is a 6U rack-mountable system containing two Oracle Linux servers and one storage shelf. Each server features two 10-core Intel® Xeon® processors E5-2630 v4, 256 GB of memory, and 10-Gigabit Ethernet (10GbE) external networking connectivity. The two servers are connected together via a redundant InfiniBand or optional 10GbE interconnect for cluster communication and share direct-attached high performance solid-state SAS storage. The storage shelf in the base system is half populated with ten solid-state drives (SSDs) for data storage, totaling 12 TB of raw storage capacity. The storage shelf in the base system also includes four 200 GB high endurance SSDs for database redo logs to improve performance and reliability. All hardware and

software components are engineered and supported by Oracle, offering customers a reliable and secure system with built-in automation and best practices.

Oracle Super Cluster M7

Oracle SuperCluster M7 is a ready-to-deploy secure cloud infrastructure for both databases and applications. It is an engineered system that leverages the SPARC M7 Processor discussed above, and combines compute, networking, and storage hardware with virtualization, operating system, and management software into a single system that is extremely easy to deploy, secure, manage and maintain.

Oracle SuperCluster M7 features the industry's most advanced security, incorporating a number of unique runtime security technologies, documented and tested system-wide security controls and best practices, and integrated automated compliance verification tools. Oracle SuperCluster M7 is the world's fastest engineered system, delivering incredible performance under a wide range of workloads ranging from traditional enterprise resource planning, to customer relationship management and data warehouses, to e-commerce, mobile applications, and real-time analytics.

Oracle MiniCluster S7-2

The Oracle MiniCluster S7-2, built on the SPARC S7 processor discussed above, is an extremely simple and efficient engineered system designed to run enterprise databases and applications with uncompromising security. Its simplicity, out-of-the-box performance, security, DISA STIG compliance, reliability, and small form factor make it an excellent choice for remote offices, small offices, and agile software development (DevOps) environments. It is ideal for highly security-sensitive applications, such as managing patient medical records, processing financial transactions, handling secure communications, running mission-critical applications, and hosting security-related services.

Oracle MiniCluster S7-2 is built with Oracle's SPARC S7 processor, all-flash database storage, and virtualization technologies that offer bare-metal performance and unique

capabilities for accelerating in-memory databases and applications. It integrates a range of unique technologies and approaches in order to provide a highly secure infrastructure with minimal effort and risk. For example, its capabilities include the following:

- Secure your applications with hundreds of security controls that are integrated into the system by default:
- Comprehensive built-in data protection with encrypted data at rest, transit, and assured secure data erasure
- Ready-to-use hardened and minimized VMs and secure access via SSH, TLS, and IPSec
- Secure Oracle Integrated Lights Out Manager (Oracle ILOM) and verified boot environment
- Built-in host-based firewalls and data link protection
- Role-based access control with least privilege and separation of duties with authorization workflow for security sensitive operations.
- Centralized key management with PKCS#11, KMIP, and FIPS-140 compliance support
- Comprehensive audit policy with dedicated auditor role and centralized audit store providing full visibility monitoring of user and system activities
- Comply with DISA-STIG, PCI-DSS, or CIS-equivalent security standards easily with the push of a button.
- Verify the compliance of VMs automatically either monthly or on demand through the built-in compliance verification tools.
- Actively protect data in memory from security exploits such as heartbleed with the SPARC S7 processor's Silicon Secured Memory capabilities.
- Encrypt end-to-end data with near-zero overhead through the SPARC S7 processor's cryptographic acceleration capability.
- Ensure application administrators and compromised applications are unable to accidently or deliberately alter the configuration of VMs in ways that would expose systems to attack , through read-only VMs.

Oracle Big Data Appliance

Oracle Big Data Appliance is an open, high performance, multi-purpose engineered system you're your data lake via Hadoop and NoSQL processing. The Big Data Appliance is designed to run diverse workloads – from Hadoop-only workloads (Yarn, Spark, Hive etc.) to interactive, all-encompassing interactive SQL queries using Oracle Big Data SQL. These capabilities are

available for on-premises deployment as well as the Oracle Cloud.

The Big Data Appliance provides an open environment for innovation while maintaining tight integration and enterprise-level support. Organizations can deploy external software to support new functionality – such as graph analytics, natural language processing and fraud detection.

It simplifies day-to-day operations by providing a simple one- command installation, update, patch and expansion utility – Mammoth – which enables rapid deployment updates (typically quarterly) to the frequently evolving Hadoop stack without incurring significant downtime. Mammoth also enables Oracle-tested, seamless upgrades between Hadoop versions and automated service management to ensure the best balance between Hadoop Master Nodes and Data Nodes.

The appliance is supported by Oracle, giving organizations a single point of support for their hardware, all integrated software (including all Cloudera software) and any additional Oracle software installed.

SUMMARY

The Oracle approach to data analytics is to create a platform that is open, feature rich, integrated, standards-based, and performant. It makes big data simple by integrating Hadoop, NoSQL and SQL across the Fast Data, Data Factory, Data Lake and the Data Lab environments. The ultimate goal is to provide a cost efficient, easy to use platform that transforms all data across the enterprise into actionable information that leads to better decisions. We are proud that 3rd party analysts agree with our innovations, as Gartner has chosen Oracle as the leader in Gartner's Magic Quadrant for data platforms for data analytics.

ORACLE

Big Data, Fast Data, All Data
Feb 2017
Author: Mark Tatum, Dan Brennan, Darryl McGowan

Oracle Corporation
World Headquarters
500 Oracle Parkway
Redwood Shores, CA 94065
U.S.A.

Worldwide Inquiries:
Phone: +1.650.506.7000
Fax: +1.650.506.7200

oracle.com

Oracle is committed to developing practices and products that help protect the environment

Hardware and Software, Engineered to Work Together