ORACLE DATABASE 12c

Oracle Big Data Management System

A Statement of Direction for Big Data and Data Warehousing Platforms

ORACLE STATEMENT OF DIRECTION | APRIL 2015

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.

Introduction: Oracle Big Data Management System Today

As today's enterprises embrace big data, their information architectures must evolve. Every enterprise has data warehouses today, but the best-practices information architecture embraces emerging technologies such as Hadoop and NoSQL. Today's information architecture recognizes that data not only is stored in increasingly disparate data platforms, but also in increasingly disparate locations: on-premises and potentially multiple cloud platforms. The ideal of a single monolithic 'enterprise data warehouse' has faded as a new more flexible architecture has emerged. Oracle calls this new architecture the Oracle Big Data Management System, and today it consists of three key components

- The data warehouse, running on Oracle Database and Oracle Exadata Database Machine, is the primary analytic database for storing much of a company's core transactional data: financial records, customer data, point- of-sale data and so forth. Despite now being part of a broader architecture, the requirements on the RDBMS for performance, scalability, concurrency and workload management are in more demand than ever; Oracle Database 12c introduced Oracle Database In-Memory (with columnar tables, SIMD processing, and advanced compression schemes) as latest in a long succession of warehouse-focused innovations. The market-leading Oracle Database is the ideal starting point for customers to extend their architecture to the Big Data Management System.
- The 'data reservoir', hosted on Oracle Big Data Appliance, will augment the data warehouse as a repository for the new sources of large volumes of data: machine-generated log files, social-media data, and videos and images -- as well as a repository for more granular transactional data or older transactional data which is not stored in the data warehouse. Oracle's Big Data Management System embraces complementary technologies and platforms, including open-source technologies: Oracle Big Data Appliance includes Cloudera's Distribution of Hadoop and Oracle NoSQL Database for data management.
- A 'franchised query engine,' Oracle Big Data SQL, enables scalable, integrated access in situ to the entire Big Data Management System. SQL is the accepted language for day-to-day data access and analytic queries, and thus SQL is the primary language of the Big Data Management System.[1] Big Data SQL enables users to combine data from Oracle Database, Hadoop and NoSQL sources within a single SQL statement. Leveraging the architecture of Exadata Storage Software and the SQL engine of the Oracle Database, Big Data SQL delivers high-performance access to all data in the Big Data Management System.

Using this architecture, the Oracle Big Data Management System combines the performance of Oracle's market-leading relational database, the power of Oracle's SQL engine, and the cost-effective, flexible storage of Hadoop and NoSQL. The result is an integrated architecture for managing Big Data, providing all of the benefits of Oracle Database, Exadata, and Hadoop, without the drawbacks of independently-accessed data repositories.

Note that the scope of this statement of direction is the data platform for Big Data. An enterprise Big Data solution would also be comprised of big data tools and big data applications built upon this data platform.

A Vision for Oracle Big Data Management System

Oracle will extend its current Big Data Management System is to provide fast, integrated, secure access to all data – not only data stored in an Oracle Exadata-based data warehouse or Oracle Big Data Appliance, but also data stored in operational NoSQL databases, transactional relational databases, streaming data sources, and more.

Oracle's future Big Data Management System will provide a framework for easily incorporating new data sources, ensuring that these new data sources can be seamlessly accessed and managed. Oracle will instantiate its BDMS

[1] For more information, see the whitepaper: "SQL – the natural language for analysis" Also, asserting that SQL is the primary language in no way precludes support for other languages for specialized use-cases, and Oracle invests in languages for analytics (e.g. R), graph processing, spatial processing, and others on its platform.

vision incrementally, in an effort spanning multiple products and release cycles, and Oracle's initial focus is upon four areas:

- Global Metadata Services – a single view of all available data across multiple data stores, exposed in a format similar to Oracle's data dictionary.
- Global Query Routing and Processing – optimized query execution across multiple data stores. A single query may possibly merge data from multiple data stores. A variety of query optimization and caching techniques will be applied to optimize performance.
- Global Resource Monitoring and Management – prioritization of workloads across the entire BDMS ecosystem.
- Global Data Optimization – the ability to automatically move data from one repository to another (for example, from a data warehouse to a data reservoir or vice versa) based on query performance requirements and/or storage costs.

In each of these areas, Oracle's strategy is to extend its existing in-database features (such as its data dictionary, SQL query engine, query optimizer, resource manager, and data optimization) in order to manage the entire Big Data Management System. By using a common SQL engine and metadata model, Oracle is delivering a unified experience to applications, tools and administrators. Moreover, by leveraging its existing database code base, Oracle is able to rapidly deliver new capabilities to the BDMS and deliver more mature and more functionally complete capabilities, as compared to building new components from scratch.

Oracle has already delivered its first installment of Big Data Management System with Big Data SQL, and this architecture underpins Oracle's strategy to deliver unified, high-performance query access across Big Data Management Systems. With Big Data SQL, the same storage level optimizations that are the foundation of Exadata have been applied to Hadoop and NoSQL Database. Future release of Big Data SQL will continue to extend query optimizations traditionally restricted to databases onto Hadoop and other platforms. This high-performance, platform-independent query capability will enable organizations to store data in the most appropriate platform (based on cost and performance considerations) without considering the undue penalties of data movement and federation.

Oracle Database as part of the Big Data Management System

The Oracle Database, as the host for the data warehouse, occupies a central role in the burgeoning Big Data Management System. Oracle Database's future directions are guided by the following three basic principles:

- Performance, concurrency and scalability are paramount for data warehousing. Every release of the Oracle Database will continue to include enhancements for these immutable data warehouse requirements. Recent Oracle releases had dozens of features for enhanced query optimization, parallelism, more efficient join and aggregation algorithms, and concurrent workload management (to name just a few key areas), and future releases will continue investment in these core areas.
- Lock-step evolution with hardware platforms. A decade ago, databases, servers, and storage platforms were developed independently, and deployed and supported by separated teams. In the interim, data warehouse appliances emerged, heralding a broader adoption of database appliances and integrated systems. Oracle fully recognized the benefits of this approach, with Oracle Exadata. Moreover, the Oracle Database has gone much farther in developing specific features tightly integrated with hardware capabilities. On the Exadata platform, Oracle has introduced IO and network database resource management, optimized database-centric network protocols over Infiniband, and increased fault-tolerance – capabilities that would be difficult to implement on generic hardware platforms. At the same time, Oracle has also enhanced its database to leverage hardware advances available on all platforms: the introduction of Oracle Database In-Memory is recognition that large-memory configurations are increasingly available to enterprise customers, and that such memory volumes enable new types of database algorithms. The use of SIMD processing for in-memory operations is the latest in a series of enhancements of the Oracle database leveraging the specific capabilities of processors (on-chip decryption

algorithms would be another recent example). Looking forward, Oracle will continue to build specific features for Exadata, the best hardware-software combination for running data warehouse workloads.[2]

- Oracle Database can make major architectural changes to address new customer requirements. A favorite hobby of new entrants to the database market is to paint Oracle, the market-leading database, as inflexible -- and promote their product on the basis that Oracle will never be able to provide the same type of functionality as their new platform. Such vendors pursue this positioning at their peril: object-oriented databases, massively-parallel databases, columnar databases, data warehouse appliances and other trends have been touted as replacements for Oracle Database only to later see their core benefits subsumed by the Oracle platform. The reason that Oracle Database, as a thirty-year-old software platform, continues to thrive is because it has successfully adapted to new generations of computing paradigms by introducing key new features, and merging these new capabilities into its existing platform. Oracle Database 12c continues this trend by delivering
  - Oracle In-Memory: Full columnar tables; SIMD processing
  - Document-store capabilities: NoSQL-style API's for document-store application development; JSON documents stored in database accessible via SQL and new API's

  Oracle will extend its database platform to address all key data management requirements. Oracle's track record has shown that even requirements, that at first glance seem incompatible with Oracle's database architecture, have been successfully incorporated in the Oracle Database. Oracle will continue to drive innovation into the Oracle Database.

The Oracle Database has a primary role in Oracle's Big Data Management System -- both as a distinct data platform for data warehouses, and also as a foundation for the Big Data Management System in which key Oracle Database capabilities are extended to the entire Big Data ecosystem.

Oracle Big Data Appliance

Oracle's Big Data Appliance hosts the data reservoir in the Big Data Management System.

Oracle Big Data Appliance (BDA) leverages state-of-the-art hardware components to build the fastest, most highly available and cost effective big data platform in the market. Following the path of Oracle's other engineered systems, Big Data Appliance generations closely follow processor update cycles; new generations of the Big Data Applience feature upgraded processors, and are released approximately every year (typically 3-6 months after a new processor becomes available, allowing time for integration and testing) . In addition to adopting new processors, every new generation of Big Data Appliance adopts state-of-the-art memory, storage and networking -- bundling together as many new hardware components as possible, in order to eliminate cumbersome piecemeal changes. The goal is to be timely enough to maintain leading performance and price-performance, while being conservative enough to ensure the highest quality.

Big Data Appliance fully supports multiple generations of hardware, both within the same rack and in multi-rack configurations, enabling the systems to grow as needed. Oracle automatically configures each component of a multi-generation Big Data Appliance to enable optimal utilization of all hardware resources; newer nodes with faster cpu's or more memory will be configured differently than nodes from a previous hardware generation.

For the software updates, Big Data Appliance includes Cloudera's Distribution including Apache Hadoop (CDH), and Big Data Appliance software updates closely follow Cloudera software updates. New software bundles for the Big Data Appliance are typically available within a few weeks of a Cloudera release. Big Data Appliance software updates include not only CDH, but also updates to the entire software stack including operating system, Oracle NoSQL Database, and firmware drivers. Oracle fully tests software upgrades on Big Data Appliance internally before

[2] For more details on Exadata's future directions, see "Oracle Exadata Statement of Direction"

releasing the entire software stack on Big Data Appliance to customers. This integrate software-release strategy maintains leading software versions on Big Data Appliance while providing high quality systems.

Additionally, Big Data Appliance provides the openness to add new software onto the core set provided with the system. Additional libraries and components (for example, new Apache projects) can be added to Big Data Appliance. While the newly added software is not supported by Oracle (it is supported as any other open source software), the existence of this additional software does not impact Oracle Support for all of the other pre-installed software components.

Big Data Management System and the Cloud

Oracle's core Big Data products will be available in the Cloud. In line with Oracle's corporate cloud strategy, the same capabilities and products that are available on-premises are also being delivered as Oracle cloud services. This allows customers to easily migrate existing Oracle systems to the cloud (or vice versa, from the cloud to on-premises).

Given the breadth of a typical enterprise Big Data ecosystem, there will be components of a BDMS that are on-premises while other components will be deployed in a cloud platform. Query access to hybrid cloud systems will be addressed with Oracle Big SQL's 'franchised query' architecture, and future components of the Big Data Management System are being built under the assumption that hybrid cloud architectures will be the most typical enterprise configuration.

Conclusion

The current transition from the Enterprise Data Warehouse centric architecture to the Big Data Management System is the most momentous architectural change in analytic infrastructure in the past 20 years. With this statement of direction, Oracle seeks to guide enterprises as they consider their long-term architectural plans under the requirements of Big Data. To meet these enterprises' requirements, Oracle is delivering innovative new technologies such as Big Data SQL, embracing new platforms such as Hadoop, and extending the capabilities of its core database – all towards a common goal of continually extending and expanding its Big Data Management System.

Oracle Corporation, World Headquarters
500 Oracle Parkway
Redwood Shores, CA 94065, USA

Worldwide Inquiries
Phone: +1.650.506.7000
Fax: +1.650.506.7200

CONNECT WITH US


blogs.oracle.com/oracle


facebook.com/oracle


twitter.com/oracle


oracle.com

Hardware and Software, Engineered to Work Together

Copyright © 2015, 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. Cloudera, Cloudera CDH, and Cloudera Manager, Cloudera BDR and Cloudera Navigator are registered and unregistered trademarks of Cloudera, Inc. 0115


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