ORACLE NOSQL DATABASE

ORACLE DATABASE MOBILE SERVER

How Oracle NoSQL Database and Oracle Database Mobile Server Enhance an Internet of Things Data Management Solution

ORACLE WHITE PAPER | JANUARY 2017

ORACLE

Table of Contents

Introduction

The Internet of Things (IoT) has become an important technology that can facilitate both personal and organizational understanding of the devices that matter. As smart devices become more ubiquitous, architectures need to be put in place that effectively and efficiently deal with the wide range of devices and environments that will exist. This includes scenarios where data is streaming over a network as well as periods where data is produced but the devices are off-line.

An IoT strategy can open up new businesses and increase the understanding of existing, intelligent products and services. The business benefits of understanding the massive amounts of data flowing into analytics systems will be impressive. IoT is bridging the gap between IT (Information Technology) and OT (operations technology) to derive insights towards controlling the costs for various industries such as manufacturing, retail, logistics etc. For all those organizations that are implementing IoT strategies, data management will be at the centre of this digital transformation and a key differentiator would be the ability to derive meaningful business insights from gathered data against organizations that will merely gather data. IT departments will have to design new systems that can handle this data, store the data and make sense of the data, both in real-time and from a historical perspective.

IoT Explained and Data Volumes

With the instrumentation of devices, whether industrial or commercial, the need to collect, analyze and make decisions based on the data has created a new industry. Data can and will be generated from wearable devices, home appliances, industrial machinery and transportation systems. Estimates are that devices created 200 exabytes of data in 2014, which will grow to 1.6 zettabytes by 2020. The data that is generated will be used in a wide range of industries, from manufacturers trying to understand product performance to healthcare organizations attempting to understand costs and benefits of treatment options. This volume brings forward the challenge of managing such data. This challenge is about managing data at all points in its lifecycle, including on the edge, at the middle-tier gateways and at the backend. To be able to derive meaningful business insights it is critical to look at end-to-end data management solutions rather than just building up the solution at each stage in isolated manner. The data that is generated by the variety of devices needs to be collected, appropriately aggregated, and stored to derive intelligent insights. Identifying right set of data models for such un-structured data and the tools that implement them that can eventually lead to business insights will be critical to the success of IoT data management.

Business Value

The value of all of the data that is captured can affect real business outcomes. Some potential uses for businesses that capture and analyze the data include:

- Optimize business processes and reduce operations cost and risks
- Predict equipment failures and maintenance needs to improve overall equipment effectiveness
- Innovate with new business models and offer products as a service

- Offer differentiated customer experiences and better product quality

Critical Components of an High Performance IoT Environment

An IoT system or service needs to be able to work with the quantities of all of this data as well as the speed of the data being produced in a manner that was not previously implemented in traditional IT architectures. The amount of data that will be created by these devices needs to be filtered and ingested at rates that exceed many of today's client/server architectures. The raw data generated on the edge cannot be directly used to derive insights. For example, binary data sent by temperature sensors need to be converted into specific data models on the gateways, it needs to be massaged before it can be sent over to the analytics eco-system either on premise or in the cloud. A new architecture needs to be defined that can meet the end-to-end data management requirements. Oracle, today, can fulfill those with specialized products that can addresse the following:

- Reduction in the data that is sent over existing data pipes from the source to the datacenter (Oracle Berkeley DB and Oracle Data Mobile Server)
- Ability to store data locally and synchronize later if the data source is not connected to a network (Oracle Data Mobile Server)
- Requirement to store data for immediate and fast inquiries. (Oracle NoSQL Database)
- Systems that can process historical comparisons (Oracle Database)

Basic Architecture

A modern IoT architecture needs to be able to handle the high volume of data that can be generated, either from a certain device, a local set of devices, or devices around the world. A data model that maintains business context has to be maintained at each point during the lifecycle in both directions. Based on the capabilities of the devices at the edge, a data management architecture needs to be defined. A temperature sensor can only emit certain binary data which is then processed or massaged at the gateway where a more capable device installed on the traffic signal can identify outliers by itself by comparing the data with locally stored history data before sending it to the gateway. At the gateway, organizations may need to analyze the data and only send a limited set of data over to the backend for further storage and analytical processing. Gateways often need to maintain local data repository in case connectivity with some or all of the devices is lost or networks go down. Gateway software would also play crucial role in maintaining consistency of the data on the edge when the connectivity is restored. As the data inflows into the data center (be it on premises or in the cloud), the data will be typically dealt with in two different methods:

- Stored in a historical, relational database – The data may be stored in a corporate database for near term and long term analysis. Since a RDBMS has the ability to compare data with historical events, even if just a few hours old, the data could be stored in a full featured RDBMS.
- Shorter term requirements would include almost real time analysis of the data from the device, perhaps grouping with similar devices, or focusing on recent events.

After the storing of the data, each of the different databases can be used to display various trends, alarms, trigger actions, or be used to determine trends and predictive analytics.

Edge Filtering

The amount of data that can be generated, even from simple devices could overwhelm existing pipes and connections to a cloud based service. Even though the data may be in a simple form, the velocity of the data generated times the number of devices that generate the data might create too much data. Not all of thisdata is necessary or of high value. For example, if the temperature of a refrigerator was being monitored, and only

notification was needed if the values exceeded a certain threshold, then most of the time the temperature of the device would not be important, as it would not have exceeded either a lower or upper bound. This testifies the need for some data analysis and actions need to be carried out at the edge. In certain cases, the edge analysis needs to be done where the connection and response from the central server is either not available or feasible. As an example, a jet engine generates 1 Terabyte of data per flight. Given that the data is both critical and voluminous, a network transmission and real-time analysis becomes a significant challenge. Edge processing of that data is a required part of the architecture in such environment.

The filtering and processing of the data needs to happen at either the device, or before the data becomes aggregated with other data or is sent to the cloud. One place that this could happen is in a smart router or a gateway, which can reduce the data before sending it over the internet to the cloud or to a data center. A smart router with local database storage would allow fast and efficient data analytics at the edge. Oracle Berkeley DB, with its zero administration, high-performance concurrent read and write operations and small footprint is an excellent fit for an edge data storage engine.

Oracle NoSQL Database Benefits

Oracle NoSQL Database is a scalable, distributed NoSQL database, designed to provide highly reliable, flexible and available data management across a configurable set of storage nodes. Oracle NoSQL Database can ingest and write data to the storage nodes with throughput of 25,000 and higher inserts per second per shard. As the number of shards increase, performance increases as well for large and growing databases. With its simple implementation and optimizations to speed the storage and retrieval of data, Oracle NoSQL Database is an ideal solution for part of any IoT environment. Oracle NoSQL Database is simple to install and use and is an excellent choice for applications that need to store massive amounts of data quickly.

Developers can choose from variety of programming languages to build their client applications for Oracle NoSQL Database. Oracle NoSQL Database is available for download for either on-site use, or as Platform-as-a-Service within the Oracle Cloud. Once installed, Oracle NoSQL Database can be expanded online, i.e. without any downtime.. Thus, if the requirement for number of IOPS increases, additional, cost effective hardware can be added fairly transparently to the datastore in order to respond to increased workloads. The Oracle NoSQL database is ideal for time series data with fast inserts given its append-only architecture and its ability to purge data based on configurable elapsed time in the store (Time-To-Live feature).

Oracle NoSQL Database Architecture

The Oracle NoSQL Database has been designed to scale with workloads as well as show excellent performance even as databases grow. Oracle NoSQL Database consists of two separate components. The first is the Application Driver which resides on the server where the actual application resides. The purpose of the Driver is to keep track of and send the data to one of the data store nodes. The driver is also aware of the cluster topology and can identify which of the nodes can serve the request in faster and efficient manner. While writing the data, the Application Driver hashes the keys to determine which data store node the data will be sent to and stored on a hard disk drive. Once the data is stored, the storage node sends a copy of the data to other storage nodes as replicas, which are used both for reading back the data, as well as for high availability in case one of the disks or nodes goes offline. Below is a diagram of a simple Oracle NoSQL Database implementation that uses 3 storage nodes with 3 shards, and a replication factor of 3 (1 master and 2 replicas per shard). Figure 1 diagrams the basic Oracle NoSQL Database logical flow.

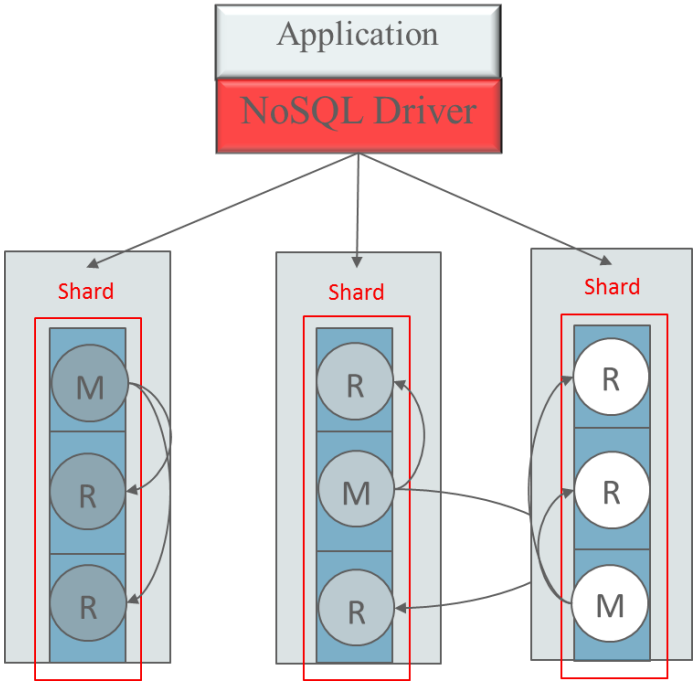

Figure 1 - Basic Oracle NoSQL Database topology with 3 shards and RF=3

Edge Filtering and the Database Mobile Server

The filtering of the data needs to happen at either the device, or before the data becomes aggregated with other data or is sent to the cloud. One place that this might happen is in a smart router, which can reduce the data before sending it over the internet to the cloud or to a data center. Oracle Database Mobile Server can meet data synchronization requirements in both directions. The core of the product is an advanced synchronization engine that is capable of keeping a large number of physically remote databases synchronized with a backend Database system. The key differentiator using the Oracle Database Mobile Server is its ability to maintain the database context. It understands the data model and its structure. This provides simplifies managing the data at the gateway than just passing it onto to the other side in canonical form.

Oracle Database Mobile Server Benefits

Where edge filtering, local storage and support for a disconnected mode are vital requirements for an IoT solution, Oracle Database Mobile Server 12c is a logical choice. It allows occasionally disconnected devices to store and process data locally or at the edge, and provides an efficient and scalable data synchronization engine for the centralized data storage and processing in either Oracle Datbase or Oracle NoSQL Database. The local or edge data can be saved in the Oracle Berkeley DB, SQLite or JavaDB databases or as JSON records in files or in memory. The client platform and storage choice depends on the volume and types of data that needs to be saved and analyzed. For example, for large volumes of structured local data, Oracle Berkeley DB is the most suitable storage option. The data can be indexed, searched and filtered using the standard SQL functions.

When network conditions are restored, A background sync process of Database Mobile Server can push filtered data to the backend, all while the device continues to collect and process new data. As an example, in a connected car scenario, multiple sensors constantly monitor vehicle state. It is predicted (automotivesensors2015.com) that the number of sensors per vehicle is going to reach 200 by year 2020. Local processing logic is the only currently possible solution for handling the large volumes of the sensor data. However, just storing this data locally reduces the value of possible insights that can be gained by a full statistical analysis. So, once the sensor data is filtered locally with averages, counts, outliers and other parameters calculated, the results can be propagated to an enterprise data center for further analysis.

Large number of IOT devices also presents a challenge for handling communication between these devices and the enterprise data centers. Oracle Database Mobile Server provides a highly optimized communication protocol that reduces the amount of metadata exchanged between client and server so that the bulk of the payload exchanged is pure data. Furthermore, Oracle Database Mobile Server is deployed as a standard J2EE container, so it can be installed in a highly distributed environment (for example on the Oracle Public Cloud) which allows it to scale out with increased workloads. Oracle Database Mobile Server is well suited for mission critical applications or any application where high performance and reliability are required.

Lambda Architecture

In IoT data storage and processing environment, the data may need to be acted on in a batch manner and a speed manner. A common term for this is called the Lamba Architecture, where incoming data is sent to two different processes, each of which can be queried and results are returned. The difference would be in the speed of storing and querying the data, compared to longer term storage and more complex queries being submitted.

A general workflow diagram would be as detailed below in Figure 2.

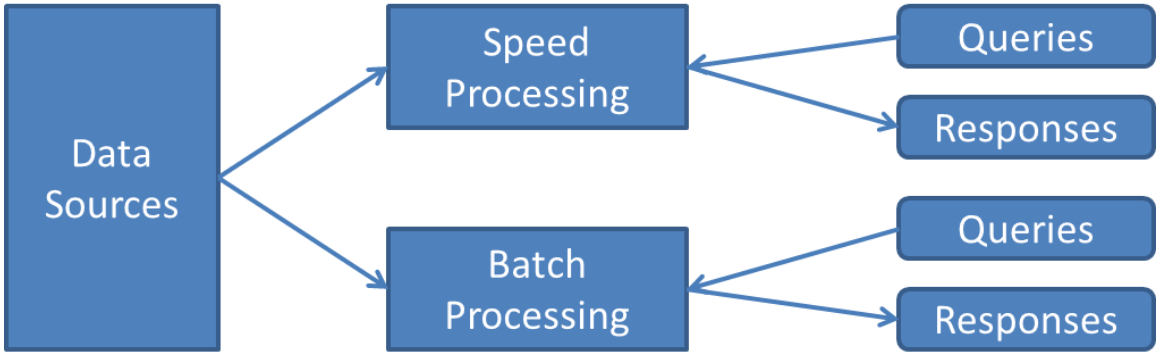

Figure 2 - Basic Lambda Architecture Layout

In the figure below, a number of Oracle technologies can be used to create an effective IoT solution. This includes Database Mobile Server, Oracle Database and Oracle NoSQL Database. The environment would look like the following:

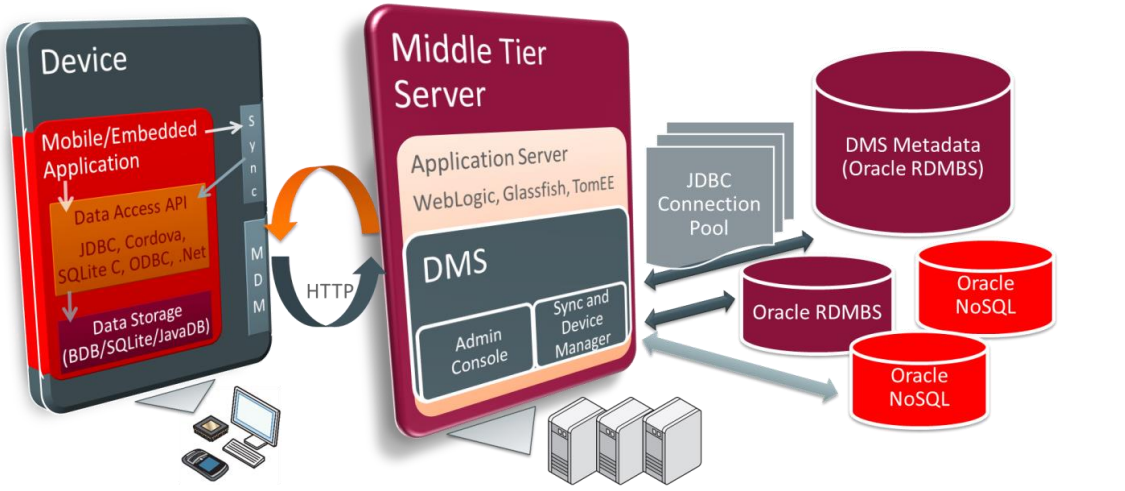

Figure 3 - Oracle Lambda Architecture

Existing Oracle IoT Cloud Service

The Oracle IoT Cloud Service is a comprehensive cloud offering that simplifies IoT planning and implementation. While it leverages Oracle NoSQL Database for data management, it also offers IoT related business applications along with integration capabilities with other Paas and Saas offering. While the Oracle NoSQL Database can ingest data such as that coming from IoT devices very fast, the Oracle solution lies with the integration with other Oracle products that are designed for Big Data environments, such as Big Data SQL, Big Data Discovery, and the Oracle Database.

Other Oracle Technologies

Oracle offers a wide range of technologies that complete an effective IoT service. These include Oracle Big Data SQL, Oracle Big Data Appliance, Oracle SQL Developer, Oracle Spatial and Graph.

Oracle Big Data SQL supports data filtering by collecting counts, maximum values, minimum values, thresholds, counts above or below user defined values, averages, shorter term averages and counts and longer time averages and counts, sliding SQL window averages and counts and comparisons of each to the other. Once the data is filtered, it can be joined to other database data using Oracle Big Data SQL and then mined inside the Oracle Database using Oracle Advanced Analytics Option. Oracle Advanced Analytics, a priced option to the Oracle Database 12c, delivers scalable, parallelized, in-database implementations of a wide library of workhorse predictive 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 SQL (Oracle Data Mining OAA SQL API component), the Oracle Data Miner “drag and drop” workflow GUI, an extension to Oracle SQL Developer 4.1 and through tight integration with open source R (Oracle R Enterprise R integration component). Because Oracle Advanced Analytics’ in-database data mining machine learning/predictive analytics algorithms are built from the inside out of the Oracle Database and take full advantage of the Oracle Database’s scalability, security, integration, cloud, structured and unstructured data mining capabilities, it makes Oracle the ideal platform for big data + analytics solutions and applications either on-premise or on the Oracle Cloud. Oracle Advanced Analytics turns the database into the ideal

platform for delivering advanced predictions and new insights into BI dashboards, reports and analytical applications.

Summary

When constructing an IoT environment, there are a number of technologies that should be represented to create a complete solution. The determination of which Oracle technologies to deploy should consider the nature and complexity of the data as well as the end user experience.

ORACLE

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

Integrated Cloud Applications & Platform Services

How Oracle NoSQL Database and Oracle Database Mobile Server Enhance an Internet of Things Data Management Solution
December 2016
Schulman/Joshi/Nirhali/Besky/Brey
Contributing Authors: [OPTIONAL]


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