Documento 6.1: Data Lakehouse on OCI

Página 1: Portada
ORACLE
University
Oracle Cloud Infrastructure
Understanding the Data Lakehouse on OCI

Página 2: Perfiles de Casos de Uso

Use Case Profiles

How to increase data productivity and liquidity?

- Data lake & data warehouse

Página 3: Perfiles de Casos de Uso

Use Case Profiles

A new broader strategy...

- Highly accurate machine learning capabilities

- Flexibility of Open Source Services

- Best in class Oracle Database and Data warehouse

Página 4: Perfiles de Casos de Uso

Use Case Profiles

- Common Identity

- Data Integration

- Oracle Lakehouse

- Orchestration

- Catalog

Página 5: Oracle Lakehouse

Oracle Lakehouse

Other Benefits...

- Cost saving

- Improving the agility of data warehouse

Página 6: Oracle Lakehouse

Oracle Lakehouse

The companies who have not adopted OCI yet...

- Reach out to data lakes using Oracle SQL
  Autonomous Database
  Analytics Cloud
  Any SQL App
  Jupyter Notebook
  Python
  REST

Página 7: Introducción al Lakehouse

Introducing the Lakehouse

- Data Warehouse

- Data Lake

- Lakehouse

Include external 3rd party data

Página 8: Cinco Elementos Clave del Lakehouse

Five Key Lakehouse Elements

- **Data Warehouse:** For curated data of known value

- **Data Lake:** For data that is raw

- **Managed Open-Source Services:** Support key open-source tools

- **Data Integration:** Moving data depending

- **Data Catalog:** Maintains a complete view data

Lakehouse

- Data Integration

- Data Warehouse

- Data Lake

- Managed Open-Source Services

- Data Catalog

Página 9: DATA To INNOVATE
DATA

To
INNOVATE
Easy access to all data

Página 10: [DIAGRAMA - Conectar y Curar Datos]

Connecting and Curating Data for Business Outcomes

- **Sources**
  - Operational Databases
  - Enterprise Applications
  - Media
  - D, AC, O
  - Energy Sensors
- **How to run and maintain**
- **Efficiently**
- **Quickly**
- **Securely**
- **Outcomes**
  - Fraud Prevention Customer $60 Recommendations Predictive Maintenance Payment Management Clinical DSS Demand Forecasting Risk Analytics Compliance Safety

Página 11: [DIAGRAMA - Solución para Resultados Deseados]

Connecting and Curating Data for Business Outcomes

- **Sources**
  - Integration
  - Persistence and Processing
  - Analysis
- **Operational Databases**
  - Enterprise Applications
  - Media
  - Events/ Sensors

Solution to get the desired outcomes

- **Outcomes**
  - Fraud Prevention
  - Customer $60
  - Recommendations
  - Predictive Maintenance
  - Payment Management
  - Clinical DSS
  - Demand Forecasting
  - Risk Analytics
  - Compliance Safety

Página 12: [DIAGRAMA - Data Warehouse para Datos Estructurados]

Data Warehouse for Structured Data

Source

- Operational Databases
- Enterprise Applications
- Media
- Events/Sensors

Integration

- Persistence and Processing
- Autonomous Data Warehouse

Analysis

- Allows for storing all types of data

Self Managed

- Analytics
- Platforms
- Automated Data Warehouse

Outcomes

- Fraud Prevention
- Customer 360
- Recommendations
- Predictive Maintenance
- Payment Management
- Clinical DSS
- Demand Forecasting
- Risk Analytics
- Compliance
- Safety

Página 13: [DIAGRAMA - Object Storage para Datos Semi y No Estructurados]

Object Storage for Semi- and Unstructured Data (Data Lake)

Source

- Operational Databases
- Enterprise Applications
- Media
- Events/Sensors

Integration

- Persistence and Processing
- Analysis
- Data Warehouse
- Object Storage Data Lake

Outcomes

- Fraud Prevention
  - Customer 360
  - Recommendations
  - Predictive Maintenance

- Payment Management

- Clinical DSS

- Demand Forecasting

- Risk Analytics

- Compliance

- Safety

Página 14: [DIAGRAMA - Data Lakehouse en OCI]

Data Lakehouse on OCI

Announcing HeatWave for analytics in MySQL

- **Sources**

    - Integration

    - Persistence and Processing

    - Analysis

    - Outcomes

- **Operational Databases**

    - Oracle

    - GoldenGate

    - Autonomous Data Warehouse

    - MySQL

    - Open Source Apache Kafka...

    - Cloud Storage Data Lake

    - MySQL

    - Open Source Spark Space

- **Enterprise Applications**

    - Oracle

    - Data Integration

    - Autonomous Data Warehouse

    - MySQL

    - Open Source Spark Space

- **Media**

- Open Source Apache Kafka...
- Open Source Spark Space
- **Purtriers**
  - Platforms
  - Management Open Source
  - Management Open Source
- **Oracle Analytics Cloud**
  - OCI Data Science
  - Platforms
  - Platforms
- **Fraud Prevention**
  - Customer 360
  - Recommendations
  - Predictive Maintenance
  - Payment Management
  - Clinical DSS
  - Demand Forecasting
  - Risk Analytics
  - Compliance
  - Safety

Página 15: [DIAGRAMA - Catálogo de Datos para Gestión de Metadatos]

Data Lakehouse on OCI

Data Catalog for metadata management

- Sources
  - **Integration**
  - Oracle GoldenGate
  - Operational Databases
    - Oracle Data Integration
  - Enterprise Applications
    - Open source Apache Kafka...

- Media
  - Partners
  - Platforms
  - Cloud Storage Data Lake
  - Cyber Storage Data Lake
  - Schema - ADW - Table
- **Persistence and Processing**
  - **Analysis**
  - Oracle Analytics Cloud
  - OCI Data Science
  - OCI Data Science
  - Partners
  - Outcomes
  - Fraud Prevention
  - Customer 360
  - Recommendations
  - Predictive Maintenance
  - Payment Management
  - Clinical DSS
  - Demand Forecasting
  - Risk Analytics
  - Compliance Safety

Página 16: [DIAGRAMA - Catálogo de Datos para Gestión de Metadatos]

Data Lakehouse on OCI

Data Catalog for metadata management

- Sources
  - Integration
  - Persistence and Processing
  - Analysis
  - Outcomes

- Operational Databases
    - Oracle GoldenGate
    - Autonomous Data Warehouse
    - MySQL Healthcare
    - Oracle Analytics Cloud
    - Fraud Prevention
    - Customer 360
    - Recommendations
    - Predictive Maintenance
    - Payment Management
    - Clinical DSS
    - Demand Forecasting
    - Risk Analytics
    - Compliance
    - Safety
- Enterprise Applications
    - Oracle Data Integration
    - Open source Apache Kafka...
    - Azure
    - Cloud Storage Data Lake
    - Data Catalog
- Media
    - Partners
    - Partner's

Will be able to access

Página 17: Integración del Catálogo de Datos con Autonomous Database

Data Catalog Integrates with Autonomous Database

Autonomous Database both a Data Catalog source and consumer

- Data Catalog is the source of truth for Object Store metadata.

  - Harvest object storage to derive schemas

  - Manage business glossary, terms, and tags

  - Discover data using powerful search

Data Catalog
Object Storage
Autonomous Database(s)

Página 18: [DIAGRAMA - Procesamiento Spark sin Servidor: Data Flow]

Data Lakehouse on OCI

New serverless Spark processing: Data Flow

- **Sources**

  - Integration

  - Persistence and Processing

  - Analysis

  - Outcomes

- **Operational Databases**

  - Oracle

  - GoldenGate

  - Oracle Analytics Cloud

  - Oracle Analytics Cloud

  - Oracle Analytics Cloud

  - Oracle Analytics Cloud

  - Oracle Analytics Cloud

- **Enterprise Applications**

  - Open source

  - Apache Kafka...

  - Open Source

  - Open Source

  - Open Source

  - Open Source

- Open Source
- **Media**
  - Partners
  - Partners
  - Partners
  - Partners
- **Data Flow**
  - Object Storage
  - Data Lake
  - Data Catalog
  - Data Catalog
  - Data Catalog
  - Data Catalog
- **OCI Data Science**
  - OCI Data Science
  - OCI Data Science
  - OCI Data Science
  - OCI Data Science
  - OCI Data Science
- **Fraud Prevention**
  - Customer 360
  - Recommendations
  - Predictive Maintenance
  - Payment Management
  - Clinical DSS
  - Demand Forecasting
  - Risk Analytics
  - Compliance
  - Safety

Página 19: [DIAGRAMA - Procesamiento Spark sin Servidor: Data Flow]

Data Lakehouse on OCI

New serverless Spark processing: Data Flow

Sources

- **Integration**
  - Oracle GoldenGate
  - Operational Databases
    - Oracle Data Integration
  - Enterprise Applications
    - Open source Apache Kafka...
  - Media
    - Partners
  - Events/Sensors

Persistence and Processing

- **Write, decode and execute code**
  - Autonomous Data Warehouse
  - Object Storage Data Lake
  - Managed Open Source
  - Analytics Cloud
  - MySQL HeatWave
  - Data Catalog
  - Partners

Analysis

- **Outcomes**
  - Oracle Analytics Cloud
  - OCI Data Science
  - Partner
  - Platform
  - Software
  - Sales
  - Customer

- Productive Maintenance

- Payment Management

- Clinical DSS

- Demand Forecasting

- Risk Analytics

- Compliance

- Safety

Página 20: [DIAGRAMA - Data Lakehouse en OCI]

Data Lakehouse on OCI

Automated, integrated, and secure for faster insights from all data

Sources

- **Integration**

  - Oracle GoldenGate

  - Operational Databases

    - Oracle Data Integration

  - Enterprise Applications

    - Open source Apache Kafka...

  - Media

    - Partners

  - Events/Sensors

Persistence and Processing

- **Data Lakehouse**

  - Autonomous Data Warehouse

  - MySQL HeatWave

  - Object Storage Data Lake

  - Managed Open Source

  - Azure

  - Data Catalog

  - Partners

Analysis

- **Analysis**

    - Oracle Analytics Cloud

    - OCI Data Science

    - Data Catalog

    - Partners

Outcomes

- **Fraud Prevention**

    - Customer 360

    - Recommendations

    - Predictive Maintenance

    - Payment Management

    - Clinical DSS

    - Demand Forecasting

    - Risk Analytics

    - Compliance Safety

Página 21: Data Lakehouse en OCI

Data Lakehouse on OCI

Open and collaborative approach

- Stores all data

- Accelerate solution development

Páginas 22-24: [Vacías]

Documento 6.2: Oracle Machine Learning Overview

Página 1: Portada
ORACLE
University
Oracle Machine Learning
Overview

Página 2: Solución Basada en la Nube para Analytics
Cloud-Based Solution

for Analytics

Data is consolidated into one system and is available for reporting.

---

1 Frontes

On-premises enterprise data warehouse

Página 3: [CONVERSACIÓN - Escenario de Implementación]

Welcome aboard, Jake.

You're going to help us implement a cloud solution for our analytical requirements.

Steve

Chief Data Scientist

You've made the right decision, Steve.

Oracle Machine Learning would be the perfect solution to deliver what you need.

Jake

Cloud Technology Architect

Página 4: Visión General de OML

Oracle Machine Learning (OML)

Overview

Oracle

Autonomous

Database

Página 5: Colaboración

Collaborate

Oracle AI and ML-Driven Environment

Analyze and Report Deploy

Machine learning
and AI are key to
analytics in every
industry.

Página 6: Flujo Básico de Machine Learning

Basic Machine Learning Flow

- Define project

- Prepare data set

- Build ML model

- Use ML model for scoring

- Deploy model in production

OML: Introduction

Enhanced performance and scalability
Simpler solution architecture and management
Empower a broader range of users with machine learning
Simple pricing structure

Machine Learning: Horizontal Use Cases

Customers

- Segmentation

- Loyalty

- Lifetime value

- Retention/churn

- Acquisition

- Lead generation and prioritization

- Factors of best customers

- Invoice anomaly detection

- Fraud detection

Products

- Next best offer

- Cross-sell

- Upsell

- Product bundling

- Demand forecasting

- Inventory forecasting

- Revenue prediction

Equipment

- Predictive maintenance

- Root cause analysis

- Demand forecasting

- Usage and sensor anomaly detection

- Failure risk assessment

Employees

- Attributes of best employees

- Retention

- Hiring

- Training needs analysis

- Survey feedback sentiment analysis

- Predicting hiring needs

Página 9: [DIAGRAMA - Casos de Uso y Técnicas de ML]
Use Cases and Machine Learning Techniques
Address important business problems that impact customers, products, operations, and employees
Classification
Association Rules
Product cross-sell and upsell
Product bundling
Customer segmentation
Biological species classification
Location-based house value analysis

Página 10: Casos de Uso Verticales por Industria
Industry: Vertical Use Cases
Financial services
Health care and life sciences
Oil and gas
Transportation
Marketing and sales
Government

Páginas 11-13: [Vacías]

Documento 6.3: Data Mesh Architecture

Páginas 1, 3: [Vacías]

Página 2: Compromiso de Oracle con Data Mesh

Oracle Engagement on Data Mesh

- Thought Leadership

- Tools for data products

- Decentralized event-driven architectures

- Streaming patterns

Página 4: Alcance de los Conceptos de Data Mesh

Scope of Data Mesh Concepts

Organizational/Systems

- **Strategy: Data Product Thinking**

  - People

  - Process

  - Technology

  - Design Thinking

  - New

  - Privacy

Design Thinking

- **Jobs to be Done Theory**

  - Business domain oriented – catalog, graph, metadata, etc.

  - Event-driven, real-time, and streaming

Data Architecture

- Anti-monolith/pro-decentralization

- Business domain oriented – catalog, graph, metadata, etc.

- Event-driven, real-time, and streaming

Página 5: [DIAGRAMA - Alcance de Conceptos de Data Mesh]

Scope of Data Mesh Concepts

Organizational/Systems

- **Strategy: Data Product Thinking**

- **Data Architecture**

- Anti-monolith/pro-decentralization

- Business domain oriented – catalog, graph, metadata, etc.

- Event-driven, real-time, and streaming

People
Process
Technology
Reviews
Date
Design Thinking
New
Data Attribution
Protocol
Database Needs
User Needs

Página 6: Capital Digital y Firmas Estelares

Digital Capital and Superstar Firms[1]

25%

of a firm's value is digital capital.

"Digital capital" refers to factors of production that are 1) complementary to IT assets, but 2) are not otherwise recorded on the balance sheet

Digital capital quantifies by market value

Digital capital is more concentrated than any other asset class.

The most value is concentrated in the top decile of firms by market value.

Digital capital accumulation predicts firm productivity three years out.

Página 7: [DIAGRAMA - Capital Digital y Firmas Estelares]

Digital Capital and Superstar Firms[1]

25%

Total clarity

Operational data availability

Faster innovation cycles

Reduction in data engineering

of a firm's value is digital capital.

"Digital capital" refers to factors of production that are 1) complementary to IT assets, but 2) are not otherwise recorded on the balance sheet

Digital capital is more concentrated than any other asset class.

The most value is concentrated in the top decile of firms by market value.

Digital capital accumulation predicts firm productivity three years out.

Página 8: Economía Oculta de Datos

Hidden Data Economy

- Wide variety of shapes and structures
  Applications Sensors Devices

- Demand side
  Analytics AI

Hidden Data Economy

Websites

Mobile Applications

- Informal Data Economy

- Command Data Economy

- Supply

- Demand

Key

- Application

- Analytic, AI

- Supply-Side Data

- Demand-Side Data

- API

- Point-to-Point Integration

Hidden Data Economy

- Market

- Data Economy

- Informal

- Data Economy

- Command

- Data Economy
  Supply

- Key

  - Application

  - Analytic, AI

  - Supply-Side Data

  - Demand-Side Data

  - API

  - Point-to-Point Integration

Página 11: Consecuencias de la Economía Oculta de Datos

What does the hidden data economy lead to?
DEAD
DATA CAPITAL
Trapped in silos
Labor intensive to repurpose
Undiscoverable
Allowed uses unclear

Página 12: Amenazas Gemelas de la Transformación Digital

Twin Threats of Digital Transformation

Loss of competitive advantage
Legal and reputational damage

Página 13: Estrategia de Datos que Refuerza la Estrategia Competitiva

Data Strategy That Reinforces Competitive Strategy
Competitive Strategy
Create unique value _____ Value your customer can only get from you in a unique way _____ Through activities your rivals cannot easily copy
Data Strategy
Create unique data assets _____ Observations only you possess Use them to enhance uniqueness _____ Differentiation, cost position, or both Protect observer and observed _____ Safeguards for all data stakeholders

Página 14: [DIAGRAMA - Sistemas de Actividad]

Activity Systems Tie Data Strategy to Competitive Strategy

- No mask

- Limited passenger service

- No inspires transfers

- No connections with other airlines

- Contact the airline to confirm that the market does not account for any airport

- Finance, reliable experience

- No seat assignments

- Is-minute gate turnaround

- Limited use of travel agents

- Standardized record (72 aircraft)

- Very low ticket prices

- Mobile union contracts

- Afton, helps producing ground and macrones

- High level of employee stock ownership

- High attraction

- Southwest Airlines' Activity System[1] (1996)
  Can we use predictive maintenance to get to 10-minute turnarounds?
  Sensor data
  Engines, subsystems
  Record data
  Aircraft utilization, repair and maintenance, arrivals and departures
  Anonymize, mask employee records
  Monitor model for performance as planes age

Página 15: Productos de Datos como Bloques de Construcción

Data Products, the Building Blocks of Data Strategy

Supply Side

Mostly passive: Data that is something
Data Sets
Sets of observations in different shapes, formats
Models
Domain objects, data models, ML features
Libraries
Inert algorithms, technical definitions of business semantics
A data product is a discrete set of observations and its complementary code designed to fulfil one or more jobs to be done.

Demand Side

Mostly active: Data that does something
Analytics
Reports and dashboards, real-time and historic
Algorithms
ML models, scoring, business rules
Data Services
Payloads, topics, authorization

Página 16: [Vacía]

Página 17: Principios para Traducir Estrategia en Arquitectura

Principles for Translating Data Strategy into Data Architecture

Data Liquidity

Ease of data reuse and recombination[1]

Data Productivity

Value created as a result of data usage per unit of work, dollar invested, or resource consumed

Data Security

Protections against external and internal threats

Data Governance

Assurances of data quality, compliance, ethics

Página 18: Definición de Data Mesh

A Data Mesh Is...

A data solution for enterprise-scale domains and/or event-driven data-centric cloud projects
A data architecture approach focused on outcomes (data products), IT agility (service mesh), and speed (streaming data)

A Data Mesh Is Not...

An alternative point-solution for data warehouse or data lakes

- Data Lake

- A finite project that can be run by a LoB Departmental IT org in isolation

- A single tool or single cloud service that you can buy

Página 19: [Vacía]

Página 20: [DIAGRAMA - Qué es un Data Mesh]

What is a Data Mesh?

A trusted Data Mesh is a data architecture approach focused on outcomes (data products), IT agility in a multicloud world (service mesh), trusted data of all kinds (polyglot data streams), and faster business innovation cycles (event-driven integration).

---

data products
streaming
intuitive, self-service user interfaces for data engineers and data product managers

- event ledgers

- systems of truth

- decentralized/distributed/multicloud/edge

API driven, secure service mesh (KSS, Docker etc) that runs anywhere

- data product KPIs, SLAs | business domain modeling | provenance & explainability

Página 21: [DIAGRAMA - Blueprint de Data Mesh Basado en OCI]

OCI-Based Data Mesh Blueprint

Data Producers
- Edge Telemetry (Devices & Things)

- App Events (Bit: Process & Logging)

- Database Events (ACID Transactions)

- Bulk Data Sources (DBs, Files, etc.)

- Customer Choice

Transient Zone
- [Lista extensa de componentes Oracle numerados del 1 al 522 - probablemente un marcador de posición para diagrama]

Páginas 22-23: [Vacías]