



# Inovação com dados em nuvem

TRILHA



## Oracle Analytics Cloud Guia para Laboratório Hands-On

Armando Moré Neto

03.12.20 10h10



# Inovação com dados em nuvem

TRILHA

#TheDevConf  
Oracle



Este trabalho está licenciado sob uma Licença Creative Commons Atribuição-Compartilhável 4.0 Internacional. Para ver uma cópia desta licença, visite <http://creativecommons.org/licenses/by-sa/4.0/>.



# Inovação com dados em nuvem

ACELERANDO O SEU NEGÓCIO COM ORACLE  
AUTONOMOUS DATA WAREHOUSE E  
ORACLE ANALYTICS CLOUD



# Guia para Laboratório *Hands-On*

<b>Introdução .....</b>	<b>5</b>
<b>Lab 1. Acessando seu ambiente .....</b>	<b>6</b>
<b>Lab 2. Criando uma instância de Autonomous Data Warehouse e Oracle Analytics Cloud .....</b>	<b>10</b>
Iniciando a criação da Instância de Autonomous Data Warehouse .....	11
Iniciando a criação da Instância de Oracle Analytics Cloud .....	16
<b>Lab 3. Explorando o Autonomous Data Warehouse .....</b>	<b>19</b>
Overview do Autonomous Data Warehouse .....	20
Acessando os detalhes do Autonomous Data Warehouse .....	21
Overview da Service Console do Autonomous Data Warehouse.....	23
Executando queries no Autonomous Data Warehouse.....	27
Carregando dados no Autonomous Data Warehouse.....	32
<b>Lab 4. Explorando e criando um projeto no Oracle Analytics Cloud .....</b>	<b>34</b>
Overview do Oracle Analytics Cloud .....	35
Criando uma conexão com o banco de dados .....	37
Criando uma base de dados e explorando o “Data Wrangling” .....	38
Criando um projeto no Data Visualization .....	44
Exportando os dados.....	55
<b>Lab 5. Explorando o Data Flow e o Machine Learning .....</b>	<b>60</b>
Conhecendo o Data Flow .....	61
Criando um modelo de Machine Learning no Autonomous Data Warehouse .....	66
Criando um modelo de Machine Learning no Oracle Analytics Cloud .....	78
<b>Lab 6. Comentários gerais sobre a versão Enterprise e Administração do OAC.....</b>	<b>85</b>
Versão Enterprise: Data Modeler .....	85
Versão Enterprise: Oracle Business Intelligence Enterprise Edition (OBIEE).....	85
Explorando o ADM Console.....	87
<b>Lab 7. Terminando e apagando instâncias.....</b>	<b>90</b>

# Introdução

Neste laboratório prático, vamos trabalhar na criação de uma instância de Oracle Analytics Cloud e uma instância de Autonomous Data Warehouse seguindo processos e boas práticas de implementação.

Exploraremos alguns dos recursos disponíveis no Oracle Analytics Cloud e no Autonomous Data Warehouse que são Plataformas como Serviço (PaaS).

Ferramentas contempladas pelo Oracle Analytics Cloud:

- Data Visualization: Criação de visualizações interativas, ferramenta para transformações de bases de dados, análises preditivas, *insights* e *forecasts*.
- Oracle Business Intelligence Enterprise Edition (OBIEE): Criação de dashboards interativos, capacidade de gerar relatórios automáticos que são enviados por e-mail, acesso ao Oracle Day by Day que é um aplicativo para smartphones que entrega insights para o usuário e modelagem e gerenciamento de acessos as informações via Data Modeler.

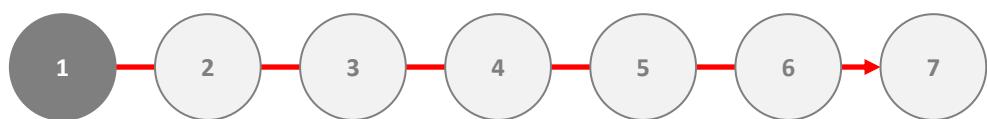
Ferramentas contempladas pelo Autonomous Data Warehouse:

- Data Visualization Desktop: Versão desktop do Data Visualization mencionado acima. Fornecido gratuitamente para o usuário que tem uma instância de Autonomous em sua Nuvem.
- SQL Developer Web: Versão Web do conhecido SQL Developer da Oracle.
- Oracle APEX: O Oracle Application Express (APEX) é uma plataforma de desenvolvimento de baixo código que permite criar aplicativos corporativos escaláveis e seguros, com recursos de classe mundial, que podem ser implantados em qualquer lugar.
- Oracle Machine Learning: O Oracle Machine Learning é uma interface de notebooks SQL para que os cientistas de dados realizem o aprendizado de máquina no ADW (Oracle Autonomous Data Warehouse). Baseado em Apache Zeppelin.

É importante que os conceitos fundamentais desses recursos estejam claros para uma boa experiência em nossa nuvem.

# Lab 1.

## Acessando seu ambiente



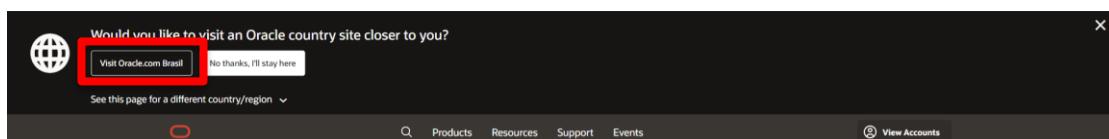
# Lab 1. Acessando seu ambiente

## Objetivos

- Acessar o console da Oracle Cloud
- Conhecer os serviços de infraestrutura e plataforma
- Familiarizar-se com o ambiente

Nesta seção você aprenderá mais sobre o acesso inicial ao ambiente.

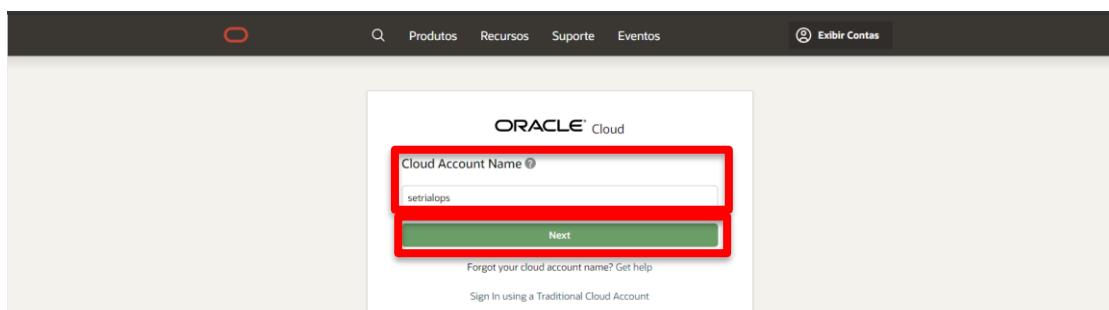
Vá para [oracle.com](https://oracle.com). Você pode alterar o idioma dessa página antes do acesso ao ambiente:



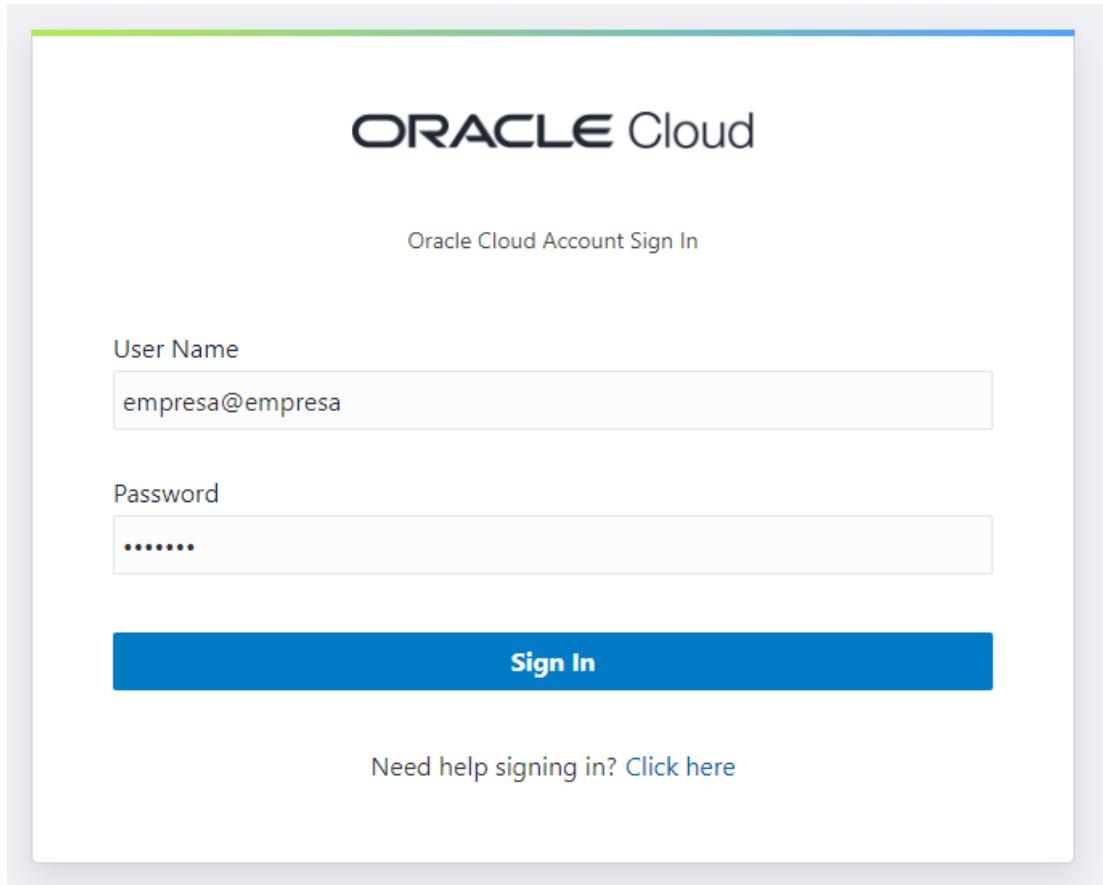
No site já em português, clique em Exibir Contas e depois em Faça Login na Nuvem:



O login deve ser feito com o “Cloud Account Name”, onde somente é necessário informar o **nome da conta** (definido no momento de solicitação do trial ou do ambiente final).



A tela de login para usuário é apresentada. O usuário administrador é identificado pelo e-mail utilizado no cadastro do ambiente.



No primeiro acesso é solicitado que sua senha seja alterada.

A tela principal do seu ambiente é apresentada. Nela, você consegue ver algumas ações rápidas para a criação de alguns recursos, alguns artigos de soluções dentro da nuvem da Oracle que possa ajudar, a parte de Learn que leva para a documentação, que é muito bem detalhada.

Na barra superior tem-se o menu que lista todas as abas da console da nuvem, a lupa para pesquisas no ambiente de nuvem, informação de qual a região que está sendo visualizada, no caso da imagem abaixo está sendo visualizado a região de Ashburn, o sino é aonde é feito os anúncios relacionados a nuvem, na interrogação é aonde tem alguns tópicos de ajuda e também onde é possível entrar em contato com o suporte ou abrir um chamado para aumentar os limites de serviço da nuvem, no mundo é aonde o usuário consegue mudar o idioma da console da nuvem e por fim no símbolo de usuário o mesmo pode encontrar as informações dele.

**Quick Actions**

- COMPUTE Create a VM instance (2-4 min)
- AUTONOMOUS TRANSACTION PROCESSING Create a database (2-3 min)
- AUTONOMOUS DATA WAREHOUSE Create a data warehouse (2-3 min)
- NETWORKING Create a virtual cloud network (2-3 min)
- OBJECT STORAGE Store data (2-3 min)
- NETWORKING SOLUTIONS Create an IPsec VPN connection (2-3 min)

**Solutions**

- Jump Start**: Hands-on use of Oracle Cloud Infrastructure with self-paced learning and demo labs, for free.
- Mission Critical Databases**: Leverage a range of modern cloud data management options, with 100% on-premises compatibility.
- Resource Manager**: Automate the provisioning of your cloud infrastructure resources using HashiCorp Terraform.
- Big Data and Analytics**: Deploy popular technologies like Cloudera, DataStax, Confluent, and Hadoop on the fastest cloud infrastructure.

**Learn**

Adding users and groups	Understanding compartments	Launching your first instance
Understanding service limits	Key concepts and terminology	Overview of Autonomous Database

**Action Center**

- All systems operational (View health dashboard)
- User Management: Add a user to your tenancy
- Billing: Total credits used: RRS\$375,39 / RRS\$1,100,00
- Total days elapsed: 15 / 30
- Log out / Log me out of my account

**What's New**

- Oracle and Microsoft announce deeper integration to support multi-cloud deployments (May 1, 2019)
- Start demand-side auto scaling and on-demand scaling for Autonomous Database (May 4, 2019)
- Add new spatial intelligence support for your Autonomous Data Warehouse (May 5, 2019)
- Leverage Border Gateway Protocol (BGP) for dynamic routing for IPsec VPNs (May 7, 2019)
- Streamline validation deployments by using Oracle Cloud Infrastructure Identity and Access Management for Kubernetes (May 29, 2019)

**Get Help**

- Contact Support
- Developer Tools
- Documentation

Copyright © 2019, Oracle and/or its affiliates. All rights reserved.

# Lab 2.

## Criando uma instância de Autonomous Data Warehouse e Oracle Analytics Cloud



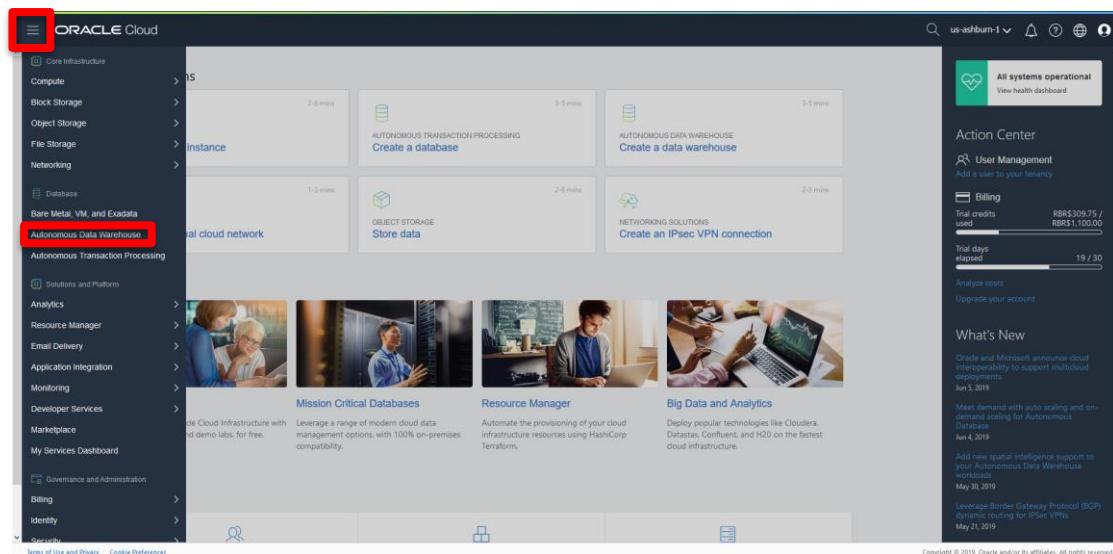
# Lab 2. Criando uma instância de Autonomous Data Warehouse e Oracle Analytics Cloud

## Objetivos

- Provisionar uma instância de Autonomous Data Warehouse de forma rápida
- Compreensão geral do Autonomous
- Provisionar uma instância de Oracle Analytics Cloud de forma rápida
- Compreender a diferença de edições

## Iniciando a criação da Instância de Autonomous Data Warehouse

Após feito o login no ambiente, serão acessados o menu e a opção Autonomous Data Warehouse, conforme a imagem abaixo.



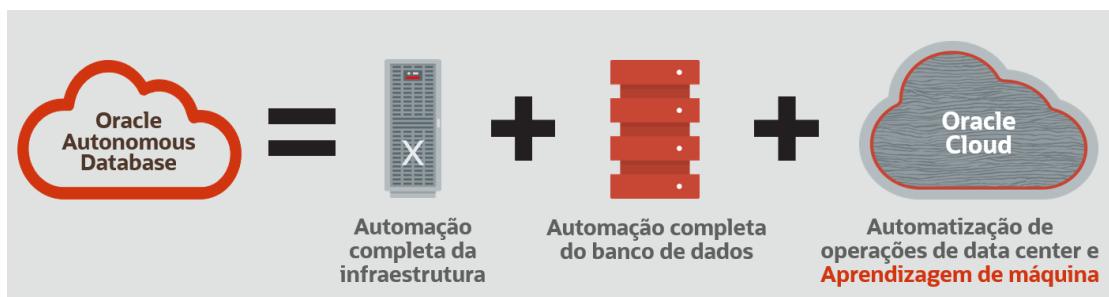
Feito o passo acima a tela abaixo aparecerá.

Feito o acesso, a primeira coisa que será feita é selecionar um Compartimento, no caso da imagem acima será acessado o compartimento “Armando\_Neto”, não é recomendado criar nenhum recurso no compartimento root, para a criação de um compartimento basta ir no menu, Identity e Compartments e criar um compartimento, mais detalhes neste link:

<https://docs.cloud.oracle.com/iaas/Content/Identity/Tasks/managingcompartments.htm>

Feito isso, como pode ser observado na imagem, é possível filtrar os bancos de dados autônomo por seu tipo de carga de trabalho e qual seu estado, ativo ou inativo por exemplo.

Além da possibilidade de criar uma instância de Autonomous *serverless*, que é a maneira que vai ser trabalhada neste laboratório, há a possibilidade de reservar um Autonomous Exadata Infrastructure e feito isso criar o Autonomous Container Database (CDB de Autonomous, é uma *feature* do banco Oracle chamada *Multitenant* introduzida na versão 12c), reservando o hardware e instanciando o CDB do Autonomous o usuário vai criar seu próprio banco Autonomous, feito isso, o usuário começa a criar os PDBs (Pluggable Databases) e como pode ser observado o Autonomous é um banco Oracle totalmente “stackado” como ilustra a imagem abaixo.



Como já mencionado, neste laboratório vamos trabalhar com o conceito *serverless* do Autonomous, clique no botão marcado na imagem “Create Autonomous Database”.

Aparecerá uma lista de informações que devem ser preenchidas para a criação deste banco conforme as imagens abaixo.

The screenshot shows the Oracle Cloud interface for creating an Autonomous Database. It consists of several stacked sections:

- Create Autonomous Database**:
  - Provide basic information for the Autonomous Database**:
    - Choose a compartment: Armando\_Neto
    - Display name: ORCLADW
    - Database name: ORCLADW
    - The name must contain only letters and numbers, starting with a letter. Maximum of 14 characters.
  - Choose a workload type**:
    - Data Warehouse**: Configures the database for a decision support or data warehouse workload, with a bias towards large data scanning operations. (Selected)
    - Transaction Processing**: Configures the database for a transactional workload, with a bias towards high volumes of random data access.
  - Configure the database**:
    - CPU core count: 1
    - Storage (TB): 1
    - The number of CPU cores to enable. Available cores are subject to your tenancy's service limits.
    - The amount of storage to allocate.
  - Create administrator credentials**:
    - New Database Preview Version 19c Available ⓘ
    - Enable Preview Mode
    - Username: READ-ONLY (ADMIN)
    - Password: (empty)
    - Confirm password: (empty)
  - Choose a license type**:
    - Bring Your Own Licence**: My organization already owns Oracle database software licenses. Bring my existing database software licenses to the database cloud service. (details) (Selected)
    - License Included**: Subscribe to new database software licenses and the Database cloud service.

[Hide Advanced Options](#)

**Tags**

Tag Namespace

No namespace (Free-Form tag)

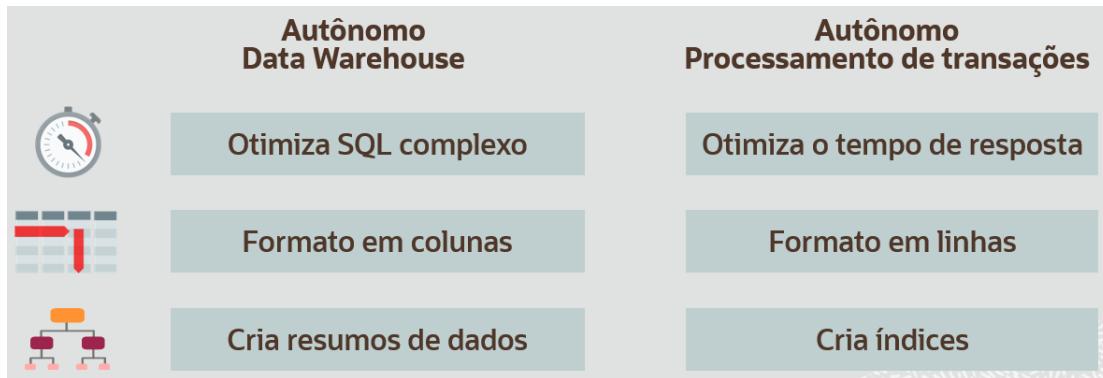
Key

Value (Optional)

+ Additional Tag

[Create Autonomous Database](#)

- Choose a compartment: Escolha o compartimento que deseja criar a instância do Autonomous, não é recomendado criar no root;
- Display Name: Aparecerá um nome gerado pela própria nuvem, mas pode ser colocado qualquer nome, será o nome dessa instância para a plataforma da nuvem. No caso da imagem foi colocado “ORCLADW”;
- Database name: Nesta opção será o nome do banco (SID do banco), só pode conter letras e números e tem que começar o nome com uma letra. No caso da imagem foi colocado “ORCLADW”;
- Choose a workload type: Aqui há duas opções, a Data Warehouse e a Transaction Processing, aqui o usuário escolhe qual Autonomous ele deseja, ele foca o banco para aprender e melhorar a performance ou de transações ou de queries analíticas, no caso deste laboratório será o workload de Data Warehouse. A imagem abaixo explica melhor as diferenças entre os workloads;



- Configure the database: Nesta parte será definido qual o poder de processamento e a quantidade de armazenamento será alocado, ambos são escaláveis tanto em processamento quanto em armazenamento, sem *downtime* para escalar, logo é recomendado começar com o mínimo que seria 1 OCPU de processamento e 1 Terabyte de armazenamento e ir escalando conforme vai crescendo as bases e o processamento;
- Auto scaling: Permitir o Auto scaling permite que o Autonomous multiplique por 3 o número de OCPU's que está inicialmente alocado a

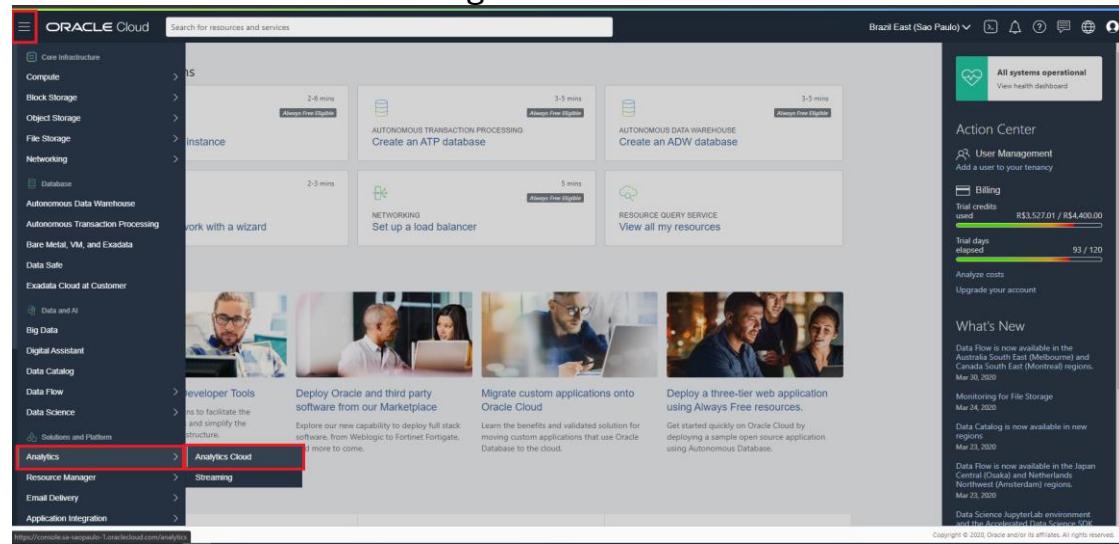
ele, por exemplo se for 1, ele irá escalar até 3, se for 2 ele vai escalar até 6 OCPU's, e assim sucessivamente. Ele irá escalar conforme ele percebe um aumento de carga, e ele escala gradualmente até atingir o seu máximo que é 3 vezes o inicialmente alocado, procurando sempre a melhor eficiência do processamento. Pode ser ativado e desativado a qualquer hora. O número de OCPU's que ele escalou será cobrado a mais pelo tanto de horas que ele ficou com mais OCPU's do que o normalmente alocado;

- New Database Preview Version 19c Available: Selecionando esta opção, ao invés do banco Autonomous usar a versão 18c como base ele usará a 19c;
- Create administrator credentials: O usuário com mais privilégios no Autonomous é o “ADMIN” e não pode ser alterado este Username, podendo criar mais usuários posteriormente quando a instância estiver pronta como qualquer outro banco, além disso a senha colocada aqui será a utilizada mais para a frente para acessar o banco, ela deve conter de 12 a 30 caracteres, uma maiúscula, uma minúscula e um número, não pode conter aspas ou escrito “admin”;
- Choose a license type: Aqui há duas opções de licença, são elas:
  - License Included: É a licença da nuvem que será emprestada ao usuário já com suporte incluso e o mesmo paga preço de lista padrão;
  - Bring Your Own License (BYOL): Nesta modalidade o usuário traz a licença de banco Oracle local dele para a nuvem para pagar um preço diferenciado, para isto estar sempre em vigor o suporte da licença deve sempre estar ativo, mais detalhes podem ser encontrados neste link:  
[https://cloud.oracle.com/en\\_US/datawarehouse/pricing](https://cloud.oracle.com/en_US/datawarehouse/pricing) ou nos chats de vendas da Oracle;
- Tags: Utilizado para monitoramento de custos de recursos ou projetos.

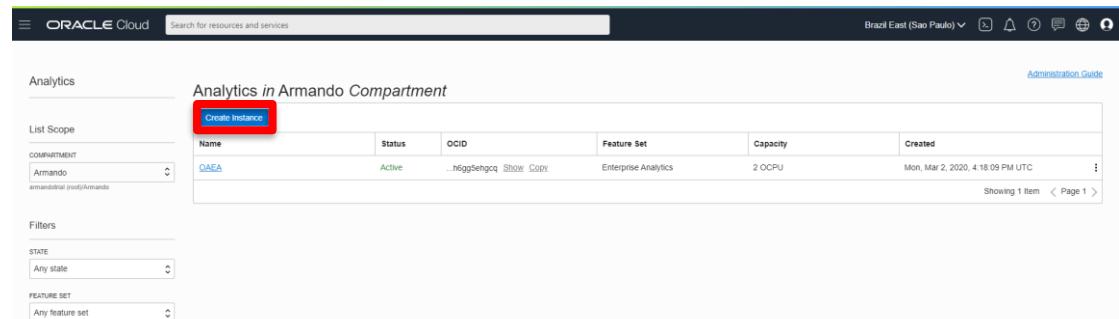
Após preenchido tudo clique em “Create Autonomous Database”, este processo de criação da Instância *serverless* do banco Autonomous demora por volta de 5 minutos.

# Iniciando a criação da Instância de Oracle Analytics Cloud

A partir da tela inicial do seu ambiente Oracle Cloud. Vá para o menu no canto superior esquerdo, clique nele, vá até Analytics e Selecione Analytics Cloud. Conforme demonstrado na imagem abaixo.



Feito o primeiro passo, clique em “Criar Instância” para abrir uma lista com as informações necessárias para a criação da Instância de Oracle Analytics Cloud. Conformas duas imagens abaixo



Create Instance

[Help](#) [Close](#)

COMPARTMENT

Armando

INSTANCE NAME [\(i\)](#)

DESCRIPTION OPTIONAL [\(i\)](#)

FEATURE SET [\(i\)](#)

Enterprise Analytics

CAPACITY [\(i\)](#)

OCPU

1 - non production

LICENSE TYPE [\(i\)](#)

MY ORGANIZATION ALREADY OWNS ORACLE MIDDLEWARE SOFTWARE LICENSES. BRING MY EXISTING MIDDLEWARE SOFTWARE LICENSE TO THE ANALYTICS CLOUD

SUBSCRIBE TO A NEW ANALYTICS CLOUD SOFTWARE LICENSE AND THE ANALYTICS CLOUD

Tagging is a metadata system that allows you to organize and track resources within your tenancy. Tags are composed of keys and values that can be attached to resources.

[Learn more about tagging](#)

TAG NAMESPACE	TAG KEY	VALUE
None (add a free-form tag)		

+ Additional Tag

**Create**

É apresentada uma seção onde se deve preencher algumas informações:

- Compartimento: Compartimento que é desejado criar a instância, não é recomendado criar no compartimento raiz.
- Nome da instância: o nome é um texto livre.
- Descrição: é opcional e é um texto livre.
- Feature Set: temos 2 feature sets que podem ser escolhidos: a Self Service Analytics contempla o Data Visualization e a versão Enterprise Analytics contempla o Data Visualization mais o Oracle Business Intelligence.
  - Data Visualization: ferramenta self-service de visualização de dados e preparação de dados, podendo ver o negócio de várias fontes de dados inclusive planilhas em .xlsx, .txt ou .csv.
  - OBIEE: ferramenta utilizada para a criação de modelos de dados, dashboards interativos, automatização de relatórios e envio automático de e-mails, além de permitir a utilização do aplicativo de celular Oracle Day by Day.
- Capacity: número de OCPUs que será dedicado à instância a ser criada, o número de OCPU's pode ser futuramente alterado, podendo

aumentar ou diminuir, o mínimo recomendado para um ambiente produtivo é 2 OCPUs.

- License Type: caso o usuário tenha uma licença e o suporte da mesma esteja ativo, ele pode selecionar a opção que a organização dele detém uma licença, com isso o valor cobrado por OCPU por hora é reduzido; caso ele não tenha, ele pode escolher a opção de se subscrever em uma licença na nuvem e o valor de lista é cobrado.
- Tag: é um campo opcional que pode ser inserido uma *tag* para rastreio dos custos ou relacionar a um projeto por exemplo.

Clique em criar e o processo de criação da instância do Oracle Analytics Cloud começará, o processo dura aproximadamente 20 minutos.

Como informativo, é importante entender o conceito de OCPU. A Oracle informa em seu [FAQ](#) o seguinte texto:

*Define-se OCPU como a capacidade da CPU equivalente a um núcleo físico de um processador Intel Xeon com Hyper threading habilitado, ou um núcleo físico de um processador Oracle SPARC. Para o processador Intel Xeon, cada OCPU corresponde a dois threads de execução de hardware, conhecidos como vCPUs. Para processadores Oracle SPARC, uma OCPU corresponde a oito threads de execução de hardware, também conhecidos como vCPUs.*

# Lab 3.

# Explorando o Autonomous Data Warehouse



# Lab 3. Explorando o Autonomous Data Warehouse

## Objetivos

- Overview do Autonomous Data Warehouse
- Conhecendo a console dos detalhes do Autonomous Data Warehouse
- Service Console

## Overview do Autonomous Data Warehouse

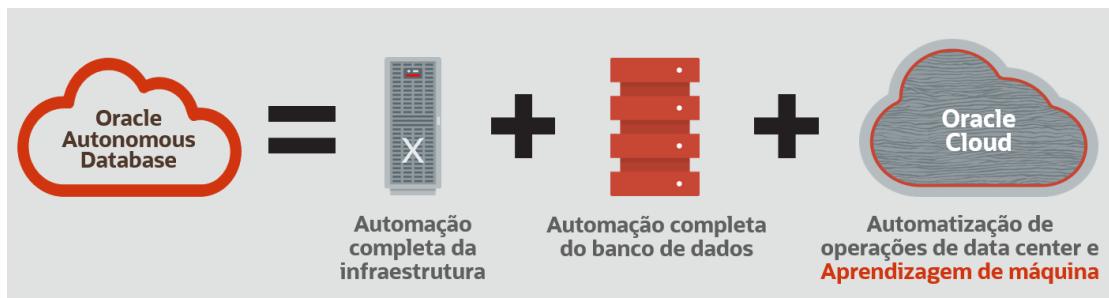
O Oracle Autonomous Data Warehouse fornece um banco de dados fácil de usar e totalmente autônomo que é escalável, oferece rápido desempenho de consultas e não exige administração de banco de dados, tudo isso sem **downtime**.

O Autonomous ele é independente, um serviço de data warehouse em nuvem totalmente gerenciado. Configuração de rede, armazenamento, aplicação de patches e upgrade em bancos de dados, além de outros serviços, são feitos pela Oracle.

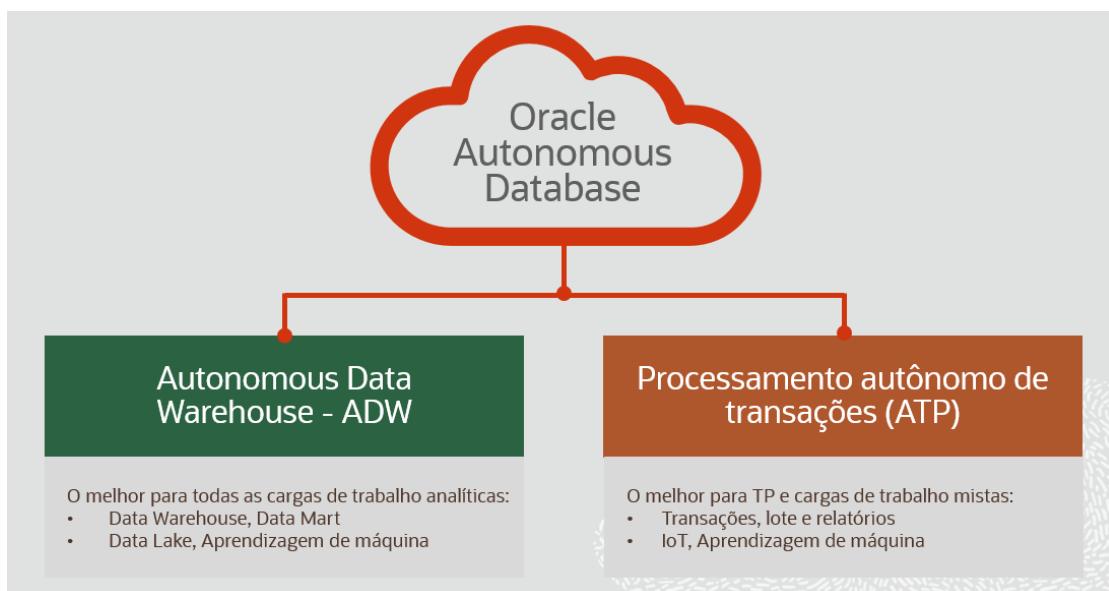
Ele se auto protege, sempre executa os patches de segurança mais recentes. Os dados armazenados são criptografados usando a criptografia TDE (Transparent Data Encryption). Os clientes de banco de dados usam SSL/TLS 1.2 criptografado e conexões mutuamente autenticadas.

Ele se autor repara, proteção automatizada contratempo de indisponibilidade. A alta disponibilidade está presente em cada componente, como por exemplo na hora de escalar ou de se aplicar um *patch* ele não terá *downtime*, e os backups são totalmente automatizados, podendo ser realizados backups manuais também.

Como já mencionado neste material, o Autonomous é um produto Oracle “stackado”, ele está em uma infraestrutura Oracle que se chama Exadata que é desenhada para melhorar a performance de bancos Oracle, o sistema operacional deste hardware é o Oracle Linux com o Unbreakable Kernel que também é um sistema operacional otimizado para banco de dados e instalado neste sistema está o banco Oracle 18c, tudo isso aliado as automações e o Machine Learning da nuvem compõem o Autonomous Data Warehouse.



Além disso o Autonomous é uma família, o Data Warehouse é focado para processamentos analíticos e o Transaction Processing para transações e cargas de trabalho mistas, ótimo para desenvolvimento de aplicações que usam ele como base, segue abaixo uma imagem que representa essas soluções. Neste laboratório será abordado apenas o Autonomous Data Warehouse



## Acessando os detalhes do Autonomous Data Warehouse

Será acessado os detalhes do Autonomous que foi criado no passo anterior deste tutorial, primeiro clique no link conforme a imagem abaixo.

The screenshot shows the Oracle Cloud interface for managing Autonomous Databases. On the left, there's a sidebar with options like 'Autonomous Database', 'Autonomous Container Database', and 'Autonomous Exadata Infrastructure'. Below that are 'List Scope' and 'COMPARTMENT' dropdowns set to 'Armando\_Neto'. Under 'Filters', there are dropdowns for 'WORKLOAD TYPE' (set to 'All') and 'STATE' (set to 'Any state'). The main area displays a table of databases. One row is highlighted with a red box: 'Name' is 'ORCLADW', 'Database Name' is 'ORCLADW', 'State' is 'Available' (green dot), 'Dedicated Infrastructure' is 'No', and 'CPU Core Count' is '1'.

Feito o passo acima, será acessado a tela da imagem abaixo.

This screenshot shows the detailed view for the 'ORCLADW' database. At the top, there's a large green button with 'ADW' and 'AVAILABLE' text. The main content area has tabs for 'Autonomous Database Information' (selected) and 'Tags'. Under 'General Information', details include: Database Name: ORCLADW, Workload Type: Data Warehouse, Compartiment: pocarmando (root)/Armando\_Neto, OCID: ...wacoxa, Storage: 100 GB, Created: Mon, 22 Jul 2019 20:51:06 GMT, CPU Core Count: 1, Storage (TB): 1, License Type: Bring Your Own Licence, Database Version: 18c, Auto Scaling: Enabled, and Lifecycle State: Available. The 'Infrastructure' section shows Dedicated Infrastructure: No. In the 'Backups' section, it says Backups are automatically created daily, and there's a table showing one backup entry: Name: Jul 25, 2019 13:46:45 UTC, State: Active, Type: Incremental, initiated by Auto Backup, Started: Thu, 25 Jul 2019 13:46:41 GMT, Ended: Thu, 25 Jul 2019 13:46:45 GMT. The bottom of the page includes links for 'Terms of Use and Privacy' and 'Cookie Preferences'.

Nesta tela é possível ter as informações gerais do banco Autonomous e ver os backups que o próprio banco criou automaticamente, as regras automáticas dele é fazer backups diários incrementais com retenção de 60 dias, atualmente a retenção não pode ser alterada mas para mitigar isso o usuário pode criar os seus backups manuais e armazená-los em um Object Storage.

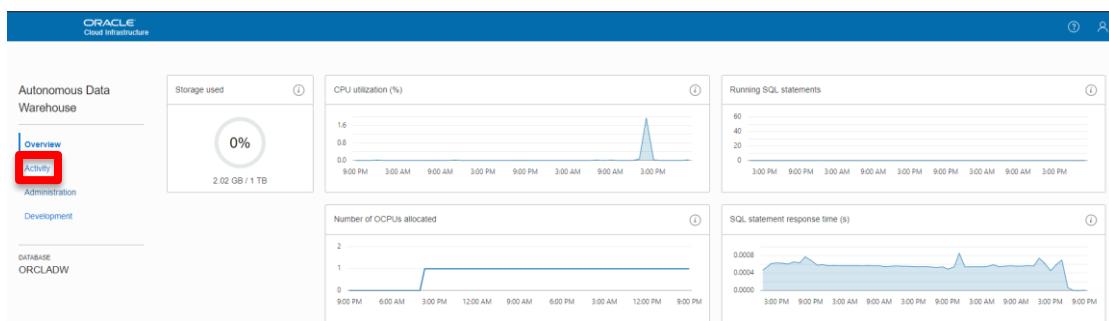
Pode-se observar que tem alguns botões no console também, e as ações deles são:

- DB Connection: Mostra uma janela com as Strings de conexão (High, Medium, Low) e permite que o usuário faça download das Credenciais do Cliente
- Performance Hub: Abre uma aba que monitora como está a performance no banco
- Service Console: Leva o usuário para a console de serviços, será explorado mais afundo no próximo tópico deste tutorial.
- Scale Up/Down: Abre uma janela que permite que o usuário escale o banco, tanto para cima quanto para baixo, tanto o armazenamento quanto o poder de processamento e todas essas alterações são independentes e **sem downtime**. Permite também o usuário ativar o Auto Scaling, *Feature* do banco que faz com que ele automaticamente aumente seu poder computacional em até 3 vezes, quando necessário, baseado no número de OCPU's que ele estava sem nenhum aumento automático.
- Stop: Para o serviço do banco e por consequência para o consumo de créditos relacionado ao poder computacional do banco, a cobrança do armazenamento continua.
- Actions
  - Restore: Restaura um backup do banco
  - Create Clone: Cria um clone do banco
  - Access Control List: Limita o acesso ao banco via um bloco de IPs (CIDR)
  - Admin Password: Altera a senha de ADMIN.
  - Update License Type: O usuário pode alterar o time de licença do banco sem precisar criar outro ou parar o mesmo.
  - Move Resource: Move o Autonomous para outro compartimento dentro do mesmo ambiente.
  - Apply Tag(s): Permite a inserção de tags no recurso para monitorar custos de projeto e afins.
  - Terminate: Deleta por completo o banco (A instância e o armazenamento).

## **Overview da *Service Console* do Autonomous Data Warehouse.**

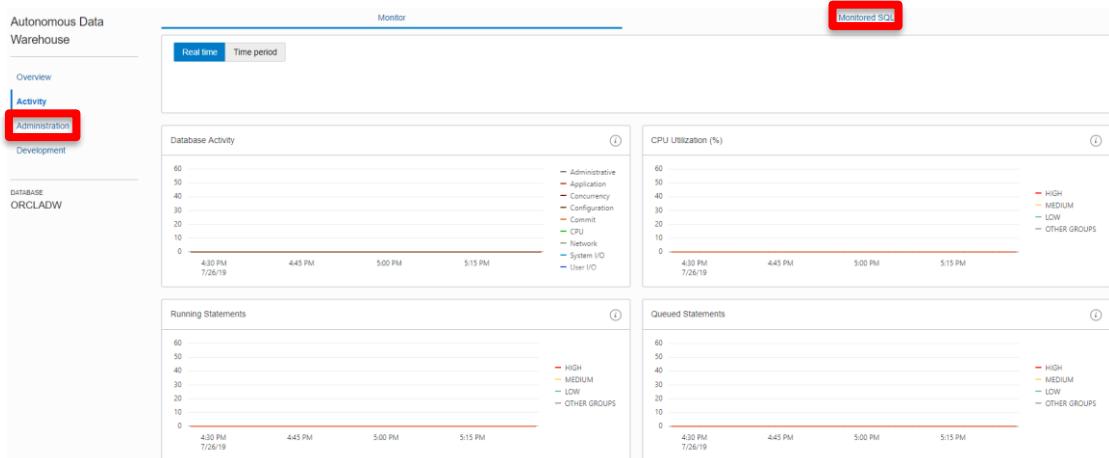
Após o overview dos detalhes do banco, será explorado a console de serviços do banco, para acessá-la, clique no botão conforme a imagem abaixo.

Feito o passo acima, o usuário irá acessar a seguinte console.



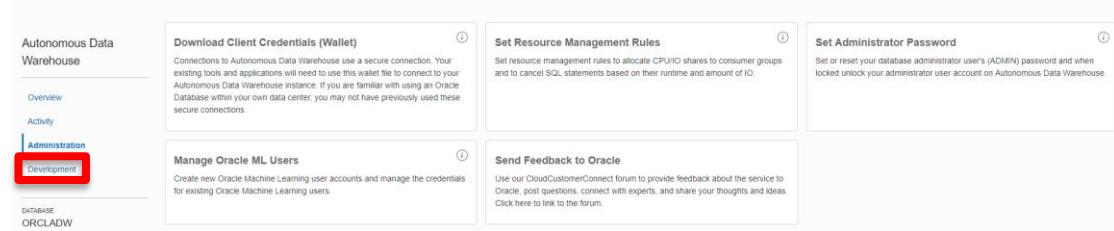
A console que aparece inicialmente logo que o usuário acessa a Service console é a de Overview, nela o usuário verifica como o Autonomous está sendo utilizado em poder computacional e armazenamento, quantas OCPU's alocadas e como está sua performance executando as queries.

Após a tela de overview será acessado a tela de “Activity”, indicada na imagem acima.



Na tela de Activity o usuário monitora as atividades do Autonomous em tempo real, e pode chegar no nível de detalhe da query quando acessar a aba “Monitored SQL” indicada na imagem.

Após a familiarização, será acessado a tela “Administration” indicada na imagem acima.



Nesta parte da console o usuário pode fazer o download das credenciais do Autonomous, da mesma maneira como mencionado anteriormente neste tutorial, o administrador do banco pode reconfigurar as regras dos recursos dedicados para cada tipo de conexão ou voltar as mesmas para o padrão, como já mencionado o Autonomous tem três maneiras de se conectar, elas são:

- “High”: Conexão que permite paralelismo de queries, tem altos recursos dedicados para essa conexão e baixa concorrência.
- “Medium”: Conexão que permite paralelismo de queries, menos recursos dedicados e mais concorrência entre os processamentos.
- “Low”: Conexão que executa as queries em série, logo utiliza do paralelismo de queries existente no banco, poucos recursos alocados e alta concorrência entre os processos.

Ainda nesta console o usuário pode alterar a senha do login ADMIN do Autonomous, criar usuários do Oracle Machine Learning, que é uma ferramenta que permite os usuários executarem scripts de Machine Learning em SQL ou PL/SQL dentro do Autonomous utilizando do Machine Learning do mesmo, está em desenvolvimento a execução de scripts em R e Python, acessado via interface web é baseado em Apache Zeppelin.

E por fim, o usuário pode enviar um feedback para a Oracle, via aquele link, ele será redirecionado para um fórum onde ele não só pode enviar feedbacks como também compartilhar suas dúvidas e casos de sucesso.

Agora será acessado a tela de Development.

The screenshot shows the Oracle Autonomous Data Warehouse console under the 'Development' tab. It lists three main tools:

- APEX**: Provides a low-code development environment for building apps in a single, extensible platform.
- SQL Developer Web**: A browser-based integrated development environment and administration interface for Oracle Autonomous Database.
- OML Notebooks**: Oracle SQL notebooks provide easy access to Oracle's parallelized, scalable in-database implementations of a library of Oracle Advanced Analytics' machine learning algorithms.

Below these, there is a section for 'Download Oracle Instant Client' which includes libraries for Node.js, Python, and PHP, and provides access to OCI, OCCI, JDBC, ODBC, and Pro\*C applications.

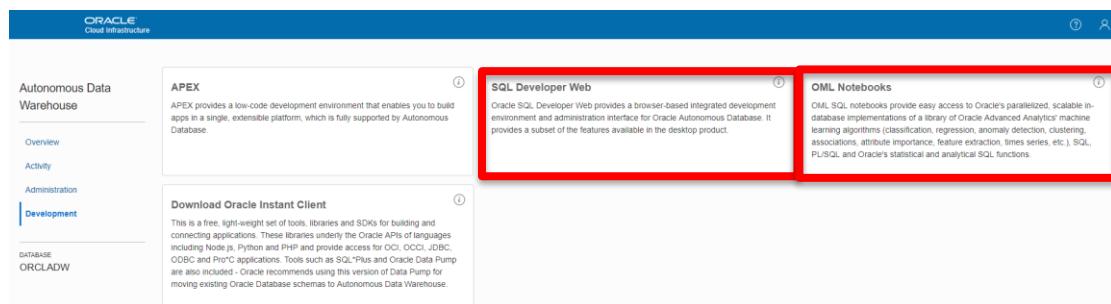
Nesta parte da console o usuário consegue acessar ferramentas de desenvolvimento no Autonomous, todas baseadas em Web, o usuário pode acessar:

- **APEX:** O Oracle Application Express (APEX) é uma plataforma de desenvolvimento low-code que permite criar aplicativos incríveis, escaláveis em uma única plataforma que podem ser implementados em qualquer lugar, que é totalmente suportado pelo Banco de Dados Autonomous.
- **SQL Developer Web:** Provê um ambiente integrado de desenvolvimento e administração baseado em navegador web para o Oracle Autonomous. Ele provê uma parte das ferramentas disponíveis no produto desktop.
- **Oracle Machine Learning:** Os notebooks Oracle Machine Learning SQL fornecem acesso fácil às implementações escalonáveis e paralelizadas da Oracle em bancos de dados de uma biblioteca de algoritmos de aprendizado de máquina do Oracle Advanced Analytics (classificação, regressão, detecção de anomalias, agrupamentos, associações, importância de atributos, extração de recursos, séries temporais etc.), SQL, PL / SQL e funções SQL analíticas e estatísticas da Oracle.
- **Download Oracle Instant Client:** O Oracle Instant Client permite que os aplicativos se conectem a um Banco de Dados Oracle local ou remoto para desenvolvimento e implementação de produção. As bibliotecas do Instant Client fornecem a conectividade de rede necessária, bem como recursos de dados básicos e de alto nível, para aproveitar ao máximo o Oracle Database. Ele é a base das APIs Oracle de linguagens e ambientes populares, incluindo Node.js, Python e PHP, além de fornecer acesso para aplicativos OCI, OCCI, JDBC, ODBC e Pro \* C. As ferramentas incluídas no Instant Client, como o SQL \* Plus e o Oracle Data Pump, fornecem acesso rápido e conveniente aos dados.

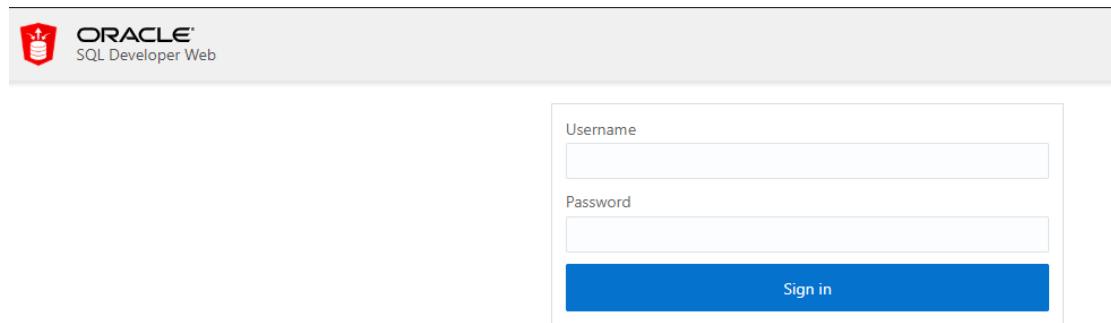
# Executando queries no Autonomous Data Warehouse

O usuário pode executar queries no Autonomous de duas maneiras, via o SQL Developer Web ou o OML Notebooks, será abordado as duas maneiras de execução de queries.

Os dois podem ser acessados conforme a imagem abaixo.



Inicialmente será acessado o SQL Developer Web, e aparecerá a tela de login.



O login pode ser feito com o usuário ADMIN ou qualquer outro usuário que tenha sido criado no Autonomous. Neste caso o usuário utilizado será o ADMIN e a senha é a que foi utilizada na criação do Autonomous.

Feito isso será acessado a tela abaixo, e nesta Worksheet será executado a seguinte query abaixo.

```
drop view DV_SH_VIEW;

create or replace view DV_SH_VIEW as select
P.PROD_NAME,
P.PROD_DESC,
P.PROD_CATEGORY,
P.PROD_SUBCATEGORY,
P.PROD_LIST_PRICE,
S.QUANTITY SOLD,
S.AMOUNT SOLD,
X.CUST_GENDER,
X.CUST_YEAR_OF_BIRTH,
X.CUST_MARITAL_STATUS,
X.CUST_INCOME_LEVEL,
X.CUST_CITY,
R.COUNTRY_NAME,
R.COUNTRY_SUBREGION,
R.COUNTRY_REGION,
T.TIME_ID,
T.DAY_NAME,
T CALENDAR_MONTH_NAME,
T CALENDAR_YEAR from
SH.PRODUCTS P,
SH.SALES S,
SH.CUSTOMERS X,
SH.COUNTRIES R,
SH.TIMES T where
S.PROD_ID=P.PROD_ID and
S.CUST_ID=X.CUST_ID and
S.TIME_ID=T.TIME_ID and
X.COUNTRY_ID=R.COUNTRY_ID;
```

```

1 drop view DV_SH_VIEW;
2 create or replace view DV_SH_VIEW as select
3   P.PROD_ID,
4   P.PROD_DESC,
5   P.PROD_CATEGORY,
6   P.PROD_SUBCATEGORY,
7   P.PROD_BRAND,
8   P.PROD_TYPE,
9   S.QUANTITY_SOLD,
10  S.AVERAGE_SOLD,
11  S.AVERAGE_PRICE,
12  X.CUST_YEAR_OF_BIRTH,
13  X.CUST_INCOME_LEVEL,
14  X.CUST_INCOME_AVG,
15  R.COUNTRY_NAME,
16  R.COUNTRY_CODE,
17  R.COUNTRY_REGION,
18  T.DAY_NAME,
19  T.MONTH_NAME,
20  T.CALENDAR_YEAR_Troe,
21  SM.PRONOUN_P,
22  SM.PRONOUN_S,
23  SM.CUSTOMERS_X,
24  SM.TYPES_T,
25  S.CUST_ID=S.X.CUST_ID and
26  S.CUST_ID=X.CUST_ID and
27  X.COUNTRY_ID=R.COUNTRY_ID;

```

Feito isso a visão dos dados que será utilizada com o Oracle Analytics já está pronta no Autonomous.

Outra alternativa para executar queries seria utilizando o OML Notebooks, para isso, primeiramente é necessário a criação de um usuário, pois o usuário administrador não é permitido que ele desenvolva no ambiente de Notebooks.

Para criar o usuário é necessário acessar o Manage Oracle ML User para criar um usuário de desenvolvimento, isto é equivalente a criar um novo Schema dentro do Banco Autonomous.

Acessando este link acima, o usuário acessará a console conforme a imagem abaixo, para criar um usuário novo clique no botão create.

User Name	Full Name	Role	Email	Created On	Status
ADMIN		System Administrator		5/6/19 9:18 AM	Open
ARMANDO		Developer	armando.neto@oracle.com	7/23/19 5:09 PM	Open

Clicando no botão create aparecerá as caixas de preenchimento conforme a imagem abaixo.

\* Username

First Name

Last Name

\* Email Address

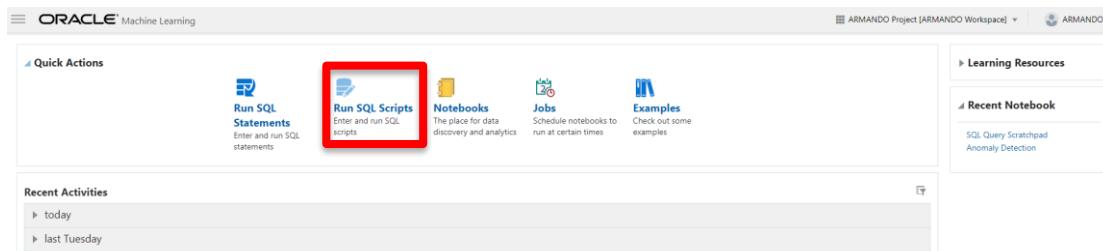
Generate password and email account details to user. User will be required to reset the password on first sign in.

\* Password

\* Confirm Password

Preencha com um nome de usuário que contenha apenas letras, um endereço de e-mail qualquer e de selecione a caixa para neste mesmo momento já inserir uma senha para o usuário, feito isso clique em create e o usuário já estará criado, é possível visualizar seu Schema no Autonomous e o mesmo já tem acesso de Desenvolvedor no OML.

Agora será feito o acesso ao OML, conforme já descrito neste guia, vá para a aba Development e clique em OML Notebooks e faça o acesso com o usuário e senha que acabou de ser criado, a console da imagem abaixo aparecerá.



Nesta console é possível processar SQLs e PL/SQL, além disso também é possível a criação de Notebooks que seriam sequencias de queries nas duas linguagens mencionadas acima para ser executadas no Autonomous.

Além disso também existe exemplos de alguns scripts de Machine Learning já prontos na parte de “Examples”.

Neste caso como será apenas executado uma query para criar uma visão, será acessado o “Run SQL Scripts” e a seguinte imagem aparecerá.



Antes de executar a query proposta, primeiro será configurado em qual conexão do Autonomous este Notebook será executado, para isso tem que acessar a imagem de engrenagem, conforme indicado na imagem acima.

Interpreter binding

Bind interpreter for this note. Click to Bind/Unbind interpreter. Drag and drop to reorder interpreters. The first interpreter on the list becomes default.

orcladw\_high %sql (default), %script

md %md (default)

orcladw\_low %sql (default), %script

orcladw\_medium %sql (default), %script

Save Cancel

Clicando na engrenagem, aparecerá a tela da imagem acima, neste caso que tem apenas um usuário neste banco será executado a query no orcladw\_high, para ter melhor performance e mais velocidade de execução da query, após selecionado, clique em Save.

Feito isso, o Notebook está conectado ao Autonomous via a conexão High e já é possível executar a query proposta no guia para a criação da visão.

```

drop view DV_SH_VIEW;
create or replace view DV_SH_VIEW as select
  P.PROD_NAME,
  P.PROD_ID,
  P.PROD_CATEGORY,
  P.PROD_BRAND,
  P.PROD_TYPE,
  P.PROD_LIST_PRICE,
  P.QUANTITY_SOLD,
  P.QUANTITY_AVAIL,
  X.CUST_GENDER,
  C.CUST_FIRSTNAME,
  C.CUST_LASTNAME,
  C.CUST_MARITAL_STATUS,
  C.CUST_INCOME_LEVEL,
  C.CUST_CITY,
  R.COUNTRY_NAME,
  C.COUNTRY_POSTCODE,
  C.COUNTRY_REGION,
  T.DAY_NAME,
  CALENDAR_MONTH_NAME,
  CALENDAR_YEAR,
  SH_PRODUCTS.P,
  SH_CUSTPERS.X,
  SH_COUNTRIES.R,
  T.MONTH_ID,
  S.PROD_ID=P.PROD_ID and
  S.CUST_ID=C.CUST_ID and
  T.MONTH_ID=T.MONTH_ID and
  X.COUNTRY_ID=R.COUNTRY_ID;
  
```

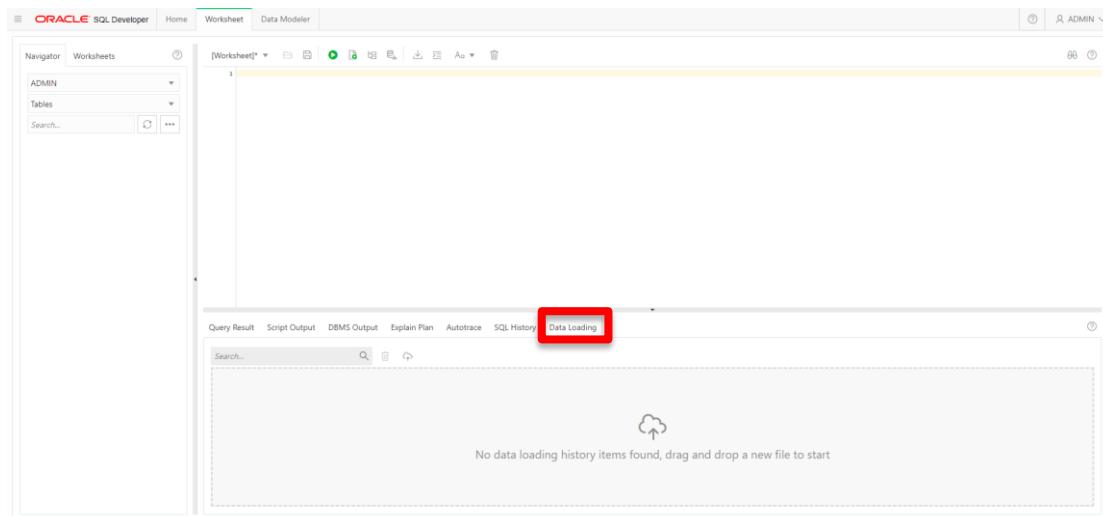
Após digitado a query no parágrafo, clique no play conforme indicado na imagem e o Autonomous processará a query e a visão estará criada no Schema do usuário que executou a query.

Assim finalizando as duas maneiras que é possível de executar queries no Autonomous via as ferramentas web.

## Carregando dados no Autonomous Data Warehouse

Será realizada uma carga de dados via SQL Developer Web para que seja trabalhado os Scripts de Machine Learning desejados.

Entre no SQL Developer Web como já foi mencionado neste material. Vá em “Data Loading” e arraste os arquivos desejados para lá.

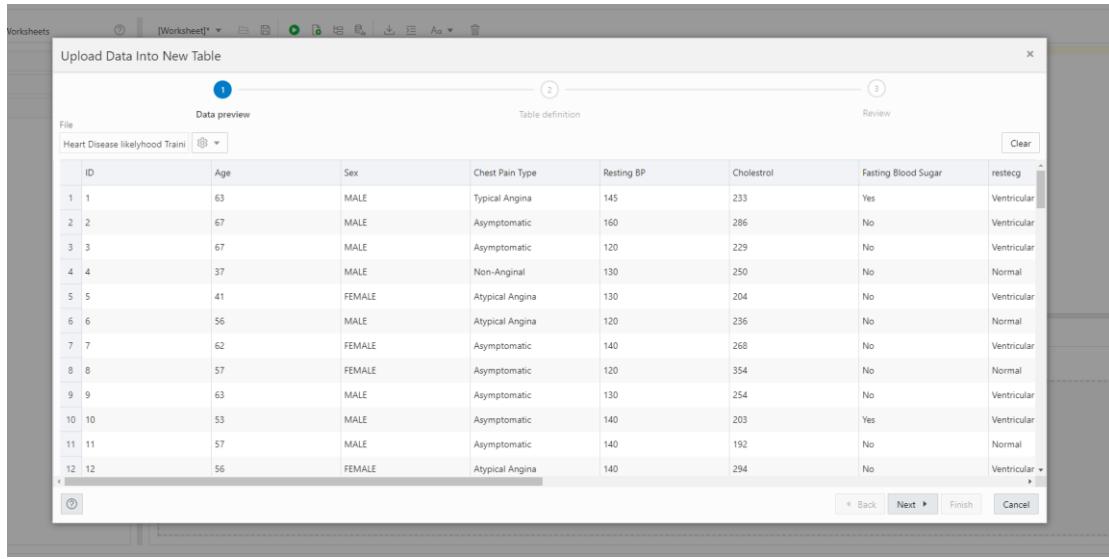


Neste material para realizar os treinamentos dos scripts de machine learning será utilizado um material de exemplo que é público e está na biblioteca do analytics. O vídeo de demonstração está neste link:

[https://youtu.be/lichF5pBt\\_U](https://youtu.be/lichF5pBt_U)

Neste caso será carregado duas bases uma chamada de “Heart Disease likelyhood Training” onde será treinado os scripts e outra base chamada “Heart Disease likelyhood Predict” onde será usado os scripts de aprendizado.

Após arrastar um arquivo conforme o primeiro passo uma outra janela se abrirá. Conforme a imagem abaixo.



A carga do dado é feita em 3 passos:

- Primeiro Passo: O SQL Developer mostra a estrutura da tabela que está sendo carregada
- Segundo Passo: É mostrado qual a definição de tabela que será criada pelo SQL Developer, o usuário pode alterar as definições
- Terceiro Passo: Revisão do que será criado e carregado no Autonomous, feito isso clique em “Finish”.

Repita os mesmos passos para qualquer outro arquivo.

# Lab 4.

## Explorando e criando um projeto no Oracle Analytics Cloud

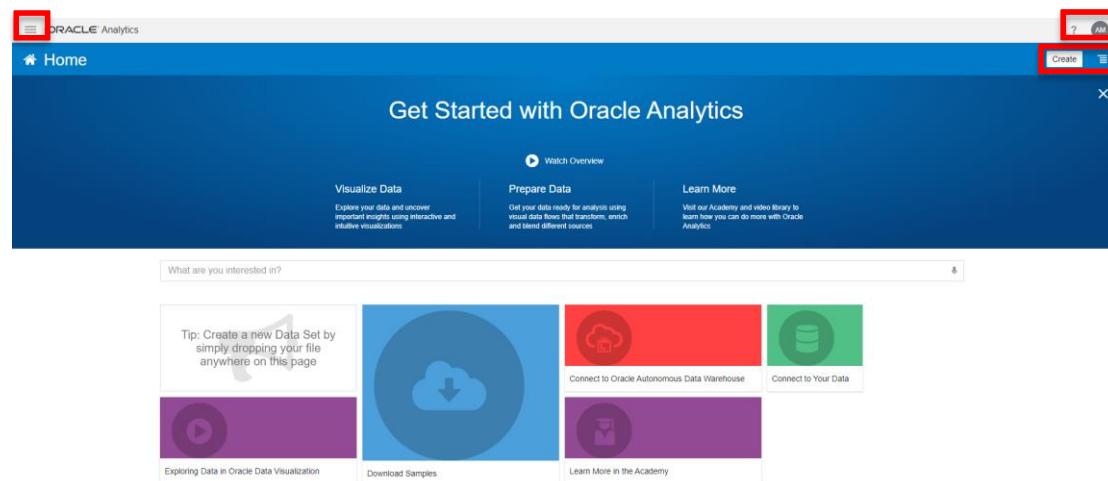


# Lab 4. Explorando e criando um projeto no Oracle Analytics Cloud

## Objetivos

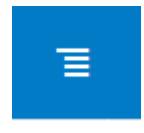
- Overview do Oracle Analytics Cloud
- Criar conexões com bancos de dados
- Carregar bases de dados diretamente na ferramenta
- Criar um projeto

## Overview do Oracle Analytics Cloud



Acima tem-se a tela inicial do Oracle Analytics Cloud. Os botões mais importantes são:

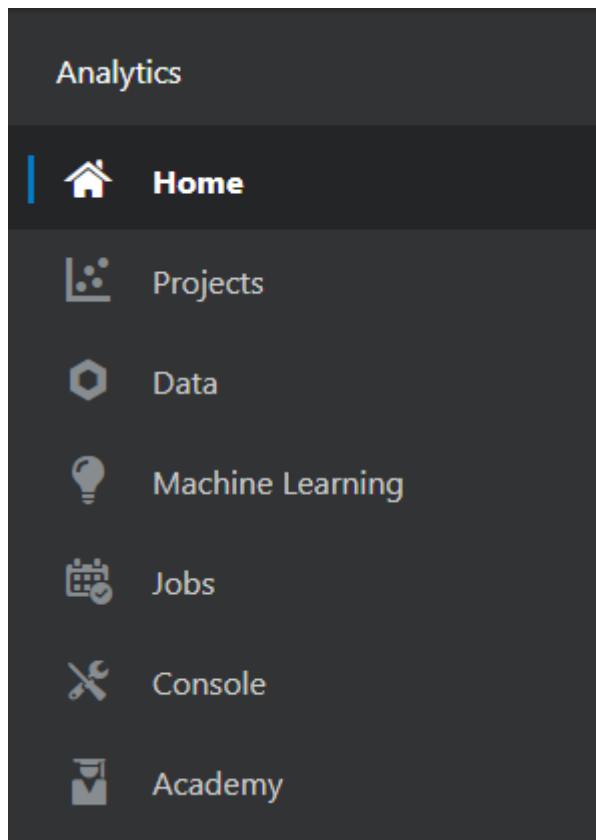
 Este botão é o menu principal. Clicando no mesmo, abrirá uma aba para navegar entre as áreas do OAC.

 Este botão é um submenu que permite a importação de projetos, visualizar o quanto está sendo consumido de "Storage" o seu usuário e onde pode-se customizar a "Home Page". Caso tenha a versão Enterprise do OAC, neste botão também é permitido acessar o Data Modeler e abrir o Classic Home (OBIEE)

 Por este botão é permitido que se crie projetos, novas bases de dados ou conexões, *Data Flows*, sequências de *Data Flows*, também criar replicações de dados ou de conexões.



Por estes dois botões o usuário consegue acessar a documentação da ferramenta, caso tenha alguma dúvida. Acessar seu perfil, trocar sua senha ou se desconectar da sua conta.



Quando o botão do menu principal é clicado uma aba aparecerá como na imagem acima. Por esta aba pode-se navegar por todas as áreas do Oracle Analytics Cloud. A parte de projetos é onde são acessados os projetos criados pelo usuário e os compartilhados com o mesmo.

Na aba “Data”, o usuário navega para a lista de todas as suas bases de dados, suas conexões, os *data flows* criados por ele, sequências e replicações de dados.

Na aba “Machine Learning”, o usuário encontra todos os modelos que ele treinou via “Data Flow”.

Na aba “Jobs” o administrador da ferramenta pode visualizar todos os *Data Flows* que os usuários da ferramenta estão executando e os que executaram, podendo ver também quais processaram com sucesso ou houve erro e quem estava executando.

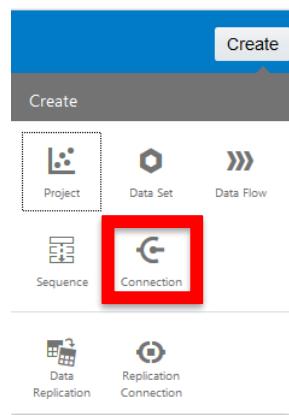
A aba “Console” é mais importante para o administrador da ferramenta do que para os usuários e desenvolvedores, já que é nesta aba que são feitas as

alterações na ferramenta, como por exemplo adicionar visualizações novas ou outros tipos de mapas.

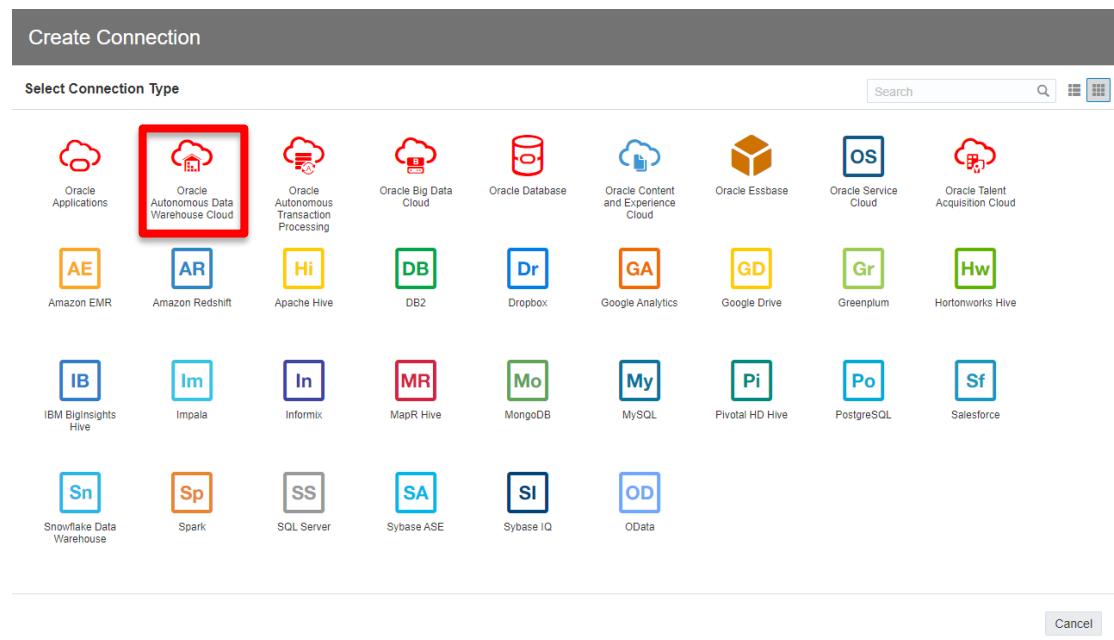
E por fim, na aba “Academy” o usuário tem acesso a documentação da ferramenta para tirar suas dúvidas.

## Criando uma conexão com o banco de dados

Partindo da tela inicial, clique em “Create” e depois em “Connection”.



Irá abrir uma janela para selecionar qual tipo de conexão. Neste caso, será criado uma conexão com um Oracle Autonomous Data Warehouse Cloud.





## Create Connection



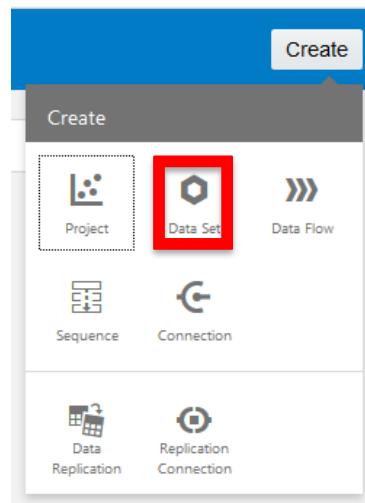
Oracle Autonomous Data Warehouse Cloud

* Connection Name	<input type="text" value="ORCLADW"/>
Description	<input type="text"/>
* Client Credentials	<input type="text" value="cwallet.sso"/> <input type="button" value="Select..."/>
* Username	<input type="text" value="admin"/>
* Password	<input type="password" value="*****"/>
* Service Name	<input type="text" value="orcladw_high"/> <input type="button" value="▼"/>

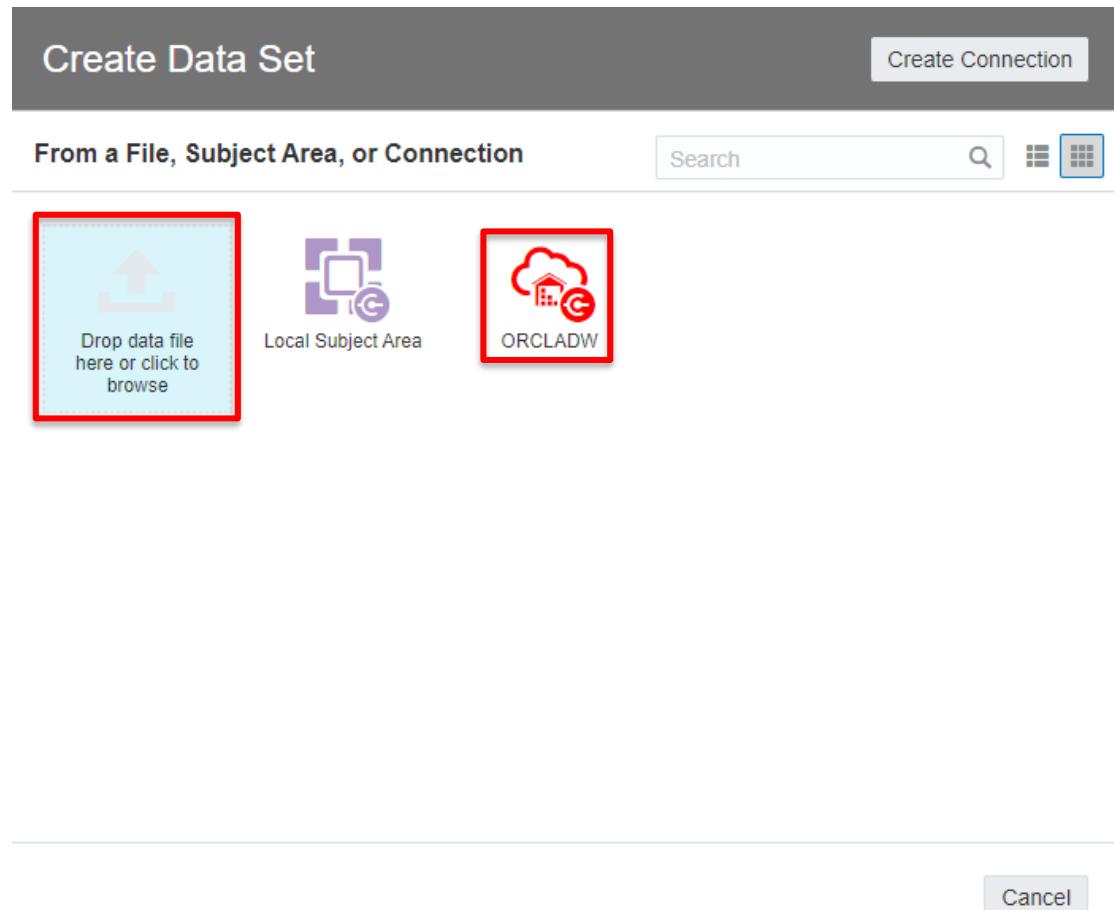
Após preenchidos todos os campos obrigatórios, clique em “Save” e, se tudo estiver correto, aparecerá uma janela verde confirmando que a conexão foi salva com sucesso. Para acessar os dados no banco, clique no menu principal, vá em “Data”, clique na aba “Connections” e clique na conexão criada para acessar os *Schemas* do banco. Procure a base de dados ou a visão desejada e salve a conexão.

## Criando uma base de dados e explorando o “Data Wrangling”

Para criar uma base de dados via console, clique em “Create” e depois clique em “Data Set”.



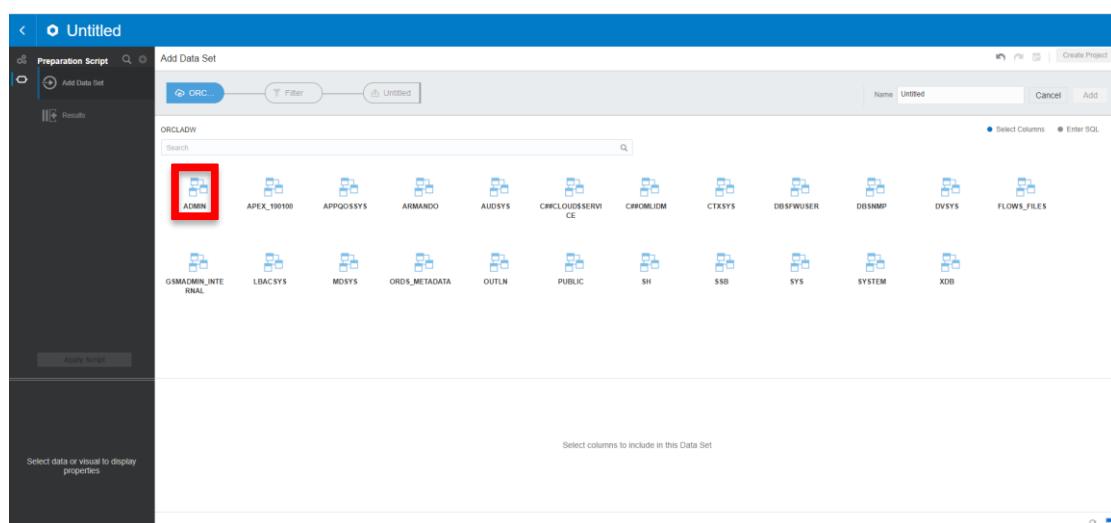
Feito isso, uma janela se abrirá. Clique na parte destacada da imagem abaixo para indicar o caminho da base desejada ou arraste a mesma para o quadrado para começar o upload de uma base de dados qualquer para a ferramenta. É recomendado que o arquivo esteja no formato xlsx (o tamanho máximo deste arquivo é 300 MB), mas também são aceitos xls, txt e csv. Caso a base de dados seja maior que 300 MB o dado deve ser carregado na ferramenta via DataSync ou carregado em um banco de dados e consumir os dados do mesmo.



Neste guia será criado uma base de dados que vem do Autonomous, quando é criado um Data Set de um banco de dados, basicamente o que o Analytics vai construir é o *SELECT* na base de dados ou na visão dentro do banco, nesse caso o Autonomous Dara Warehouse. Então será clicado na conexão destacada na imagem acima.

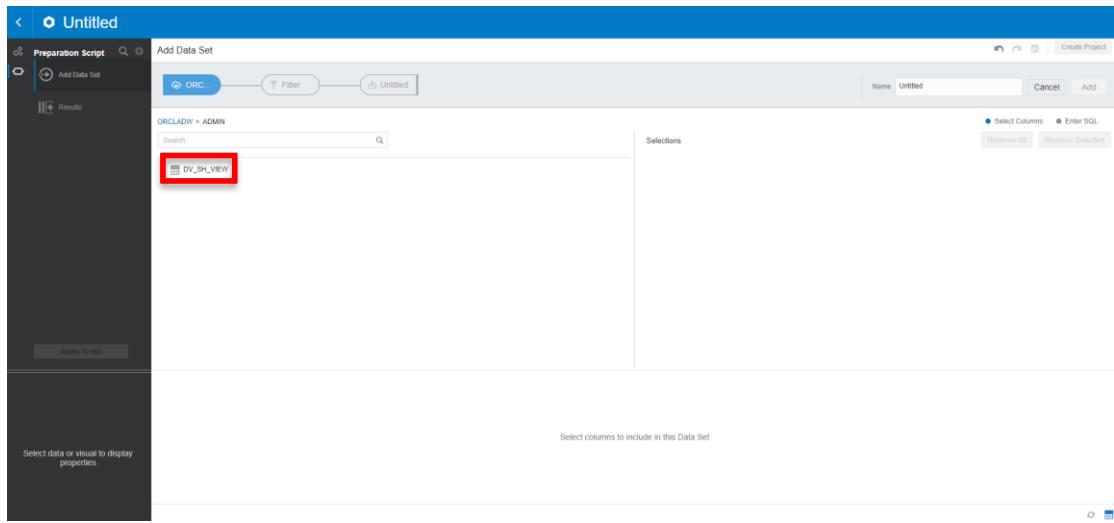
Após clicado na conexão, o Analytics mapeia todos os Schemas do banco que o usuário conectado pode visualizar, neste caso o usuário conectado foi o ADMIN, logo ele pode visualizar todos os Schemas do banco.

Anteriormente neste tutorial foi executado a query que será necessária para criar as visualizações, uma vez ela foi executada com o usuário ADMIN e outra vez foi executada com o usuário criado na ferramenta do Oracle Machine Learning, os dois podem ser usados como fonte, neste passo será utilizado o do usuário ADMIN.

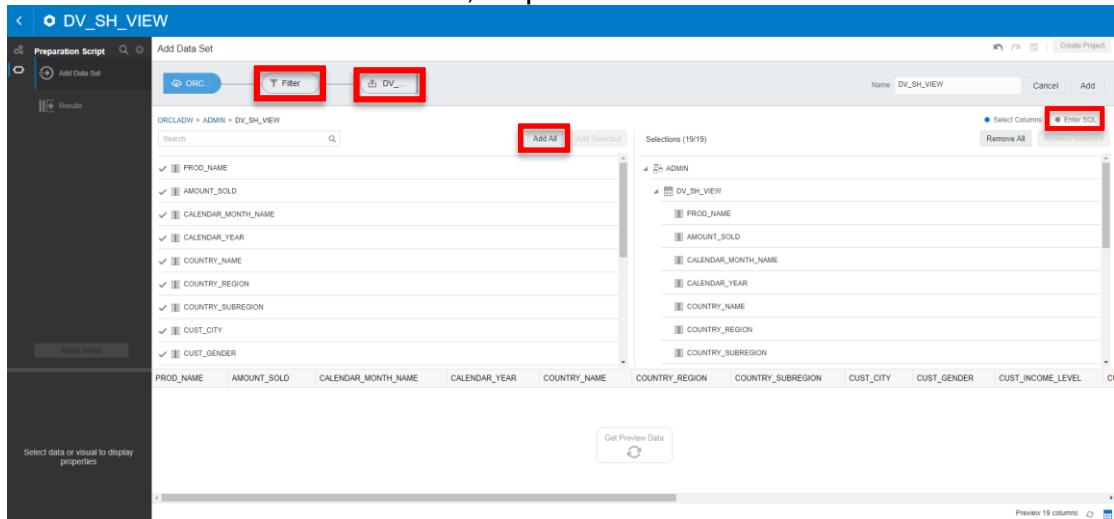


The screenshot shows the Oracle Analytics interface with the title 'Untitled'. On the left, there's a sidebar with 'Preparation Script' and 'Add Data Set' buttons. The main area is titled 'Add Data Set' with a 'CRC...' button, a 'Filter' button, and an 'Untitled' placeholder. Below this is a search bar with 'Search' and a magnifying glass icon. The main content area is titled 'ORCLADW' and contains a grid of schema names. The 'ADMIN' schema is highlighted with a red box. Other schemas listed include APEX\_190100, APPQOSYS, ARMANDO, AUDSYS, CIRCLOUDSERVICE, CHROMIUM, CTXSYS, DBAPRUSER, DBSNMP, DVTSYS, FLOWF\_FILES, GIMADMIN\_INTE\_RNAL, LBACSYS, MOSYS, ORDS\_METADATA, OUTLN, PUBLIC, SH, SSB, SYS, SYSTEM, and XDB. At the bottom, there's a note 'Select data or visual to display properties' and a 'Select columns to include in this Data Set' section.

Clicando no Schema do usuário ADMIN o Analytics mapeia todas as visões e bases de dados que tem no mesmo.



Neste caso tem apenas a visão que foi criada via a query que foi executada anteriormente neste laboratório, clique na visão destacada.



Feito isso o Analytics mapeia todas as colunas que estão na visão criada, clicando em “Add All”, ele seleciona todas as colunas, pode-se criar uma base com menos colunas selecionadas também, basicamente o que o Analytics está fazendo é criar o *SELECT* no Autonomous, que pode ser visto no botão “Enter SQL”.

Caso o usuário deseje filtrar a base de dados antes de finalizar a criação do *SELECT*, ele deve clicar no botão “Filter” destacado na imagem acima.

Para finalizar a criação da base, clique na parte destacada ao lado do botão “Filter”.

DV\_SH\_VIEW

Add Data Set

DV...

Name: DV\_SH\_VIEW

Description: External Data Set

Connection: ORCLADW

Data Access: Live

New Data Indicator: TIME\_ID

A PROD_NAME	A PROD_DESC	A PROD_CATEGORY	A PROD_SUBCATEGORY	PROD_LIST_PRICE	QUANTITY SOLD	AMOUNT SOLD	CUST
Home Theatre Package with DVD-Audio/Video Play	Home Theatre Package with DVD-Audio/Video Play	Electronics	Home Audio	599.99	1	629.99 M	
Home Theatre Package with DVD-Audio/Video Play	Home Theatre Package with DVD-Audio/Video Play	Electronics	Home Audio	599.99	1	629.99 F	
Home Theatre Package with DVD-Audio/Video Play	Home Theatre Package with DVD-Audio/Video Play	Electronics	Home Audio	599.99	1	561.47 M	
Home Theatre Package with DVD-Audio/Video Play	Home Theatre Package with DVD-Audio/Video Play	Electronics	Home Audio	599.99	1	628.89 M	
Mouse Pad	Mouse Pad	Software/Other	Accessories	9.99	1	9.99 M	
External 8X CD-ROM	External 8X CD-ROM	Peripherals and Accessories	CD-ROM	49.99	1	49.99 M	
Standard Mouse	Standard Mouse	Peripherals and Accessories	Accessories	22.99	1	22.99 M	
Model SM26273 Black Ink Cartridge	Model SM26273 Black Ink Cartridge	Peripherals and Accessories	Printer Supplies	27.99	1	27.99 M	
18" Flat Panel Graphics Monitor	18" Flat Panel Graphics Monitor	Peripherals and Accessories	Monitors	899.99	1	899.99 M	
SIMM- 16MB PCMCIAI card	SIMM- 16MB PCMCIAI card	Peripherals and Accessories	Memory	149.99	1	149.99 M	
External 8X CD-ROM	External 8X CD-ROM	Peripherals and Accessories	CD-ROM	49.99	1	49.99 M	
Mouse Pad	Mouse Pad	Software/Other	Accessories	9.99	1	9.99 M	

Nesta parte, o usuário escolhe como será o acesso aos dados, que pode ser feito de duas maneiras:

- Live: O último dado sempre é retornado da fonte viva. Alguns passos de preparação de dados e enriquecimento não estão disponíveis neste modo. Recomendado para grandes bases de dados
- Automatic Caching: O dado da fonte viva será colocado em cache no Analytics para uma melhor performance. Recomendado para pequenas bases de dados.

O usuário coloca qual a coluna que indica que novos dados foram inseridos na tabela e na parte destacada pode ser alterado o nome do Data Set criado. E por fim para finalizar a criação, clique em “Add”.

Após a base adicionada na ferramenta, um *script* de *Machine Learning* será executado sobre a base para identificar recomendações e algumas opções para enriquecimento da base.

DV\_SH\_VIEW2

Add DV\_SH\_VIEW2

Results

A PROD\_NAME A PROD\_DESC A PROD\_CATEGORY A PROD\_SUBCATEGORY PROD\_LIST\_PRICE QUANTITY SOLD AMOUNT SOLD CUST

17" LCD built-in-HDTV Tuner	17" LCD built-in-HDTV Tuner	Peripherals and Accessories	Monitors	999.99	1	1,108.99 M	
SMP Telephoto Digital Camera	SMP Telephoto Digital Camera	Photo	Cameras	899.99	1	1,193.88 M	
SMP Telephoto Digital Camera	SMP Telephoto Digital Camera	Photo	Cameras	899.99	1	1,265.99 F	
SMP Telephoto Digital Camera	SMP Telephoto Digital Camera	Photo	Cameras	899.99	1	1,222.87 M	
SMP Telephoto Digital Camera	SMP Telephoto Digital Camera	Photo	Cameras	899.99	1	1,205.99 M	
SMP Telephoto Digital Camera	SMP Telephoto Digital Camera	Photo	Cameras	899.99	1	1,250.25 M	
SMP Telephoto Digital Camera	SMP Telephoto Digital Camera	Photo	Cameras	899.99	1	1,042.55 M	
SMP Telephoto Digital Camera	SMP Telephoto Digital Camera	Photo	Cameras	899.99	1	1,036.64 M	
SMP Telephoto Digital Camera	SMP Telephoto Digital Camera	Photo	Cameras	899.99	1	1,055.31 F	
SMP Telephoto Digital Camera	SMP Telephoto Digital Camera	Photo	Cameras	899.99	1	1,210.21 M	
SMP Telephoto Digital Camera	SMP Telephoto Digital Camera	Photo	Cameras	899.99	1	1,045.04 M	
SMP Telephoto Digital Camera	SMP Telephoto Digital Camera	Photo	Cameras	899.99	1	1,205.99 M	
17" LCD built-in-HDTV Tuner	17" LCD built-in-HDTV Tuner	Peripherals and Accessories	Monitors	999.99	1	1,267.35 F	
SMP Telephoto Digital Camera	SMP Telephoto Digital Camera	Photo	Cameras	899.99	1	994.06 M	
SMP Telephoto Digital Camera	SMP Telephoto Digital Camera	Photo	Cameras	899.99	1	1,210.21 M	
SMP Telephoto Digital Camera	SMP Telephoto Digital Camera	Photo	Cameras	899.99	1	1,036.64 F	
SMP Telephoto Digital Camera	SMP Telephoto Digital Camera	Photo	Cameras	899.99	1	1,014.06 F	
SMP Telephoto Digital Camera	SMP Telephoto Digital Camera	Photo	Cameras	899.99	1	1,075.12 M	
SMP Telephoto Digital Camera	SMP Telephoto Digital Camera	Photo	Cameras	899.99	1	970.56 M	
SMP Telephoto Digital Camera	SMP Telephoto Digital Camera	Photo	Cameras	899.99	1	1,001.70 M	
SMP Telephoto Digital Camera	SMP Telephoto Digital Camera	Photo	Cameras	899.99	1	1,169.03 F	
SMP Telephoto Digital Camera	SMP Telephoto Digital Camera	Photo	Cameras	899.99	1	867.93 F	

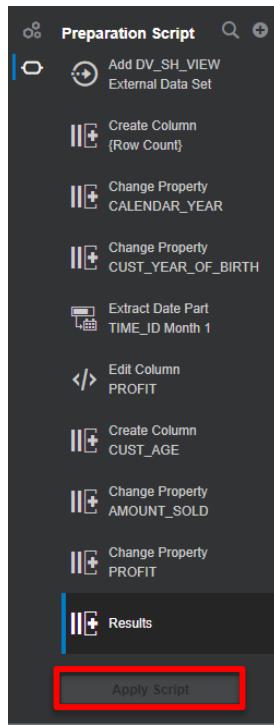
Recommendations (60)

Nesta mesma tela da imagem acima é uma tela de preparação de dados, é possível que o usuário faça inúmeras ações de “Data Wrangling”. Clicando no símbolo de “+” destacado na imagem é possível que seja criada uma nova coluna que pode ser personalizada conforme a necessidade do usuário, basta criar um cálculo utilizando os operadores disponíveis na ferramenta.

Cada coluna existente também tem um menu em seu cabeçalho, onde é possível fazer inúmeras ações. Na imagem abaixo temos um exemplo do que é possível fazer em uma coluna com os dados em formato de data.

Na coluna é extraído o dado “Month”, que cria uma coluna nova com os dados de mês e ano reduzindo a granularidade da coluna para algumas análises.

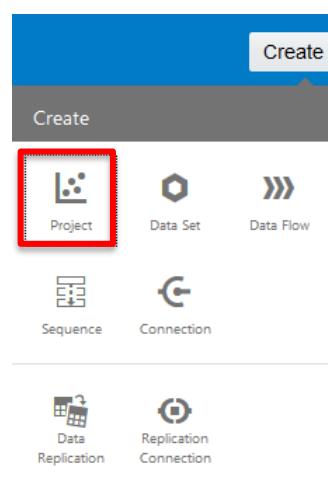
The screenshot shows the Power BI Data Editor interface. A context menu is open over the 'Order Date' column header. The menu includes options like 'Extract Day of Month', 'Rename...', 'Duplicate', 'Convert to Text', 'Concatenate...', 'Extract', 'Create...', 'Edit...', 'Hide', and 'Delete'. A sub-menu under 'Extract' is visible, showing 'Extract Year from' and other options. The main table view shows various date columns such as 'Ship Date', 'Order Date', and numerous timegranularity columns derived from the original date.



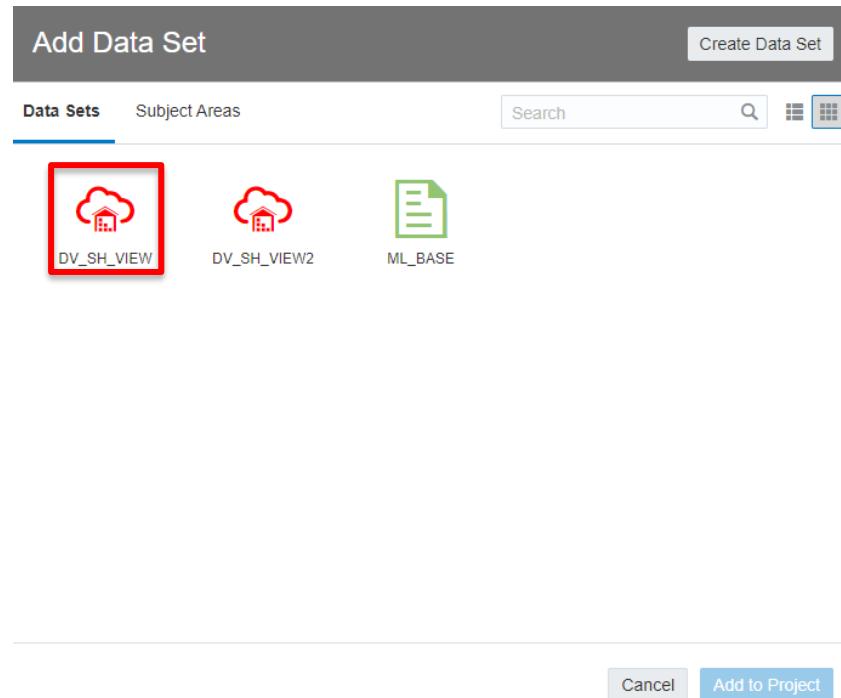
Depois de feitos todos os passos de preparação dos dados, clique em “Apply Script” para salvar todas as alterações feitas na base, como mostrado na imagem. Nesta parte da tela também é possível checar qual o passo a passo feito na base de dados. Os passos executados na base foram, mudar as propriedades das colunas CALENDAR\_YEAR e CUST\_YEAR\_OF\_BIRTH, indicando para o Analytics interpretar elas como Atributos e não como colunas de métrica, extrair o mês de TIME\_ID, criar a coluna PROFIT (Formula: QUANTITY\_SOLD\* (AMOUNT\_SOLD - PROD\_LIST\_PRICE)), e por fim criar a coluna CUST\_AGE (Formula: CALENDAR\_YEAR – CUST\_YEAR\_OF\_BIRTH).

## Criando um projeto no Data Visualization

Clique em “Create” e depois em “Project”.



Uma janela se abrirá para o usuário selecionar a base de dados que ele deseja explorar ou criar visualizações. Neste caso serão utilizadas as bases de dados que estão sendo selecionadas do Autonomous.



Após selecionada a base de dados desejada, na imagem abaixo está a tela de criação de projetos. Será feito um *overview* sobre as três telas: “Visualize”, “Prepare” e “Narrate”. Primeiro será explicado a parte do “Visualize”.

A screenshot of the "Visualize" tab in a project creation interface. On the left is a sidebar with "Data Elements" and a search/filter section. The main area is a canvas labeled "Canvas 1" with a "Drop Visualizations or Data Here" placeholder. At the bottom is a navigation bar with tabs: "Prepare", "Visualize" (highlighted with a red box), and "Narrate".

Nestas três abas é por onde o usuário pode navegar por 3 ferramentas na hora de criar um projeto.

O “Prepare” é onde o usuário consegue fazer alterações nas bases de dados, fazer *joins* entre múltiplas bases de dados e fazer o “Data Wrangling”.

No “Visualize” é onde o usuário cria suas visualizações e seus canvas, além de criar métricas calculadas, aplicar algumas práticas de *Analytics* diretamente nas visualizações e além disso usar a ferramenta do Oracle Analytics que é chamada de “*Explain*”, para acioná-la, o usuário faz clique direito em qualquer coluna de sua base de dados e aparecerá uma opção de *Explain*, quando acionada o Analytics processa um algoritmo de Machine Learning na base de dados e relaciona a coluna selecionada, linha a linha, com todas as outras colunas da base de dados.

No “Narrate” é onde o usuário narra a sua história, podendo escolher quais *canvas* ele utilizará para criar seu *storytelling*, podendo também colocar anotações em cada página para destacar algo e colocar em um formato de apresentação.



Nesta parte temos o símbolo de menu na esquerda que seria o menu dos filtros que vão ser colocados no canvas todo. Nele é possível escolher se os filtros vão ser aplicados automaticamente ou não e também é utilizado para criar um filtro de expressão.

As setas são o ir e voltar: caso tenha feito algo não desejado, clique em voltar, e caso tenha voltado sem vontade, pode-se clicar na seta do ir.

O outro botão é o símbolo de *Share*, com ele pode-se exportar uma visualização, o canvas todo ou o projeto todo em vários formatos. Também é possível a extração de dados via arquivo .csv.

O último botão da direita, que tem o símbolo de menu, é permitido ao usuário alterar as propriedades do projeto, criar ações de dados, definir se deseja ou não que as visualizações se sincronizem, atualizar os dados do projeto, voltar às cores para o padrão da ferramenta, voltar os símbolos para o padrão da ferramenta, criar um projeto novo, reverter alguma alteração e acessar o console de desenvolvedor da ferramenta.



Nestes botões é possível selecionar entre os elementos de dados que o usuário quer utilizar (engrenagens), as visualizações que ele quer utilizar para criar seu *canvas* (barras) e algumas funções de *Analytics* que podem ser colocadas nas visualizações (barras com a lupa).



Na lupa o usuário pode pesquisar o nome da coluna na base de dados que ele deseja utilizar no projeto e no símbolo de adição pode ser adicionado ou um novo *data set* ao projeto ou uma métrica calculada, além de também criar cenários, opção utilizada quando se deseja testar um modelo de *Machine Learning*.



Nesta barra da tela são mostrados todos os canvas que estão contidos no projeto. O símbolo de adição é utilizado para adicionar novos canvas em branco ao projeto, além disso clicando na seta o usuário pode renomear o canvas, mudar as propriedades do mesmo, duplicar, limpar, e deletar o canvas.

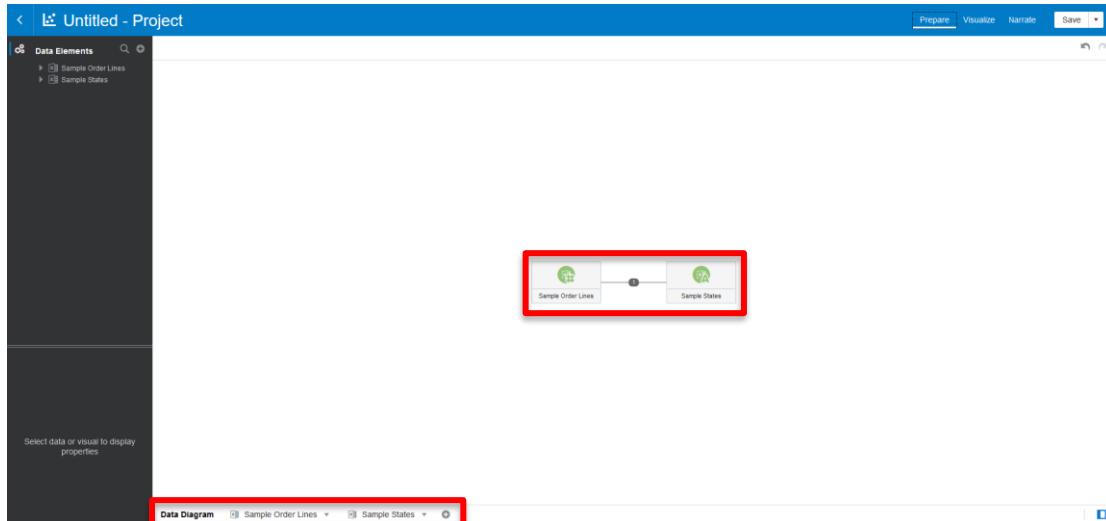
A screenshot of a software interface titled 'Data'. On the left, there's a tree view with nodes like 'DV\_SH\_VIEW', 'My Calculations', and 'Value Labels'. In the center, a column from a table is selected, highlighted with a red border. The properties for this column, 'PROD\_NAME', are displayed in a table:

Name	PROD_NAME
Treat As	Attribute
Data Type	Text
Aggregation	None

Na imagem acima está destacada uma parte importante da ferramenta. Sempre neste local aparecerá as propriedades das visualizações ou dos dados que estão selecionados pelo usuário, então sempre quando o usuário precisar ver ou alterar alguma propriedade da visualização ou de um dado será neste local que terá as informações sobre eles.

Na imagem abaixo tem-se a visão da aba “Prepare”. Nela pode-se checar os *joins* que a ferramenta pode ter feito automaticamente ou o usuário pode desfazer e criar novos *joins*, basta clicar na linha que está interligando as bases de dados no diagrama de dados abaixo na parte destacada.

Também na imagem abaixo, na barra que está destacada na parte inferior, é onde pode-se adicionar novas bases de dados ao projeto clicando no sinal de adição, ou checar os dados das bases já presentes e fazer o “Data Wrangling” na camada do projeto.

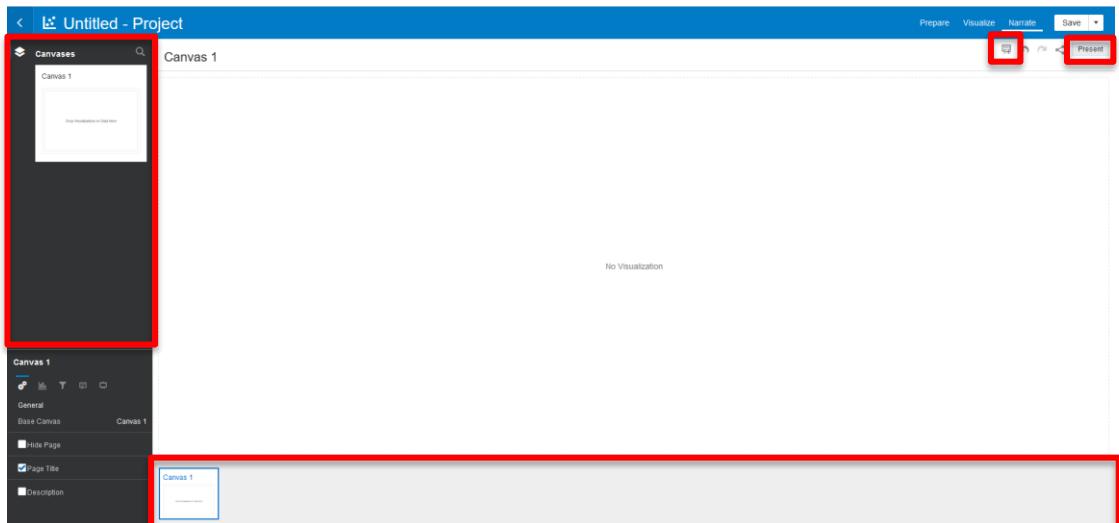


Na imagem abaixo, por último tem-se a visão da aba “Narrate”, onde é possível que o usuário crie sua narrativa, sua história, evitando que seja necessária uma segunda ferramenta para apresentar as análises criadas no Analytics.

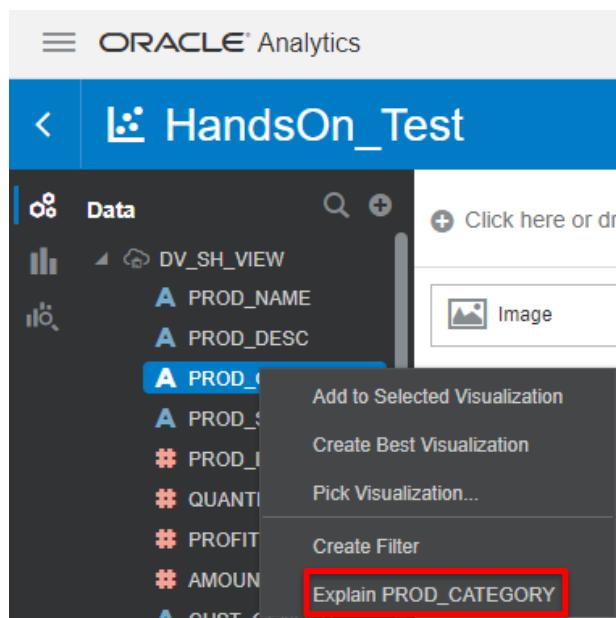
Para a construção da história o usuário arrasta os canvas desejados que foram criados no “Visualize”. Eles estão disponíveis na esquerda da tela em destaque e deve ser arrastado para a parte inferior.

Para a adição de anotações na narrativa, deve-se clicar no botão com o símbolo de anotação em destaque. Com isso, aparecerá uma caixa de texto flutuante onde se pode colocar a anotação em qualquer local que o usuário desejar.

Por fim, para que fique em formato apresentação, basta que o usuário clique no botão “Present” em destaque.

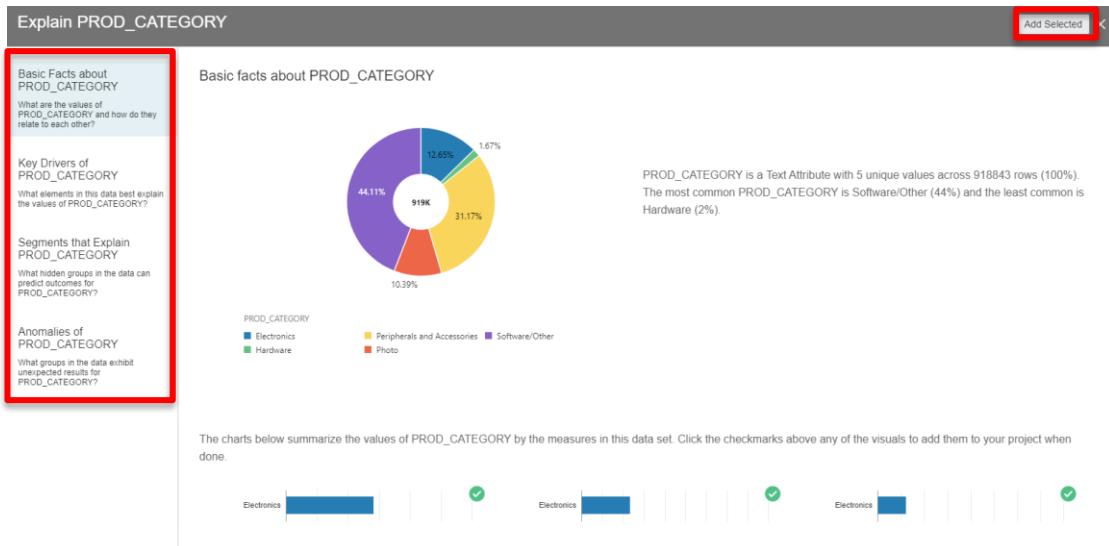


Após o *overview* de todas as abas da criação de um projeto, para um *discovery* da base de dados, criação rápida de dashboards e *insights*, será executado a *feature Explain* na coluna PROD\_CATEGORY.



Feito isso, a ferramenta abrirá uma janela que relacionará o atributo com todas as outras métricas e atributos da base de dados, trazendo *insights* para o usuário e também algumas possíveis anomalias, como pode ser observado na imagem abaixo.

Além disso, caso o usuário goste dos gráficos gerados pelo algoritmo, ele pode selecionar os gráficos e clicar no botão “Add Selected”.



Para a construção do projeto, inicialmente será desenhado um canvas que mostra um *overview* da base de dados, como resultado a imagem abaixo.



Como pode ser observado na imagem acima, na caixa de “Values” de cada visualização será colocada a métrica AMOUNT\_SOLD e PROFIT, respectivamente. Nas propriedades da visualização, é possível alterar como o número pode aparecer, por exemplo, como nesses dois casos, o assunto é monetário. O usuário pode configurar para o número aparecer como moeda.

Para enriquecer o *overview* dos dados no projeto, foram criadas algumas visualizações em gráficos de pizza, formato de tabela dinâmica, gráficos de linha, gráfico de Treemap e uma visualização de imagem.

Para a inserção da imagem no *canvas*, o usuário pode tanto clicar no “Select Image” quanto ir nas propriedades da visualização e indicar o caminho da imagem.

Nas visualizações de “*Big Numbers*”, que são quatro, foram colocados as métricas QUANTITY\_SOLD, AMOUNT\_SOLD, PROFIT e uma métrica calculada que foi chamada de # Average Ticket.

Os gráficos de pizza foram construídos com a coluna de dados PROD\_CATEGORY junto com a coluna de medida AMOUNT\_SOLD e PROFIT, a coluna de dados está no campo da visualização chamado “Color”, por isso cada fatia da pizza tem uma cor diferente da outra.

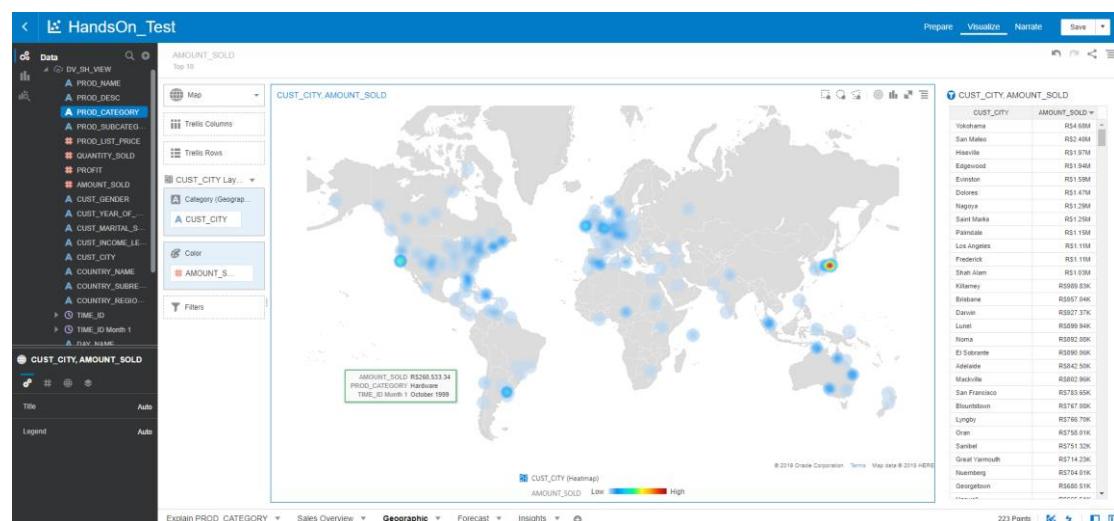
A visualização de *pivot*, nas linhas foi colocado o atributo PROD\_CATEGORY, e na caixa “Values” foi colocado a métrica QUANTITY\_SOLD, AMOUNT\_SOLD, # Average Ticket que; e uma métrica calculada de AMOUNT\_SOLD dividido por QUANTITY\_SOLD, PROFIT como valor e PROFIT mostrando como porcentagem do todo e por fim foi criado uma métrica chamada de % Profit Ratio é uma métrica que é a divisão de PROFIT por AMOUNT\_SOLD.

A visualização de linha foi construída com o atributo PROD\_CATEGORY nas caixas “Trellis Columns” e “Color”, com AMOUNT\_SOLD na caixa “Values” e TIME\_ID Month na caixa “Category”.

E, por fim, a visualização de *Treemap*, foi construída com AMOUNT\_SOLD na caixa “Values”, CUST\_INCOME\_LEVEL em “Category” e PROD\_CATEGORY em “Color”.

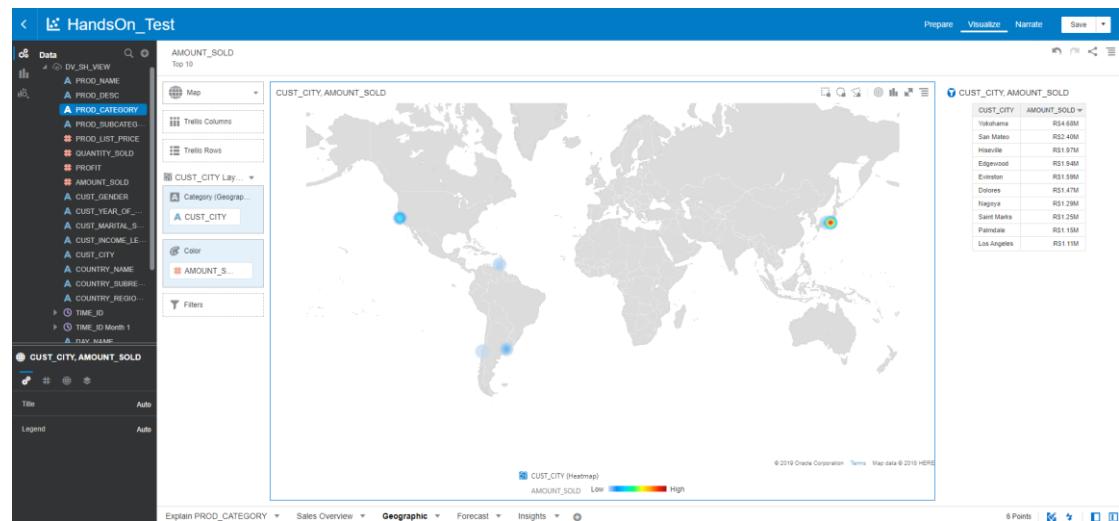
Este é um resumo de como foram construídas as visualizações neste *canvas* de *overview*.

No próximo *canvas* será construído um Mapa de Calor e mais uma visualização de tabela, para checar nominalmente em qual cidade a empresa da amostra mais vendeu. Segue abaixo o canvas com os dados populados:



Neste *canvas* para a visualização de mapa de calor, a caixa de “Category” é populado com o atributo CUST\_CITY; para a opção de mapa de calor aparecer nas propriedades da visualização, a granularidade do dado deve estar em nível de cidade; em “Values” foi colocada a métrica AMOUNT\_SOLD e nas propriedades da visualização foi indicado para o mapa aparecer como Mapa de Calor (*Heatmap*). Ao lado, foi feita uma visualização de tabela para checar os dados do mapa: ela foi populada com CUST\_CITY e AMOUNT\_SOLD.

Observa-se que na barra superior foi criado um filtro com a métrica AMOUNT\_SOLD. Todo o filtro também tem um menu de propriedades: neste caso, o filtro foi alterado do modelo de *Range* para o modelo de *Top Bottom*. Segue a imagem abaixo com o filtro ativo das dez cidades que a empresa exemplo mais vendeu:



No próximo *canvas*, será explorado o Forecast, que pode ser feito na própria camada de visualização. Segue abaixo o canvas com os dados populados e o Forecast na camada de visualização:



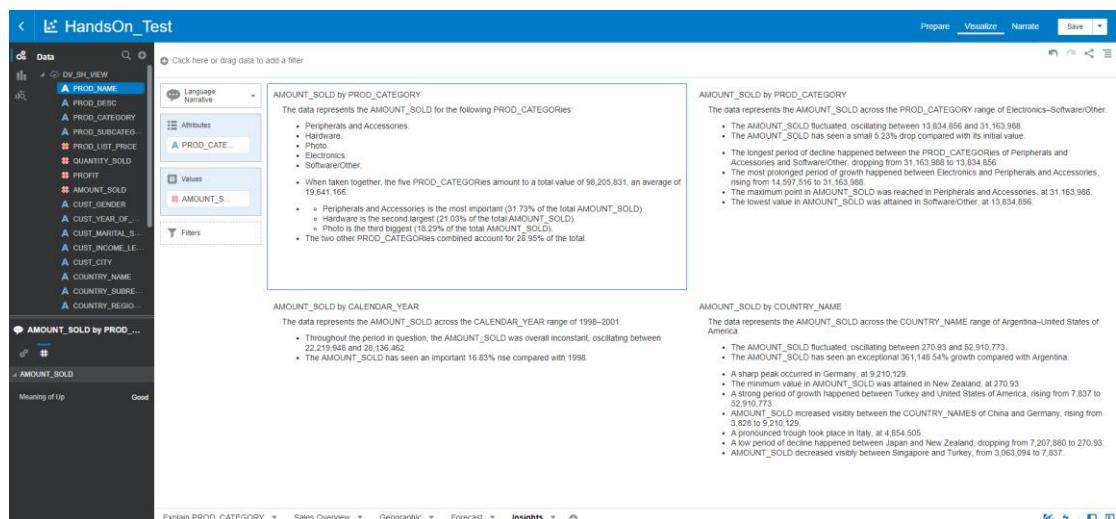
Neste canvas, para a construção do gráfico de combo, foi utilizado duas métricas de valor, uma métrica é AMOUNT\_SOLD e a outra métrica é QUANTITY\_SOLD, a métrica AMOUNT\_SOLD foi colocada no formato de barra e QUANTITY\_SOLD foi colocado como linha no gráfico, para o usuário fazer esse tipo de alteração de como a métrica será impressa no gráfico, ele deve apertar o botão direito do mouse na métrica dentro da caixa “Values” e escolher a visualização desejada para a mesma, ou até se deseja um segundo eixo Y.

Feito isso, foi selecionado a visualização para que apareça as propriedades da imagem no canto inferior esquerdo, vá na parte em destaque na imagem acima, o símbolo é exatamente igual ao símbolo de onde estão as operações de Analytics, clique no símbolo de adição e adicione a operação de Forecast.

Na imagem acima o Forecast realizado foi de três períodos, baseado na granularidade de data do gráfico, o modelo escolhido foi o padrão “Seasonal ARIMA”, mas além desse é possível escolher ARIMA e ETS também, o intervalo de confiança foi designado como 95%.

No próximo e ultimo canvas, será explorado a visualização chamada de “Language Narrative”, é uma visualização que cria insights em forma de texto baseado no seus dados, a visualização consegue relacionar, uma coluna de métrica com uma coluna de atributo, duas de métrica e uma de atributo ou duas de atributo e uma de métrica.

Os idiomas suportados pela visualização hoje são Inglês e Francês, o usuário pode controlar qual o nível de detalhe que a visualização vai gerar o texto, pode escolher se esse texto vai conter uma análise de detalhamento ou de tendência e também se o teor do conteúdo vai ser um teor positivo ou negativo. Segue a imagem do canvas abaixo.



Foram criadas quatro visualizações de “Language Narrative”, todas com o conteúdo positivo, duas estão relacionando a métrica AMOUNT\_SOLD com PROD\_CATEGORY, a diferença entre elas é que em uma a análise é de detalhamento e a outra a análise é de tendências.

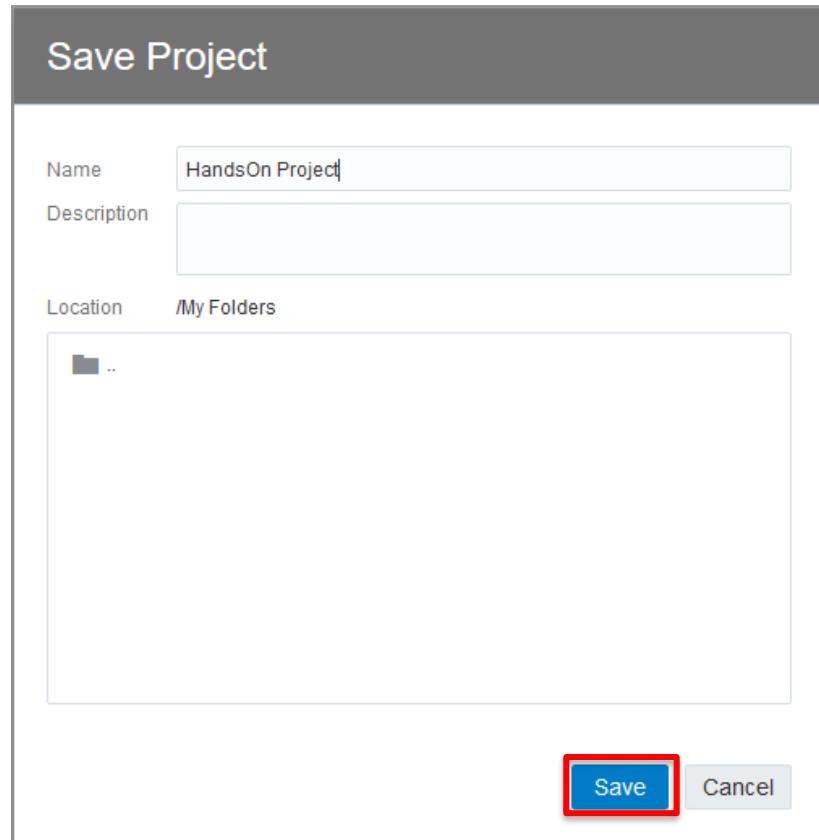
A terceira visualização foi relacionado AMOUNT\_SOLD com CALENDAR\_YEAR em uma análise de tendências.

A quarta e última visualização foi relacionado AMOUNT\_SOLD com COUNTRY\_NAME em uma análise de tendências.

Após terminada a construção do projeto, clique no botão Save para salvar o projeto. Quando for salvar o projeto pela primeira vez, a janela abaixo se abrirá.

O projeto pode ser salvo na pasta pessoal do usuário ou em uma pasta compartilhada para que o projeto seja compartilhado.

O nome pode ser qualquer nome que o usuário desejar e a descrição não é obrigatória.



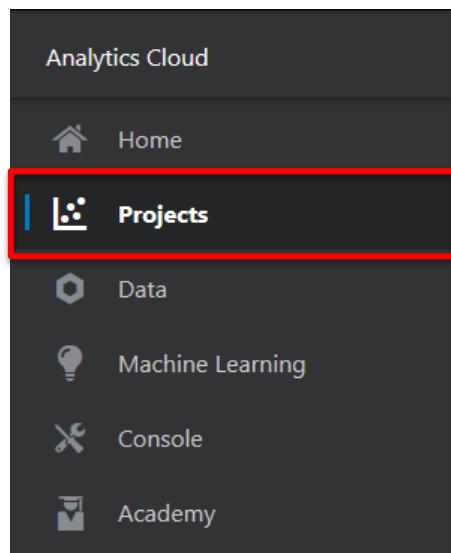
Clique em “Save” e o projeto, neste caso, estará salvo apenas na pasta do usuário.

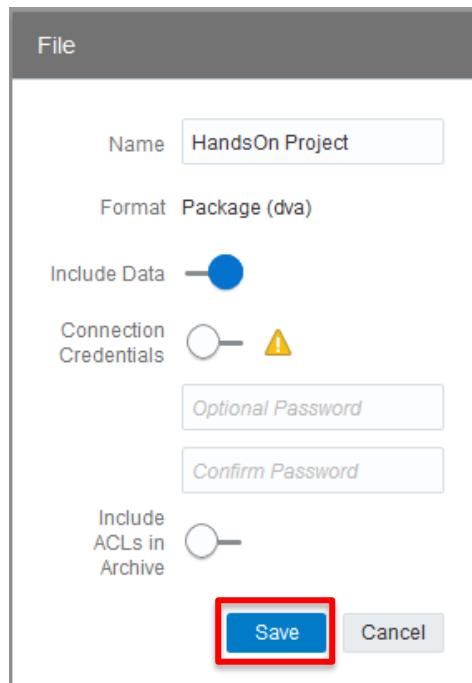
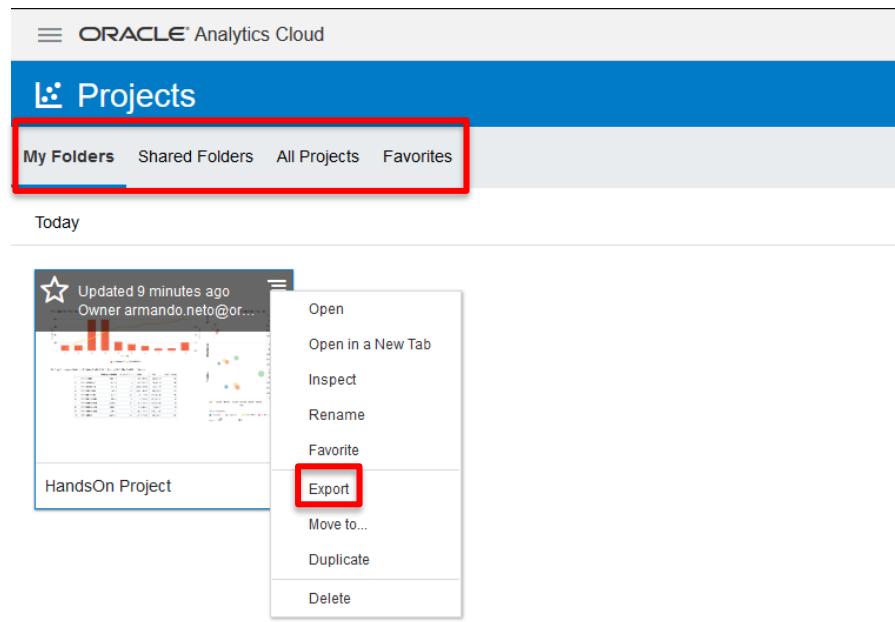
O usuário também pode salvar este projeto no local intitulado como “Shared Folders”, neste local o usuário não pode salvar diretamente na pasta raiz, ele deve criar pastas para já começar a criar a segmentação de quem pode ver o que, todos os recursos do Analytics têm as opções de permissionamento, desde a conexão com a base de dados, as pastas e projetos.

Para fazer a governança a nível dos dados, o usuário pode criar múltiplas bases de dados e visualizações ou com a versão Enterprise do Oracle Analytics Cloud, ele pode criar “Subject Areas” e fazer toda essa governança via uma ferramenta chamada de Data Modeler que será detalhado mais a frente neste guia.

## Exportando os dados

Para a exportação do projeto, o usuário pode abrir o menu, ir em “Projects” e ir na aba que está o projeto desejado, como as imagens abaixo.





Com este passo a passo, o usuário irá exportar o projeto da nuvem no formato .dva, que é o formato do Oracle Analytics Cloud. Com isso, ele poderá compartilhar esse projeto de uma maneira *On-Premises*.

É possível também exportar o projeto em outras extensões, como por exemplo em:

- Powerpoint (.ppt), Todos os Canvas, Canvas selecionado, Visualização selecionada.
- Acrobat (.pdf), Todos os Canvas, Canvas selecionado, Visualização selecionada.
- Imagem (.png), Canvas selecionado, Visualização selecionada.
- Dados (.csv), Visualização selecionada.
- Pacote (.dva), o Projeto todo.

Segue um tutorial abaixo.

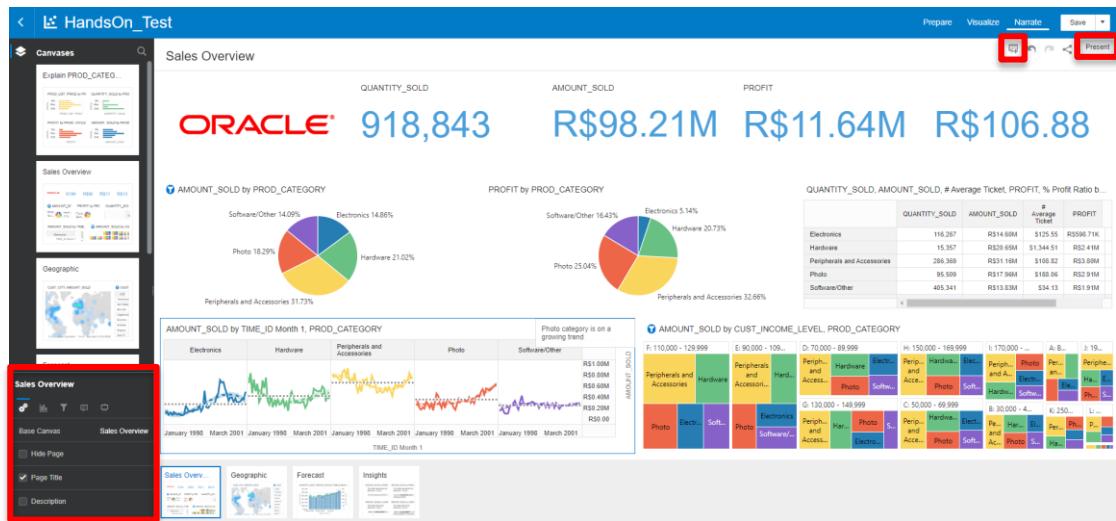
The screenshot shows the Oracle Analytics Cloud interface. At the top, there's a navigation bar with tabs: Prepare, Visualize (which is selected), Narrate, and Save. Below the navigation bar is a toolbar with icons for search, refresh, and other functions. The main area displays a dashboard titled "Profit" with a large blue text "1M \$1.30M". Below the dashboard is a table titled "# Order ID, Sales, Profit, % Profit, % Profit Ratio by Product Category" with the following data:

	# Order ID	Sales	Profit	% Profit	% Profit Ratio
Furniture	1,726	\$2.89M	\$156.63K	12.05%	5.41%
Office Supplies	3,949	\$2.13M	\$413.41K	31.80%	19.39%
Technology	2,072	\$3.47M	\$729.96K	56.15%	21.01%

At the bottom, a modal dialog box is open with the title "File". It contains fields for "Name" (set to "HandsOn Project"), "Format" (set to "Powerpoint (.pptx)"), "Include" (set to "All Canvas"), "Size" (set to "US Letter (8.5" x 11")"), and "Orientation" (set to "Landscape"). The "Save" button is highlighted with a red box.

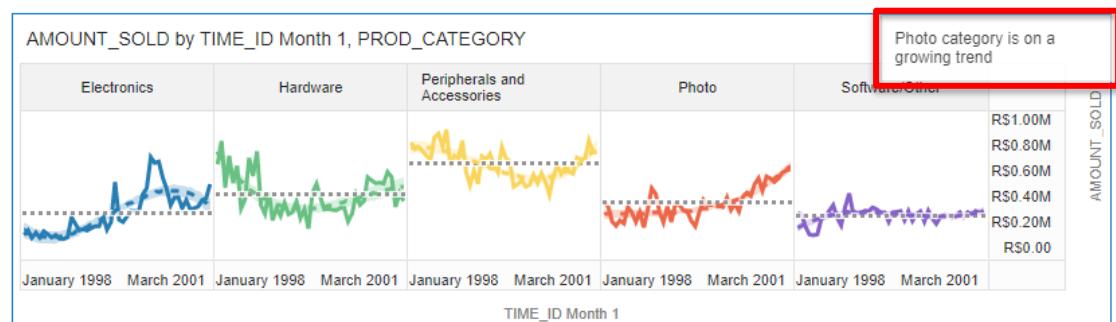
Neste caso acima, quando clicar em “Save”, começará um download de um Powerpoint com todos os *canvas* do projeto.

Na imagem abaixo, tem-se a tela do “Narrate”, com todos os canvas adicionados para criarmos a história desejada.

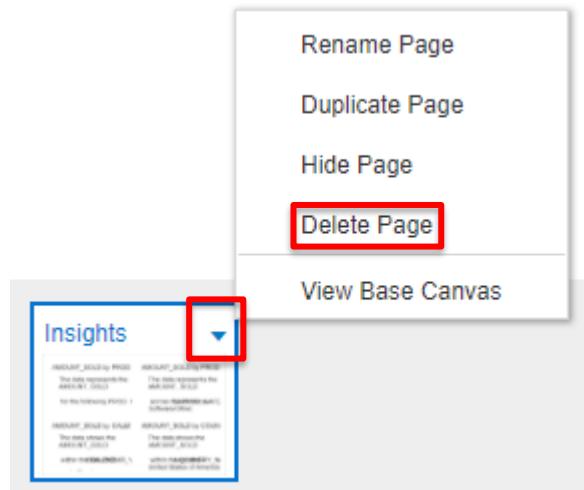


Na parte inferior, na esquerda, temos as propriedades da história, neste lugar é onde se permite fazer alterações nesta tela.

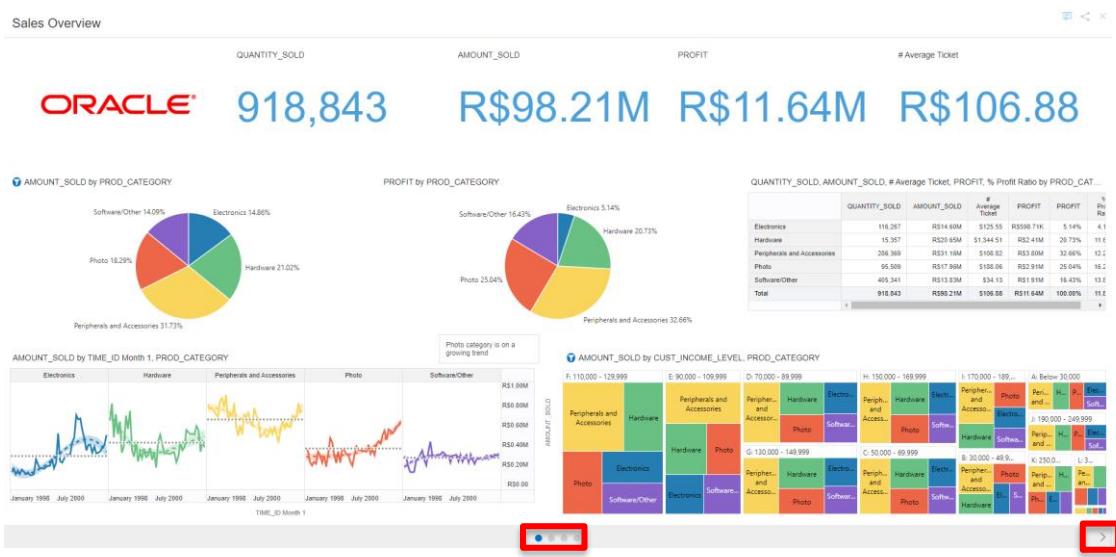
Se o usuário clicar no símbolo da nota, pode-se adicionar um *insight* a história. Como na imagem abaixo.



Caso o usuário veja que algum *canvas* não é importante para a história, ele pode retirar ele *canvas*, como na imagem abaixo.



Quando clicar em “Present”, a tela ficará da maneira como está na imagem abaixo.



A navegação pode ser feita clicando nos círculos azuis na parte inferior da tela, ou pelas setas do teclado.

# Lab 5.

## Explorando o Data Flow e o Machine Learning



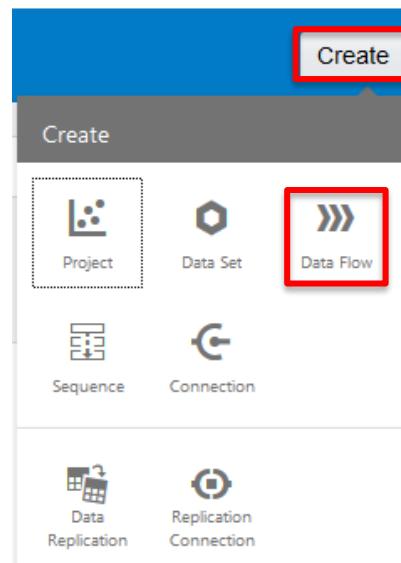
# Lab 5. Explorando o Data Flow e o Machine Learning

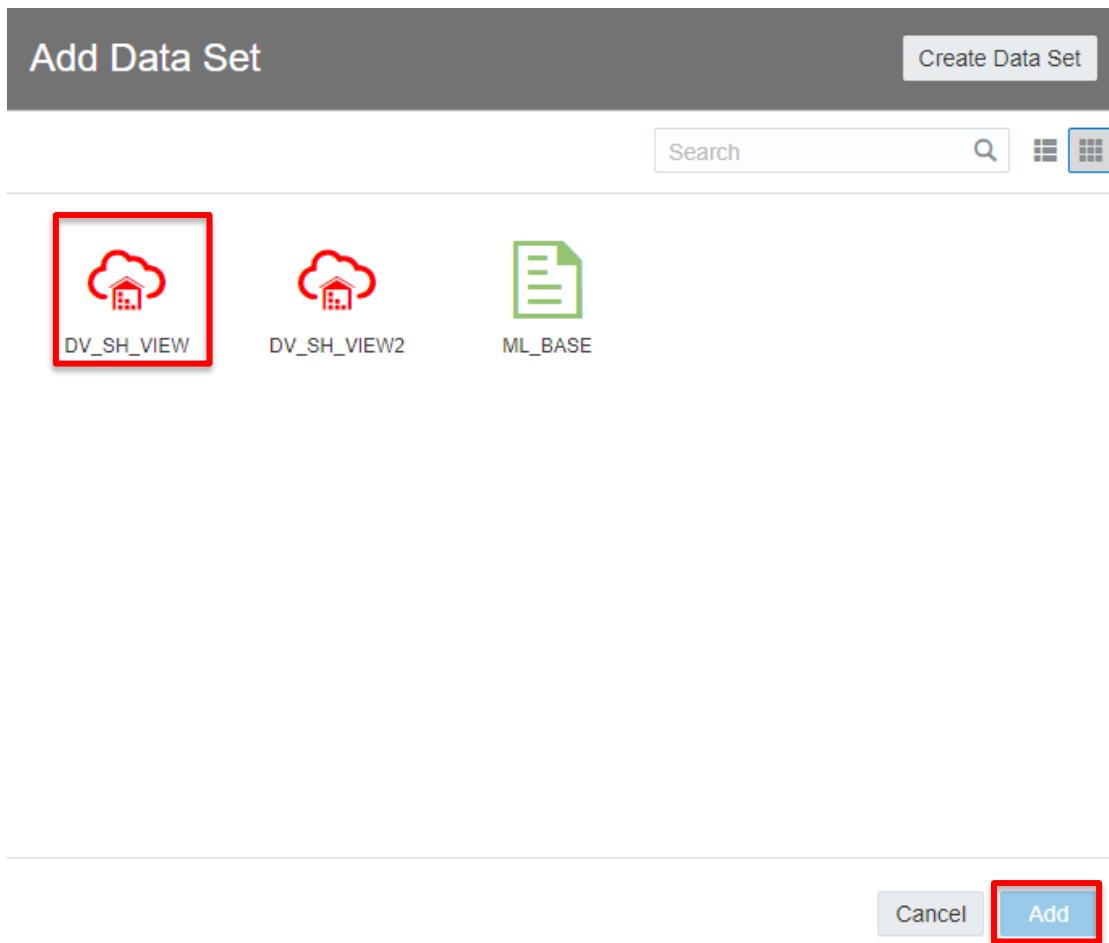
## Conhecendo o Data Flow

### Objetivos

- Entender o funcionamento do Data Flow
- Como treinar uma base para *Machine Learning*.

Para criar um Data Flow no Oracle Analytics Cloud siga o passo a passo abaixo.

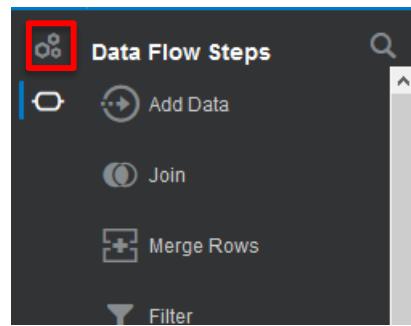




Feito isso, o usuário terá uma tela como na imagem abaixo.

PROD_NAME	PROD_DESC	PROD_CATEGORY	PROD_SUBCATEGORY	PROD_LIST_PRICE	QUANTITY SOLD	PROFIT
Home Theatre Package with DVD-Audio/Video Play	Home Theatre Package with DVD-Audio/Video Play	Electronics	Home Audio	599.99	1	30.00
Home Theatre Package with DVD-Audio/Video Play	Home Theatre Package with DVD-Audio/Video Play	Electronics	Home Audio	599.99	1	30.00
Home Theatre Package with DVD-Audio/Video Play	Home Theatre Package with DVD-Audio/Video Play	Electronics	Home Audio	599.99	1	-38.52
Home Theatre Package with DVD-Audio/Video Play	Home Theatre Package with DVD-Audio/Video Play	Electronics	Home Audio	599.99	1	28.90
Mouse Pad	Mouse Pad	Software/Other	Accessories	9.99	1	0.00
External EX CD-ROM	External EX CD-ROM	Peripherals and Accessories	CD-ROM	49.99	1	0.00
Standard Mouse	Standard Mouse	Peripherals and Accessories	Accessories	22.99	1	22.99

Na imagem acima é a parte da tela onde aparecerá a parte gráfica do passo a passo da transformação da base, conforme adicionarmos passos do Data Flow, novas caixas vão aparecer.



Nesta parte, temos o local onde está listado todos os passos possíveis de serem executados no Data Flow dentro da ferramenta. Caso clique nas engrenagens na imagem acima, será listado as colunas da tabela adicionada ao Data Flow.

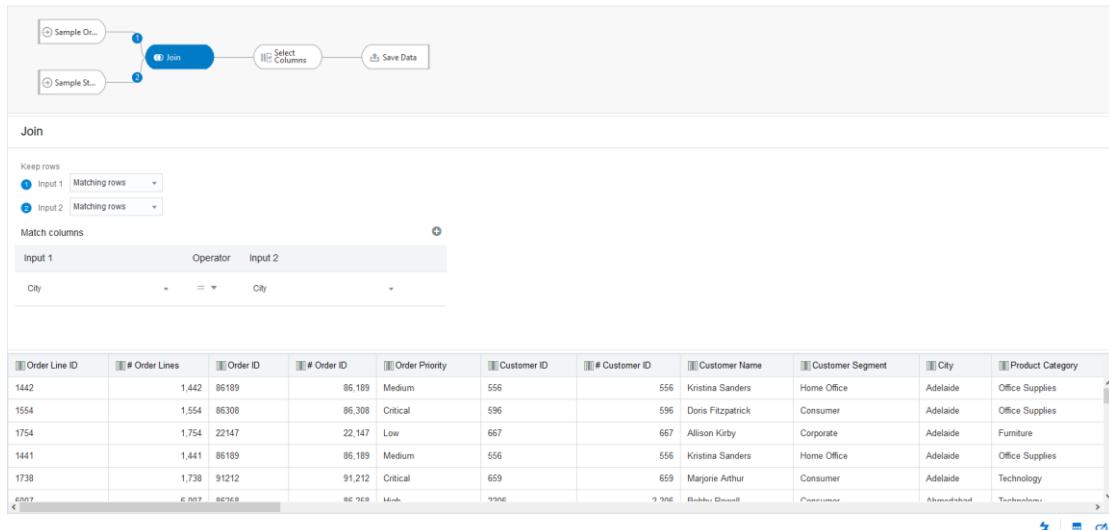
Column	Type
PROD_NAME	String
PROD_DESC	String
PROD_CATEGORY	String
PROD_SUBCATEGORY	String
PROD_LIST_PRICE	Decimal

Nesta parte, temos o editor do passo, isso seria o local onde o usuário edita o passo criado por ele. Por exemplo, caso ele crie um *Join*, será nesse editor que ele indicará as bases e qual a coluna que será a coluna de referência para o *Join*.

PROD_NAME	PROD_DESC	PROD_CATEGORY	PROD_SUBCATEGORY	PROD_LIST_PRICE	QUANTITY SOLD	PROFIT	AM
Home Theatre Package with DVD-Audio/Video Play	Home Theatre Package with DVD-Audio/Video Play	Electronics	Home Audio	599.99	1	30.00	629.99
Home Theatre Package with DVD-Audio/Video Play	Home Theatre Package with DVD-Audio/Video Play	Electronics	Home Audio	599.99	1	30.00	629.99
Home Theatre Package with DVD-Audio/Video Play	Home Theatre Package with DVD-Audio/Video Play	Electronics	Home Audio	599.99	1	-38.52	561.47
Home Theatre Package with DVD-Audio/Video Play	Home Theatre Package with DVD-Audio/Video Play	Electronics	Home Audio	599.99	1	28.50	628.89
Mouse Pad	Mouse Pad	Software/Other	Accessories	9.99	1	0.00	9.99
External BX CD-ROM	External BX CD-ROM	Peripherals and Accessories	CD-ROM	49.99	1	0.00	49.99
Standard Mouse	Standard Mouse	Peripherals and Accessories	Accessories	22.99	1	0.00	22.99

Nesta última parte da tela, é onde aparece uma amostra da base de dados que está sendo transformada.

Por exemplo, segue abaixo a imagem de como seria fazer um *Join* entre duas bases de dados para ficar apenas uma unificada.



Como pode ser observado, a parte do editor muda conforme o passo que está sendo executado.

Os passos possíveis de serem executados pelo *Data Flow* são:

- *Add Data*: adiciona novas bases de dados
- *Join*: faz a junção entre duas tabelas de dados, é necessário indicar a coluna que une as duas bases de dados.
- *Union Rows*: acrescenta uma base de dados a outra.
- *Filter*: faz um filtro na base com qualquer informação da mesma.
- *Aggregate*: agrupa a base de dados.
- *Save Data Set*: função para salvar a saída do Data Flow, a saída pode ser salva no *Storage* da própria ferramenta ou na conexão com a base de dados.
- *Create Essbase Cube*: salva a saída do Data Flow em uma aplicação em formato de Cubo de dados.
- *Add Columns*: adiciona colunas a base de dados, a coluna a ser adicionada é customizada pelo próprio usuário.
- *Select Columns*: permite selecionar as colunas da base de dados.
- *Rename Columns*: permite renomear o nome das colunas.
- *Transform Column*: permite transformar os dados de uma coluna sem precisar duplicá-la.
- *Merge Columns*: concatena duas colunas.
- *Bin*: agrupa em intervalos uma coluna de métrica.
- *Group*: agrupa uma coluna de atributo.
- *Branch*: permite criar “galhos” no Data Flow para o mesmo ter duas bases de dados diferentes como saída.
- *Cumulative Value*: acumula um valor no tempo (aplica a fórmula RSUM).
- *Time Series Forecast*: aplica scripts padrões de *Forecast* para prever valores, o script deve ser executado na coluna de métrica que deseja fazer a previsão e a mesma deve estar acumulada.

- *Analyze Sentiment*: script padrão da ferramenta que analisa o sentimento de uma coluna de atributo que contenha um texto livre, podendo ele ser Positivo, Negativo ou Neutro.
- *Train Numeric Prediction (ML)*: script de *Machine Learning* para treinar previsões numéricas na base de dados, scripts padrões disponíveis:
  - *Random Forest, Linear Regression, Elastic Net Linear Regression e CART*.
- *Train Multi-Classifier (ML)*: script de *Machine Learning* para multi-classificar a base de dados, scripts padrões disponíveis:
  - *Naive Bayes, Neural Network, Random Forest, SVM e CART*.
- *Train Clustering (ML)*: script de *Machine Learning* para treinar a criação de clusters na base de dados, scripts padrões disponíveis:
  - *K-Means e Hierarchical Clustering*.
- *Train Binary Classifier (ML)*: script de *Machine Learning* para treinar classificações binárias na base de dados, scripts padrões disponíveis:
  - *Logistic Regression, Naive Bayes, Neural Network, Random Forest, SVM e CART*.
- *Apply Model (ML)*: aplica o modelo previamente treinado em uma base que não tem a coluna treinada pelo modelo.

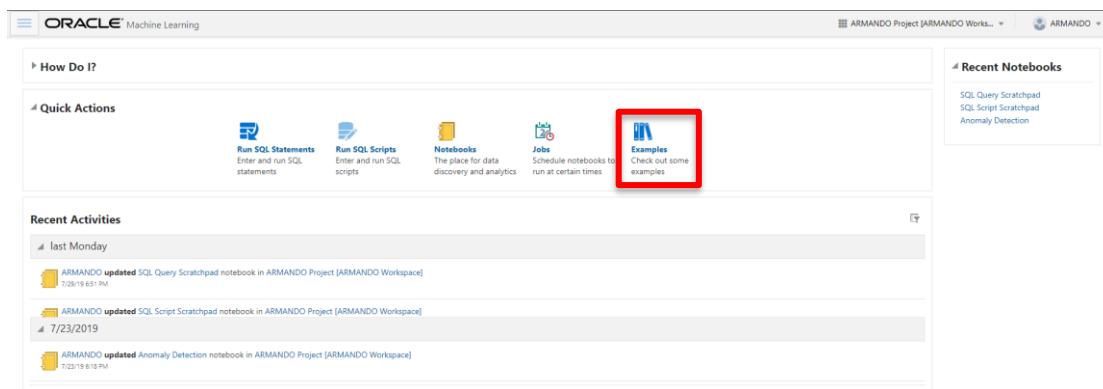
Para treinar um bom script de *Machine Learning*, é recomendado que a base de dados não tenha dados muito granulares, como por exemplo para atributos em nível de cidade devem-se reduzir essa granularidade para o nível de região ou país. Caso os dados temporais estejam com muita granularidade, uma sugestão seria criar uma coluna de meses que vai de 1-12 e também uma coluna com informação de ano. Essas são algumas sugestões para treinar um bom modelo.

Com o modelo treinado, ele automaticamente terá selecionado algumas colunas da base de dados que ele utiliza para “prever” a coluna aprendida, como por exemplo prever se uma pessoa pode ter uma doença cardíaca. Tem-se uma base de dados com todas as informações das pessoas e o que ocorreu com elas, o usuário treina o modelo em cima da coluna do que ocorreu com essas pessoas e utiliza esse modelo treinado em uma outra base com as mesmas informações, sendo que ainda não sabemos o que pode ocorrer. O modelo nos dá o que pode ocorrer com essas pessoas, com uma certa porcentagem de precisão.

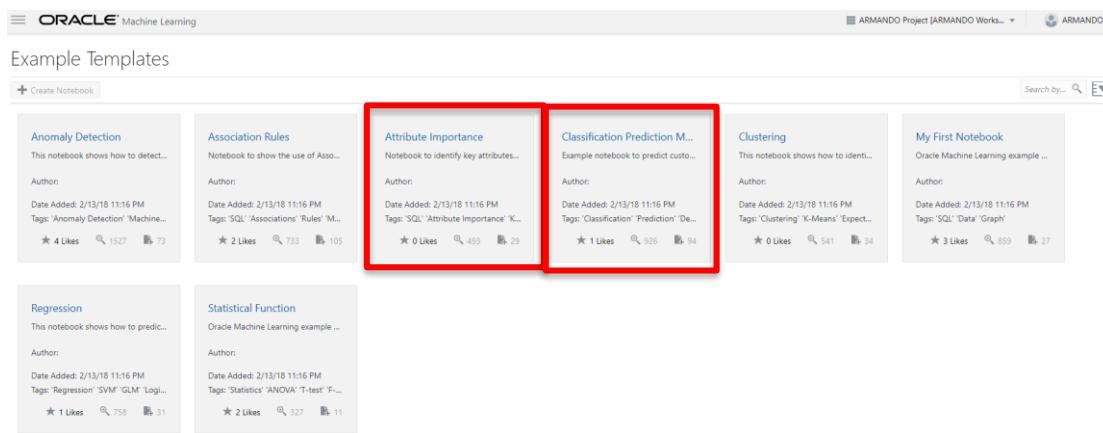
# Criando um modelo de Machine Learning no Autonomous Data Warehouse

Para utilizar do Machine Learning do Autonomous Data Warehouse, o usuário utilizará o OML com o usuário que já foi criando anteriormente neste guia.

Feito o login no OML, será utilizado um dos Notebooks exemplo para mostrar a capacidade de execução de um script de Machine Learning no Autonomous, conforme já dito neste guia, atualmente as linguagens aceitas são SQL e PL/SQL e futuramente Python.



Será utilizado como exemplo os scripts de importância de atributos e Classificação.



É necessário fazer download do exemplo e fazer o upload do mesmo dentro da área de trabalho do usuário. Imagem ilustrativa abaixo.

Anomaly Detection

Anomaly Detection to Detect Suspicious or Rare Occurrences

This notebook shows how to detect rare records, customers or transactions using an unsupervised learning algorithm (1-Class Support Vector Machine). The notebook first builds a 1-Class SVM model and then applies the model to flag unusual or suspicious records. The anomaly detection model can also be applied to "score" new records. The entire machine learning methodology runs inside the Oracle Autonomous Data Warehouse Cloud (ADWC).

By Charlie Berger, updated Oct 18, 2018

Task 0 sec. Last updated by CHARLIE at October 18 2018, 2:05:05 PM (optional)

For more information, check the Oracle ADWC Documentation <https://docs.oracle.com/en/cloud/paaS/autonomous-data-warehouse-cloud/index.html>, Oracle Machine Learning folder on Oracle on Github <https://github.com/oracle/OracleAdvancedAnalytics>, Oracle Technology Network and Introducing Oracle Machine Learning blog post

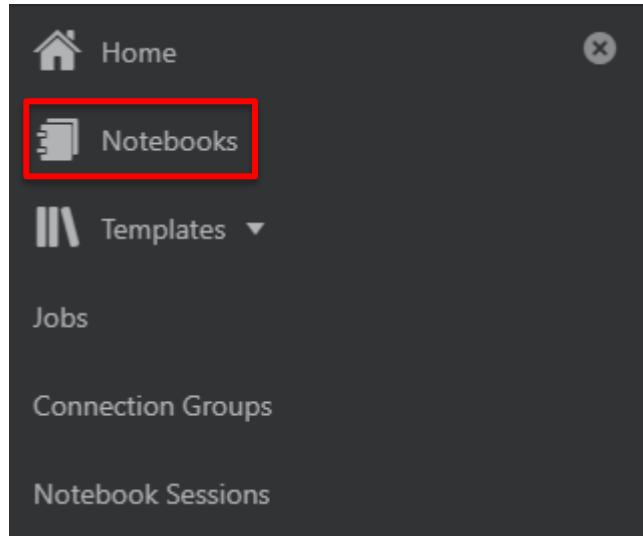
Task 0 sec. Last updated by CHARLIE at October 18 2018, 2:43:54 PM

Clean up and drop any table if previously exists for notebook reproducibility

```

1-->1
2BEGIN
3  EXECUTE IMMEDIATE 'DROP Table SUPPLEMENTARY_DEMOGRAPHICS2';
4  EXCEPTION
5    WHEN OTHERS THEN NULL;
6END;
7
8PL/SQL procedure successfully completed.
9
10Task 0 sec. Last updated by CHARLIE at October 18 2018, 2:43:54 PM
11
12Create SUPPLEMENTARY_DEMOGRAPHICS2 table that remove COMMENTS unstructured data for simplicity.
13
14-->1
15CREATE Table SUPPLEMENTARY_DEMOGRAPHICS2
16AS (SELECT AFFINITY_CARD, BOOKKEEPING_APPLICATION, BULK_PACK_DISKETTES, CUST_ID, EDUCATION, FLAT_PANEL_MONITOR, HOME_THEATER_PACKAGE, HOUSEHOLD_SIZE, OCCUPATION, OS_DOC_SET_XANII, PRINTER_SUPPLIES, YRS_RESIDENCE, YR_BI_SAVES
17  FROM SH.SUPPLEMENTARY_DEMOGRAPHICS);
18
19Updated 4580 row(s).
20
21Task 0 sec. Last updated by CHARLIE at October 18 2018, 2:43:55 PM.

```



Name	Comment	Last Update	Updated By	Connection Group
SQL Query Scratchpad	Scratch pad notebook to run SQL statements.	7/29/19 6:51 PM	ARMANDO	Global
SQL Script Scratchpad	Scratch pad notebook to run SQL scripts.	7/29/19 6:50 PM	ARMANDO	Global

A extensão de importação dos Notebooks é no formato JSON.

Feito a importação a console informa que foi importado com sucesso.

Name	Comment	Last Update	Updated By	Connection Group
Anomaly Detection		7/31/19 11:42 PM	ARMANDO	Global
SQL Query Scratchpad	Scratch pad notebook to run SQL statements.	7/29/19 6:51 PM	ARMANDO	Global
SQL Script Scratchpad	Scratch pad notebook to run SQL scripts.	7/29/19 6:50 PM	ARMANDO	Global

Clique no nome do Notebook recém adicionado e realize duas ações, primeiro conecte o Notebook em alguma das conexões, High, Medium ou Low, neste caso será utilizado a High.

Anomaly Detection

Settings

Interpreter binding

Bind interpreter for this note. Click to Bind/Unbind interpreter. Drag and drop to reorder interpreters. The first interpreter on the list becomes default.

md %md (default)

orcladw\_low %sql (default), %script

orcladw\_medium %sql (default), %script

orcladw\_high %sql (default), %script

Save Cancel

Limpe os *outputs* do Notebook pois as informações de importação vieram junto com o JSON, será executado a limpeza dos *outputs* para que o Autonomous execute esta sequência de queries.

Depois disso execute todos os parágrafos e veja os resultados, tudo o que foi criado por esta interface gráfica está registrado no banco, para checar isso basta acessar o usuário que executou as queries pelo SQL Developer.

Anomaly Detection

▶ ✎ ⌂ ⌂ 🔍 ⌂ 🔍 default ▾

No próprio Notebook o usuário pode ver quais queries foram executadas e se foram executadas com sucesso e futuramente utilizar a mesma estrutura e adaptar o script para o seu modelo de dados, até mesmo melhorar este script de exemplo.

Conforme o tutorial acima, será utilizado os exemplos de importância de atributo e Classificação, ambos serão carregados conforme o tutorial descrito acima. Em ambos o template foi usado como referência, todos os modelos que serão executados houve alterações.

As bases de dados a serem utilizadas serão as que foram carregadas no Autonomous no passo anterior neste mesmo guia, são as “Heart Disease likelyhood Training” e “Heart Disease likelyhood Predict”.

Primeiro foi executado o modelo de importância de atributo. Inicialmente é feito uma exploração dos dados que serão usados para o script que reconhece qual atributo é mais importante. A base utilizada será a de predisposição para ter doença do coração e o que queremos descobrir é quais atributos são mais importantes para o resultado da coluna “LIKELYHOOD”.

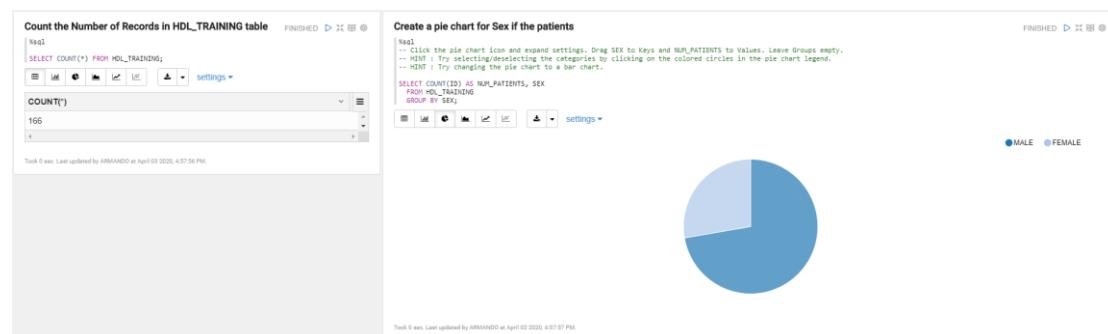
Query: SELECT \* FROM HDL\_TRAINING;

Display the HDL_TRAINING table												
ID	AGE	SEX	CHEST_PAIN_TYPE	RESTING_BP	CHOLESTROL	FASTING_BLOOD_SUGAR	RESTECG	MAXIMUM_HEART RATE	EXERCISE_INDUCED_ANGINA	OLDPEAK	LIKELYHOOD	
1	63	MALE	Typical Angina	145	233	Yes	Ventricular Hypertrophy	150	No	2.3		
2	67	MALE	Asymptomatic	160	286	No	Ventricular Hypertrophy	108	Yes	1.5		
3	67	MALE	Asymptomatic	120	229	No	Ventricular Hypertrophy	129	Yes	2.6		
4	37	MALE	Non-Anginal	130	250	No	Normal	187	No	3.5		
5	41	FEMALE	Atypical Angina	130	204	No	Ventricular Hypertrophy	172	No	1.4		

Query: SELECT COLUMN\_ID ID, COLUMN\_NAME NAME, DATA\_TYPE TYPE, DATA\_LENGTH LENGTH, NULLABLE FROM ALL\_TAB\_COLUMNS WHERE OWNER='ARMANDO' AND TABLE\_NAME='HDL\_TRAINING' ORDER BY COLUMN\_ID;

List column in the HDL_TRAINING table					
ID	NAME	TYPE	LENGTH	NULLABLE	
1	ID	NUMBER	22	Y	
2	AGE	NUMBER	22	Y	
3	SEX	VARCHAR2	26	Y	
4	CHEST_PAIN_TYPE	VARCHAR2	26	Y	
5	RESTING_BP	NUMBER	22	Y	
6	CHOLESTROL	NUMBER	22	Y	
7	FASTING_BLOOD_SUGAR	VARCHAR2	26	Y	
8	RESTECG	VARCHAR2	26	Y	

Query\_1: SELECT COUNT(\*) FROM HDL\_TRAINING; Query\_2: SELECT COUNT(ID) AS NUM\_PATIENTS, SEX FROM HDL\_TRAINING GROUP BY SEX;



Query\_1: CREATE OR REPLACE VIEW HDLTRAININGAI\_V AS (SELECT \* FROM HDL\_TRAINING);

Query\_2:

BEGIN

DBMS\_DATA\_MINING.DROP\_MODEL('AI\_EXPLAIN\_OUTPUT');

EXCEPTION WHEN OTHERS THEN NULL; END;

/

DECLARE

v\_setlst DBMS\_DATA\_MINING.SETTING\_LIST;

BEGIN

v\_setlst('ALGO\_NAME') := 'ALGO\_AI\_MDL';

V\_setlst('PREP\_AUTO') := 'ON';

DBMS\_DATA\_MINING.CREATE\_MODEL2(

MODEL\_NAME => 'AI\_EXPLAIN\_OUTPUT',

MINING\_FUNCTION => 'ATTRIBUTE\_IMPORTANCE',

DATA\_QUERY => 'select \* from HDLTRAININGAI\_V',

SET\_LIST => v\_setlst,

CASE\_ID\_COLUMN\_NAME => 'ID',

TARGET\_COLUMN\_NAME => 'LIKELYHOOD');

END;

```

Create HDLTRAININGAI_V view
-----
--> /nscript
-----/nscript
CREATE OR REPLACE VIEW HDLTRAININGAI_V
AS (SELECT *
     FROM HDL_TRAINING);
-----/nscript

View HDLTRAININGAI_V created.

-----/nscript

Took 1 sec. Last updated by ARMANDO at April 03 2020, 4:57:58 PM.
-----/nscript

Drop, then compute Attribute Importance to predict LIKELYHOOD
-----
BEGIN
  DBMS_DATA_MINING.DROP_MODEL('AI_EXPLAIN_OUTPUT');
  EXCEPTION WHEN OTHERS THEN NULL; END;
/
DECLARE
  v_setlst DBMS_DATA_MINING.SETTING_LIST;
BEGIN
  v_setlst('ALGO_NAME') := 'ALGO_AI_MDL';
  V_setlst('PREP_AUTO') := 'ON';

  DBMS_DATA_MINING.CREATE_MODEL2(
    MODEL_NAME          => 'AI_EXPLAIN_OUTPUT',
    MINING_FUNCTION    => 'ATTRIBUTE_IMPORTANCE',
    DATA_QUERY          => 'select * from HDLTRAININGAI_V',
    SET_LIST            => v_setlst,
    CASE_ID_COLUMN_NAM => 'ID',
    TARGET_COLUMN_NAME  => 'LIKELYHOOD');
END;
-----/nscript

PL/SQL procedure successfully completed.

-----/nscript

Took 13 sec. Last updated by ARMANDO at April 03 2020, 4:58:11 PM.
-----/nscript
  
```

Query: SELECT \* FROM DM\$VAAI\_EXPLAIN\_OUTPUT;

PARTITION_NAME	ATTRIBUTE_NAME	ATTRIBUTE_SUBNAME	ATTRIBUTE_IMPORTANCE_VALUE	ATTRIBUTE_RANK
	CHEST_PAIN_TYPE		0.06323281388788007	1
	NUMOF_VESSELS_COLORED_BY_FLOROSCOPY		0.04296810792134553	2
	MAXIMUM_HEARTRATE		0.0	3
	THAL		0.0	3
	EXERCISE_INDUCED_ANGINA		0.0	3
	SLOPE		0.0	3
	CHOLESTROL		0.0	3
	AGE		0.0	3

Query: SELECT \* FROM DM\$VAAI\_EXPLAIN\_OUTPUT  
WHERE ATTRIBUTE\_IMPORTANCE\_VALUE > 0;



Depois foram realizados mais dois modelos, ambos de Classificação, um é um script de Decision Tree o outro é um SVM (Support Vector Machine) o treino desses modelos ensinará para o Machine Learning a coluna de “LIKELYHOOD”.

Inicialmente é feito uma exploração dos dados, feito isso o modelo é treinado e depois testado nos dados conforme as querys e imagens abaixo. O primeiro modelo treinado foi o de Decision Tree.

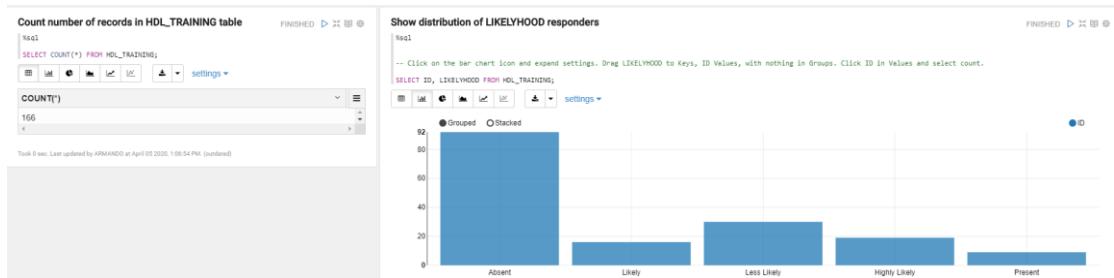
Query: SELECT \* FROM HDL\_TRAINING;

Display the HDL\_TRAINING data

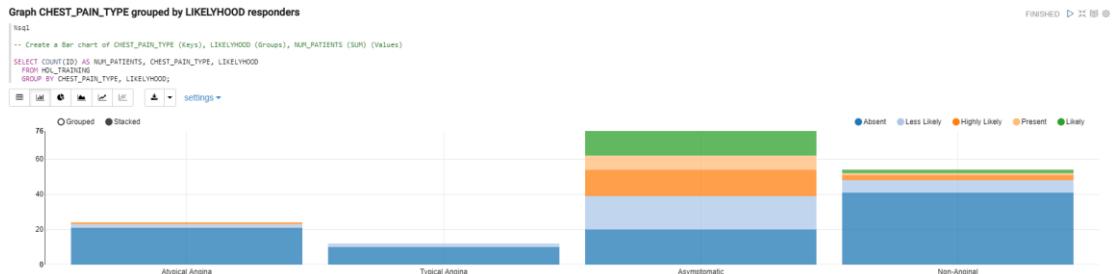
```
-- Click the table icon. Drag ID to Keys, AGE, SEX, CHEST_PAIN_TYP.., RESTING_BP, CHOLESTROL, FAS... to Values, with nothing in Groups.
SELECT * FROM HDL_TRAINING;
```

ID	AGE	SEX	CHEST_PAIN_TYP..	RESTING_BP	CHOLESTROL	FAS...	RESTECG	MAXIMUM_HEARTRATE	EXERCISE_INDUCED_ANGINA	OLDPEAK
1	60	MALE	Typical Angina	145	233	Yes	Ventricular Hypertrophy	150	No	2.3
2	67	MALE	Asymptomatic	160	286	No	Ventricular Hypertrophy	108	Yes	1.5
3	67	MALE	Asymptomatic	120	229	No	Ventricular Hypertrophy	129	Yes	2.6
4	37	MALE	Non-Anginal	130	250	No	Normal	187	No	3.5
5	41	FEMALE	Atypical Angina	130	204	No	Ventricular Hypertrophy	172	No	1.4

Query\_1: SELECT COUNT(\*) FROM HDL\_TRAINING; Query\_2: SELECT ID, LIKELYHOOD FROM HDL\_TRAINING;



Query: SELECT COUNT(ID) AS NUM\_PATIENTS, CHEST\_PAIN\_TYPE, LIKELYHOOD FROM HDL\_TRAINING GROUP BY CHEST\_PAIN\_TYPE, LIKELYHOOD;



```

Query: DECLARE
  v_sql varchar2(200);
  v_setlst DBMS_DATA_MINING.SETTING_LIST;

BEGIN
  -- Drop apply result
  BEGIN
    v_sql := 'DROP TABLE N1_APPLY_RESULT PURGE';
    EXECUTE IMMEDIATE v_sql;
    DBMS_OUTPUT.PUT_LINE (v_sql ||': succeeded');
  EXCEPTION
    WHEN OTHERS THEN
      DBMS_OUTPUT.PUT_LINE (v_sql ||': drop unnecessary - no table exists');
  END;

  -- Drop lift result
  BEGIN
    v_sql := 'DROP TABLE N1_LIFT_TABLE PURGE';
    EXECUTE IMMEDIATE v_sql;
    DBMS_OUTPUT.PUT_LINE (v_sql ||': succeeded');
  EXCEPTION
    WHEN OTHERS THEN
      DBMS_OUTPUT.PUT_LINE (v_sql ||': drop unnecessary - no table exists');
  END;
END;

```

## Drop existing objects in preparation for model building

```
%script

DECLARE
  v_sql varchar2(200);
  v_setlst DBMS_DATA_MINING.SETTING_LIST;

BEGIN
  -- Drop apply result

  BEGIN
    v_sql := 'DROP TABLE N1_APPLY_RESULT PURGE';
    EXECUTE IMMEDIATE v_sql;
    DBMS_OUTPUT.PUT_LINE (v_sql || ': succeeded');
  EXCEPTION
    WHEN OTHERS THEN
      DBMS_OUTPUT.PUT_LINE (v_sql || ': drop unnecessary - no table exists');
  END;

  -- Drop lift result

  BEGIN
    v_sql := 'DROP TABLE N1_LIFT_TABLE PURGE';
    EXECUTE IMMEDIATE v_sql;
    DBMS_OUTPUT.PUT_LINE (v_sql || ': succeeded');
  EXCEPTION
    WHEN OTHERS THEN
      DBMS_OUTPUT.PUT_LINE (v_sql || ': drop unnecessary - no table exists');
  END;
END;

DROP TABLE N1_APPLY_RESULT PURGE: succeeded
DROP TABLE N1_LIFT_TABLE PURGE: drop unnecessary - no table exists

PL/SQL procedure successfully completed.
```

Query: BEGIN

```
EXECUTE IMMEDIATE 'CREATE OR REPLACE VIEW N1_TRAIN_DATA AS
SELECT * FROM HDL_TRAINING SAMPLE (60) SEED (1)';
DBMS_OUTPUT.PUT_LINE ('Created N1_TRAIN_DATA');
EXECUTE IMMEDIATE 'CREATE OR REPLACE VIEW N1_TEST_DATA AS
SELECT * FROM HDL_TRAINING MINUS SELECT * FROM N1_TRAIN_DATA';
DBMS_OUTPUT.PUT_LINE ('Created N1_TEST_DATA');

END;
```

**Create train and test data sets (60/40) for model build and test**

```
%script
-- Build a classification model and then generate a lift test result and an apply result.
-- Split the Data into N1_TRAIN_DATA (randomly selected 60% of the data) and N1_TEST_DATA (remaining 40%)
BEGIN
EXECUTE IMMEDIATE 'CREATE OR REPLACE VIEW N1_TRAIN_DATA AS SELECT * FROM HDL_TRAINING SAMPLE (60) SEED (1)';
DBMS_OUTPUT.PUT_LINE ('Created N1_TRAIN_DATA');
EXECUTE IMMEDIATE 'CREATE OR REPLACE VIEW N1_TEST_DATA AS SELECT * FROM HDL_TRAINING MINUS SELECT * FROM N1_TRAIN_DATA';
DBMS_OUTPUT.PUT_LINE ('Created N1_TEST_DATA');
END;
Created N1_TRAIN_DATA
Created N1_TEST_DATA

PL/SQL procedure successfully completed.
```

```
Query: BEGIN DBMS_DATA_MINING.DROP_MODEL('N1_CLASS_MODEL');
EXCEPTION WHEN OTHERS THEN NULL; END;
/
DECLARE
  v_setlst DBMS_DATA_MINING.SETTING_LIST;

BEGIN
  v_setlst('PREP_AUTO') := 'ON';
  v_setlst('ALGO_NAME') := 'ALGO_DECISION_TREE';

  DBMS_DATA_MINING.CREATE_MODEL2(
    'N1_CLASS_MODEL',
    'CLASSIFICATION',
    'SELECT * FROM N1_TRAIN_DATA',
    v_setlst,
    'ID',
    'LIKELYHOOD');
END;
```

## Drop, then build a Decision Tree model

```
%script

BEGIN DBMS_DATA_MINING.DROP_MODEL('N1_CLASS_MODEL');
EXCEPTION WHEN OTHERS THEN NULL; END;
/
DECLARE
    v_setlst DBMS_DATA_MINING.SETTING_LIST;

BEGIN
    v_setlst('PREP_AUTO') := 'ON';
    v_setlst('ALGO_NAME') := 'ALGO_DECISION_TREE';

    DBMS_DATA_MINING.CREATE_MODEL2(
        'N1_CLASS_MODEL',
        'CLASSIFICATION',
        'SELECT * FROM N1_TRAIN_DATA',
        v_setlst,
        'ID',
        'LIKELYHOOD');
END;
```

PL/SQL procedure successfully completed.

-----  
PL/SQL procedure successfully completed.

Query: DECLARE

```
v_sql varchar2(200);
BEGIN
    v_sql := 'DROP TABLE N1_APPLY_RESULT PURGE';
    EXECUTE IMMEDIATE v_sql;
    EXCEPTION WHEN OTHERS THEN NULL;
END;
/

-- Score the data
BEGIN

DBMS_DATA_MINING.APPLY('N1_CLASS_MODEL','N1_TEST_DATA','ID','N1_
APPLY_RESULT');
END;
```

## Evaluate the model

```
%script

-- Drop apply result table
DECLARE
    v_sql varchar2(200);
BEGIN
    v_sql := 'DROP TABLE N1_APPLY_RESULT PURGE';
    EXECUTE IMMEDIATE v_sql;
    EXCEPTION WHEN OTHERS THEN NULL;
END;
/

-- Score the data
BEGIN
    DBMS_DATA_MINING.APPLY('N1_CLASS_MODEL','N1_TEST_DATA','ID','N1_APPLY_RESULT');
END;
```

PL/SQL procedure successfully completed.

-----  
PL/SQL procedure successfully completed.

Query: `SELECT ID, PREDICTION, ROUND(PROPABILITY,3) PROB,  
ROUND(COST,2) COST  
FROM N1_APPLY_RESULT WHERE PREDICTION =  
${PREDICTION='Absent','Absent'|'Less Likely'|'Likely'|'Highly Likely'|'Present'}  
AND PROPABILITY > 0.5  
ORDER BY PROPABILITY DESC;`

Display all patients with likelihood > 50% to be HDL responders

```
sql
SELECT ID, PREDICTION, ROUND(PROPABILITY,3) PROB, ROUND(COST,2) COST
FROM N1_APPLY_RESULT WHERE PREDICTION = ${PREDICTION='Absent','Absent'|'Less Likely'|'Likely'|'Highly Likely'|'Present'} AND PROPABILITY > 0.5
ORDER BY PROPABILITY DESC;
```

FINISHED D X 单 ①

ID	PREDICTION	PROB	COST
5	Absent	0.839	0.16
165	Absent	0.839	0.16
17	Absent	0.839	0.16
21	Absent	0.839	0.16
26	Absent	0.839	0.16
28	Absent	0.839	0.16
35	Absent	0.839	0.16
43	Absent	0.839	0.16

Query: `SELECT A.* , B.*  
FROM N1_APPLY_RESULT A, N1_TEST_DATA B  
WHERE PREDICTION = ${PREDICTION='Absent','Absent'|'Less  
Likely'|'Likely'|'Highly Likely'|'Present'} AND A.ID = B.ID;`

Interactively show patient probability to be HDL responders

```
sql
SELECT A.* , B.*
FROM N1_APPLY_RESULT A, N1_TEST_DATA B
WHERE PREDICTION = ${PREDICTION} AND A.ID = B.ID;
```

PREDICTION

ID	PREDICTION	PROBABILITY	COST	AGE	SEX	CHEST_PAIN_TYP...	RESTING_BP	CHOLESTROL	FASTING_BLOOD_SUG...	RESTECG	MAXIMUM_HI
80	Absent	0.2765957446808511	0.7234042553191489	58	MALE	Asymptomatic	150	270	No	Ventricular Hypertrophy	111
165	Absent	0.8392857142857143	0.1607142857142857	48	MALE	Non-Anginal	124	255	Yes	Normal	175
55	Absent	0.2765957446808511	0.7234042553191489	60	MALE	Asymptomatic	130	253	No	Normal	144
67	Absent	0.8392857142857143	0.1607142857142857	60	MALE	Non-Anginal	140	185	No	Ventricular Hypertrophy	155
97	Absent	0.2765957446808511	0.7234042553191489	59	MALE	Asymptomatic	110	239	No	Ventricular Hypertrophy	142
60	Absent	0.8392857142857143	0.1607142857142857	51	MALE	Typical Angina	125	213	No	Ventricular	125

Query: SELECT A.\* , B.\*  
 FROM N1\_APPLY\_RESULT A, N1\_TEST\_DATA B  
 WHERE CHEST\_PAIN\_TYPE = \${CHEST\_PAIN\_TYPE='Typical Angina','Typical Angina'|'Atypical Angina'|'Non-Anginal'|'Asymptomatic'}  
 AND A.ID = B.ID;

Interactive selection of likely high HDL responders selected by CHEST\_PAIN\_TYPE

```
sql
-- View the probability of AFFINITY_CARD responders by selecting HOUSEHOLD_SIZE
SELECT A.* , B.*
FROM N1_APPLY_RESULT A, N1_TEST_DATA B
WHERE CHEST_PAIN_TYPE = ${CHEST_PAIN_TYPE='Typical Angina','Typical Angina'|'Atypical Angina'|'Non-Anginal'|'Asymptomatic'}
AND A.ID = B.ID;
```

CHEST\_PAIN\_TYPE

ID	PREDICTION	PROBABILITY	COST	AGE	SEX	CHEST_PAIN_TYP...	RESTING_BP	CHOLESTROL	FASTING_BLOOD_SUG...	RESTECG	MAXIMUM_HI
80	Absent	0.2765957446808511	0.7234042553191489	58	MALE	Asymptomatic	150	270	No	Ventricular Hypertrophy	111
80	Likely	0.1702127659574468	0.8297872340425532	58	MALE	Asymptomatic	150	270	No	Ventricular Hypertrophy	111
80	Present	0.1276595744680851	0.8723404255319149	58	MALE	Asymptomatic	150	270	No	Ventricular Hypertrophy	111
80	Less Likely	0.1702127659574468	0.8297872340425532	58	MALE	Asymptomatic	150	270	No	Ventricular Hypertrophy	111
80	Highly Likely	0.2553191489561702	0.846808510638299	58	MALE	Asymptomatic	150	270	No	Ventricular Hypertrophy	111

Para o treinamento do Segundo modelo, usando o algoritmo de SVM, o mesmo template foi usado, as mesmas querys foram executadas, a única diferença é que na hora do treino do modelo, muda-se um parâmetro da query, conforme na query e na imagem abaixo.

Query: BEGIN DBMS\_DATA\_MINING.DROP\_MODEL('N1\_CLASS\_MODEL');  
 EXCEPTION WHEN OTHERS THEN NULL; END;

/

DECLARE

  v\_setlst DBMS\_DATA\_MINING.SETTING\_LIST;

BEGIN

  v\_setlst('PREP\_AUTO') := 'ON';

  v\_setlst('ALGO\_NAME') := 'ALGO\_SUPPORT\_VECTOR\_MACHINES';

DBMS\_DATA\_MINING.CREATE\_MODEL2(

  'N1\_CLASS\_MODEL',

  'CLASSIFICATION',

  'SELECT \* FROM N1\_TRAIN\_DATA',

  v\_setlst,

  'ID',

```
'LIKELYHOOD');
```

```
END;
```

### Drop, then build a Support Vector Machine model

FINISHED ▶ ✎ 📄 ⚙

```
%script

BEGIN DBMS_DATA_MINING.DROP_MODEL('N1_CLASS_MODEL');
EXCEPTION WHEN OTHERS THEN NULL; END;
/
DECLARE
    v_setlst DBMS_DATA_MINING.SETTING_LIST;
BEGIN
    v_setlst('PREP_AUTO') := 'ON';
    v_setlst('ALGO_NAME') := 'ALGO_SUPPORT_VECTOR_MACHINES';

    DBMS_DATA_MINING.CREATE_MODEL2(
        'N1_CLASS_MODEL',
        'CLASSIFICATION',
        'SELECT * FROM N1_TRAIN_DATA',
        v_setlst,
        'ID',
        'LIKELYHOOD');

    END;
```

PL/SQL procedure successfully completed.

-----

PL/SQL procedure successfully completed.

## Criando um modelo de Machine Learning no Oracle Analytics Cloud

Para a criação de um modelo de Machine Learning no Oracle Analytics Cloud, deve seguir o procedimento mencionado acima e criar um Data Flow.

Neste exercício será usada as bases que foram carregadas no Autonomous Data Warehouse, são elas: “Heart Disease likelihood Training” e “Heart Disease likelihood Predict”.

Será construído dois Data Flows, um para treinar um script de CART e o segundo para treinar um script de SVM.

O primeiro Data Flow na imagem abaixo, o de treinamento de uma multi classificação com o algoritmo CART.



O Flow é simples, no primeiro passo é selecionado as colunas da base que serão trabalhadas, deixe todas selecionadas.

No segundo passo é inserido o passo que executa o treinamento do algoritmo, o treinamento escolhido foi de multi classificação, após a inserção do passo, é aberta uma janela para escolher qual script será utilizado para o treinamento do modelo, conforme a imagem abaixo, neste flow será utilizado o de CART. A coluna desejada para o aprendizado será a “LIKELYHOOD”. Todos os outros valores foram deixados como padrão.

### Train Multi-Classifier

**Model Training Script** **CART for model training**

\* Target **LIKELYHOOD**

target, the target(label) to learn/predict

Positive Class in Target **Present**

Positive class in the target value. Default is Yes.

Minimum Node Size **50**

Minimum Node Size

Maximum Depth **5**

Maximum Depth

Maximum Confidence **80**

Maximum Confidence

Train Partition Percent **80**

The percentage of original data used for training, default is 80%.

Balancing Method **Over Sample**

The mode method for balancing unbalanced set.

Feito isso no último passo o modelo será salvo, é apenas necessário dar um nome ao modelo de saída e uma descrição sugestiva, a descrição é opcional.

Agora será criado o segundo Data Flow para o treinamento do algoritmo de SVM será feita uma comparação dos modelos.



No primeiro passo é selecionado as colunas da base que serão trabalhadas, deixe todas selecionadas.

No segundo passo é inserido o passo que executa o treinamento do algoritmo, o treinamento escolhido foi de multi classificação, após a inserção do passo, é aberta uma janela para escolher qual script será utilizado para o treinamento do modelo, conforme a imagem abaixo, neste flow será utilizado o de SVM, a coluna que o script irá aprender é a “LIKELYHOOD”. Todos os outros valores foram deixados como padrão

### Train Multi-Classifier

Model Training Script [SVM for Classification](#)

\* Target **LIKELYHOOD**  
target, the target(label) to learn/predict

Positive Class in Target **Present**  
Positive class in the target value. Default is Yes.

Categorical Column Imputation **Most Frequent**  
The mode method for categorical features to fill NA. Two options: mostFrequent and leastFrequent. Default is mostFrequent.

Numerical Column Imputation **Mean**  
The mode method for numeric features to fill NA. Four options: mean, max, min, median. Default is mean.

Encoding Method **Indexer**  
Encoding method.

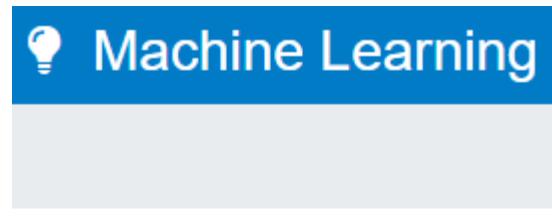
Maximum Null Value Percent **80**  
Maximum Null Value Percent

Train Partition Percent **80**  
The percentage of original data used for training, default is 80%.

Kernel Method **Linear**  
Kernel Method for SVM Classifier.

Feito isso no último passo o modelo será salvo, é apenas necessário dar um nome ao modelo de saída e uma descrição sugestiva, a descrição é opcional

Após executado os Data Flow, na aba de Machine Learning no menu principal da ferramenta, o modelo fica disponível para visualizar suas propriedades.

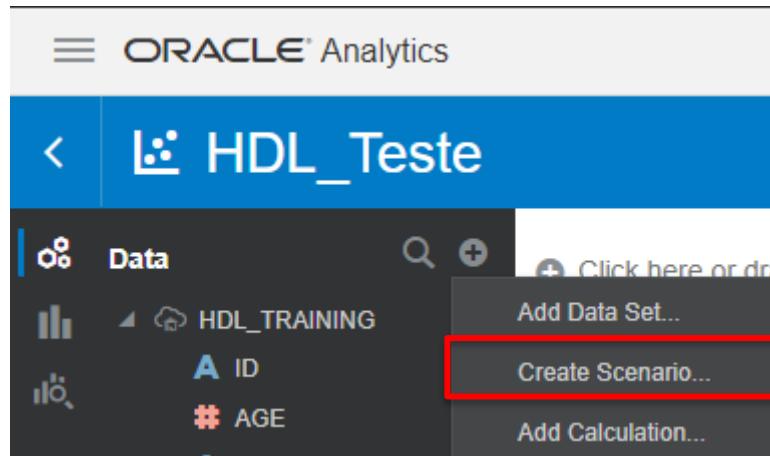


Type Name

	MC_SVM_HDL
	MC_CART_HDL

Clicando com o botão direito no modelo o usuário pode clicar em Inspect e ver todas as propriedades desse modelo, sua confiança, além de ver quais foram as matrizes de dados que ele vai utilizar e utilizou para os cálculos.

Para verificar como o modelo está funcionando, crie um projeto com a base de dados do banco que foi utilizada durante todo este guia, após isso crie um cenário.



Feito isso será aberto uma janela para indicar as colunas equivalentes da base que o modelo utiliza, na maioria das vezes o Oracle Analytics Cloud identifica automaticamente as colunas equivalentes entre o modelo e a base de dados. Mas caso a ferramenta não identifique será aberto a janela abaixo.

## Edit Scenario - Map Your Data

x

Select which Data Set you want to use with the Model

Data Set HDL\_TRAINING ▾

For each model input listed on the left, select a corresponding data element from your project

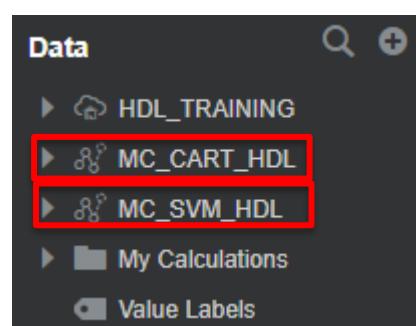
Model Input	Map To
CHEST_PAIN_TYPE	* CHEST_PAIN_TYPE
EXERCISE_INDUCED_ANGINA	* EXERCISE_INDUC...
FASTING_BLOOD_SUGAR	* FASTING_BLOOD_...
MAXIMUM_HEARTRATE	* MAXIMUM_HEART...
NUMOF_VESSELS_COLORED_BY_FLOUROSCOPY	* NUMOF_VESSELS_...
OLDPEAK	* OLDPEAK
RESTING_BP	* RESTING_BP
SLOPE	* SLOPE
AGE	Select Data

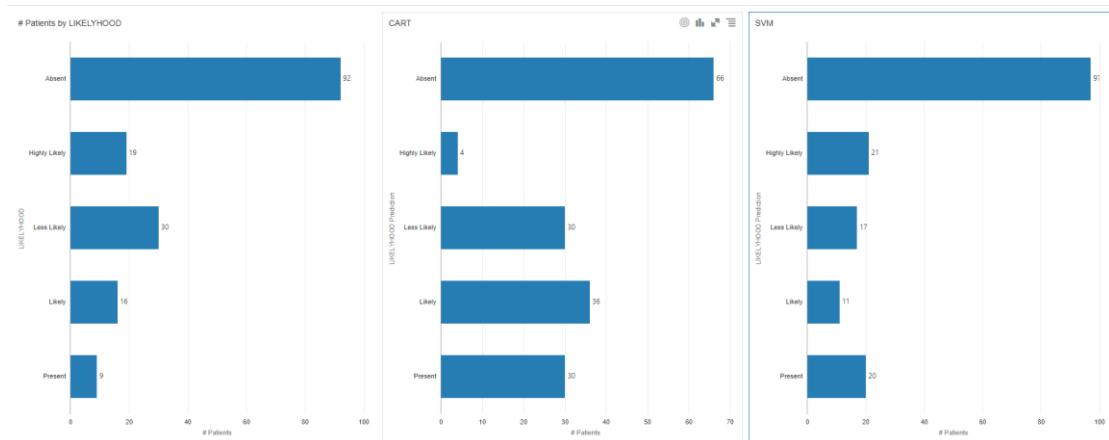
\* Required Fields

Done

Cancel

Feito isso o modelo aparecerá na parte dos elementos de dados e ele pode ser testado como um cenário no projeto, como pode ser visto na imagem abaixo.





Os modelos foram testados, como pode ser visto na imagem acima, da esquerda para a direita, o primeiro gráfico é o resultado real que está contido na própria base, o segundo gráfico é o modelo CART predizendo as pessoas que tem potencial para a doença de coração e o terceiro é o modelo SVM fazendo o mesmo.

## Lab 6.

# Comentários gerais sobre versão Enterprise e Administração do OAC



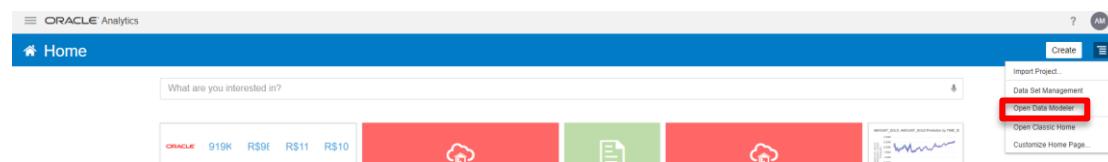
# Lab 6. Comentários gerais sobre a versão *Enterprise* e Administração do OAC

## Objetivos

- Entender as diferenças entre modelos *Professional* e *Enterprise*.
- Conceitos básicos para administrar a ferramenta.

## Versão *Enterprise*: Data Modeler

Para acessar o Data Modeler das suas bases de dados, vá em opções na “Home” e “Select Manage Models”.



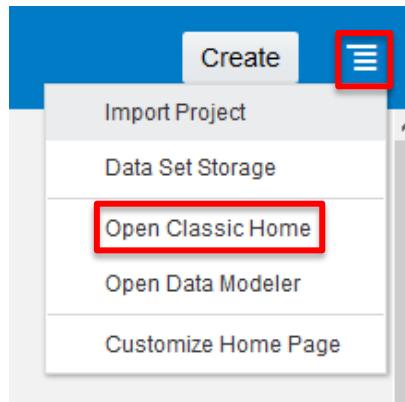
O Data Modeler é uma ferramenta da versão *Enterprise* do Oracle Analytics Cloud que auxilia ao usuário na modelagem de dados, na criação de esquemas estrela, e até sugerindo como o usuário pode construir o esquema.

Nele também é possível criar colunas novas e cálculos nas bases de dados diretamente dentro do banco, pois ele sempre sincroniza os dados.

O Data Modeler, é utilizado para a publicação de áreas de dados, é uma maneira de facilitar o *self-service BI* mantendo a integridade dos dados, pois o usuário final não tem acesso as bases em si, apenas à área publicada pelo administrador, permitindo que aumente a geração de *insights* corretos nas empresas, pois vários usuários poderão utilizar sem fazer alterações errôneas nas bases de dados.

## Versão *Enterprise*: Oracle Business Intelligence Enterprise Edition (OBIEE)

O antigo BICS/OBIEE hoje é acessado pelo botão Classic Home dentro do Analytics Cloud. É acessado na *Home Page*.



O OBIEE é uma ferramenta de criação de análises interativas que são agregadas em um *dashboard*, que também é interativo.

Recomendado para visualizações tabulares, já que o Data Visualization não permite mudar quase nada nas tabelas. Para visualizações gráficas, o Data Visualization já supera o OBIEE, quando utilizado a versão *Enterprise* do Oracle Analytics Cloud. É recomendado trabalhar com o Data Visualization e o OBIEE em conjunto, dependendo da necessidade.

O OBIEE tem a funcionalidade do “lbot”, gera relatórios automaticamente e os envia por e-mail para uma lista definida pelo desenvolvedor do “lbot”, possui geração automatizada de relatórios ad-hoc, que auxilia muito em entregas de relatórios periódicos nos quais o usuário final quer fazer uma análise ad-hoc (linha a linha).

Imagen da tela inicial do OBIEE.

A screenshot of the OBIEE home page. At the top, there's a header bar with the 'ORACLE Analytics' logo, a search bar, and navigation links like 'Home', 'Catalog', 'Favorites', 'Dashboards', 'New', 'Open', and 'Signed In As: Armando More'. The main content area has several sections: 'Create...' with icons for 'Analysis and Interactive Reporting', 'Published Reporting', and 'Actionable Intelligence'; 'Recent' with a 'Most Popular' section showing items like 'HandsOn\_ML' and 'HandsOn\_Test'; and a 'Jobs' section with links to 'Report Jobs' and 'Report Job History'. The overall layout is clean and modern, typical of enterprise BI software.

# Explorando o ADM Console

Nesta parte será dado um *overview* do console administrativo, que pode ser acessado conforme a imagem abaixo. Apenas administradores da ferramenta terão acesso a todas as opções que serão citadas.

The screenshot displays the Oracle Analytics Console interface. At the top left, there is a navigation bar with the following items:

- Analytics
- ☰ (Menu icon)
- ORACLE® Analytics

The main menu on the left side includes:

- Home
- Projects
- Data
- Machine Learning
- Jobs
- Console (highlighted with a red box)
- Academy

The right side of the interface is divided into two main sections:

- Visualizations**: A large area currently empty.
- Configuration**: A grid of configuration and administration options:

Visualizations and Sharing			
Maps	Extensions	Social	

Configuration and Administration			
Search Index	Safe Domains	Users and Roles	Snapshots
Connections	Virus Scanner	Session and Query Cache	Issue SQL
Monitor Deliveries	Mail Server	System Settings	Remote Data Connectivity

Segue a explicação geral de cada aba do console para agilizar a resolução de problemas com a ferramenta:

- *Maps*: permite carregar mapas que o *Data Visualization* não tem por padrão, como por exemplo um mapa detalhado das regiões do Brasil, além de ser o local onde é colocado as URLs das APIs do Google ou da Baidu para que o *Data Visualization* consiga utilizar seus fundos de mapa.
- *Extensions*: o usuário pode fazer *upload* de extensões para o *Data Visualization* que foram criadas por ele mesmo ou baixadas da biblioteca do OAC (URL: <https://www.oracle.com/solutions/business-analytics/data-visualization/extensions.html>)
- *Social*: permite ao administrador, conectar o Analytics a um espaço na Oracle Cloud que irá armazenar os projetos que serão compartilhados nas redes sociais. Nativamente o Analytics tem conectores com o Oracle Content, Slack, Twitter e LinkedIn.
- *Search Index*: permite ao administrador ver como está a indexação do seus dados e *dashboards*, além de monitorar os *Crawls*.
- *Safe Domains*: o administrador indica para o Analytics quais domínios são seguros, para que o OAC permita a impressão do *dashboard* na *webpage* destino, por padrão nenhum domínio vem permitido
- *Users and Roles*: Administrar usuários e qual o nível que eles podem desenvolver.
- *Snapshots*: Gerenciar os *Snapshots* (*Backups* da ferramenta, salva o estado no qual a ferramenta está).
- *Connections*: Gerenciar as conexões com outros *Databases* ou *Data Warehouses*, para utilizar o Data Modeler a conexão com o banco deve estar mapeada nesta parte da ferramenta.
- *Vírus Scanner*: Configurar um Antivírus para verificar todos os arquivos que foram carregados na ferramenta e impedir que ela seja atacada por algum arquivo malicioso (opcional).
- *Session and Query Cache*: Monitorar as *queries* solicitadas no Analytics, quais estão em cache e quais sessões estão ativas na ferramenta.
- *Issue SQL*: Solicitar alguma *query* para o banco no qual o Analytics está instalado para depurar algo, este banco não é visível para o usuário já que o Analytics é entregue para o cliente como PaaS.
- *Monitor Deliveries*: Monitora o compartilhamento de arquivos que foi feito via e-mail pelo Analytics.
- *Mail Server*: Configura o servidor SMTP para o Analytics realizar os disparos de e-mail.
- *System Settings*: Configurações avançadas do Analytics, consultar a documentação antes de alterar qualquer uma dessas opções.
- *Remote Data Connectivity*: Habilita a conexão com um banco de dados local onde foi instalado o Data Gateway ou Remote Data Connector.

# Lab 7.

## Terminando e apagando instâncias



# Lab 7. Terminando e apagando instâncias

## Objetivos

- Apagar a instância criada

Quando a instância for apagada, tudo desenvolvido nela será deletado para sempre. O usuário, para manter isso, pode fazer um *Snapshot* ou exportar todos os seus projetos. Nos dois casos, o usuário pode salvá-los no seu próprio computador.

Segue o passo a passo abaixo de como terminar sua instância de Oracle Analytics Cloud:

The screenshot shows the Oracle Cloud homepage with the navigation bar "ORACLE Cloud" and "Brazil East (Sao Paulo)". The main content area features several quick-start cards: "Instance" (2-6 mins, Always Free Eligible), "AUTONOMOUS TRANSACTION PROCESSING Create an ATP database" (3 mins, Always Free Eligible), "AUTONOMOUS DATA WAREHOUSE Create an ADW database" (3-5 mins, Always Free Eligible), "NETWORKING Set up a load balancer" (5 mins, Always Free Eligible), and "RESOURCE QUERY SERVICE View all my resources". On the left sidebar, under "Analytics", the "Analytics Cloud" option is highlighted with a red box. The URL in the address bar is <https://console.us-east-1.oraclecloud.com/analytics>.

The screenshot shows the "Analytics in Armando Compartiment" page. It includes a "List Scope" section with a dropdown set to "Armando" and a "Filters" section with dropdowns for "STATE" (Any state) and "FEATURE SET" (Any feature set). A table lists one instance: "QA&A" (Status: Active, OCID: 1...5gg5ehgqa, Feature Set: Enterprise Analytics, Capacity: 2 OCPU, Created: Mon, Mar 2, 2020, 4:18:09 PM UTC). To the right of the table are buttons for "Start", "Stop", "Update", and a red-bordered "Delete" button. The URL in the address bar is <https://console.us-east-1.oraclecloud.com/analytics/instances>.

Seguindo este passo a passo, a instância começará o processo de término.

Parabéns!  
Você completou o Laboratório  
*Hands-On Oracle Analytics Cloud*

