ORACLE

Modern Data Warehouse

Gain Insights from All Your Data

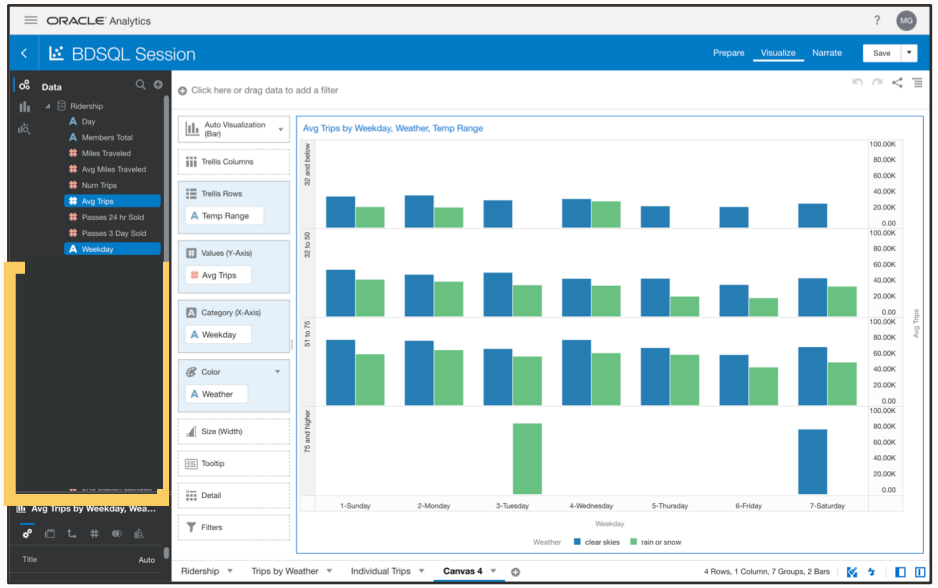
Name
Marty Gubar
Cloud SQL Product Management
Global Leaders - September 2020

Safe harbor statement

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

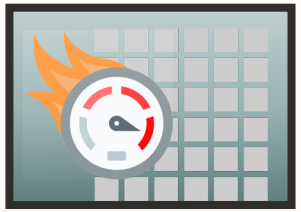
Modern Data Warehouse - Key Requirements

Expand depth and breadth of analyses



Breadth

Depth



Historical

Data Warehouse contains the critical data run your business

- Autonomous, self-driving, self-securing, self-repairing

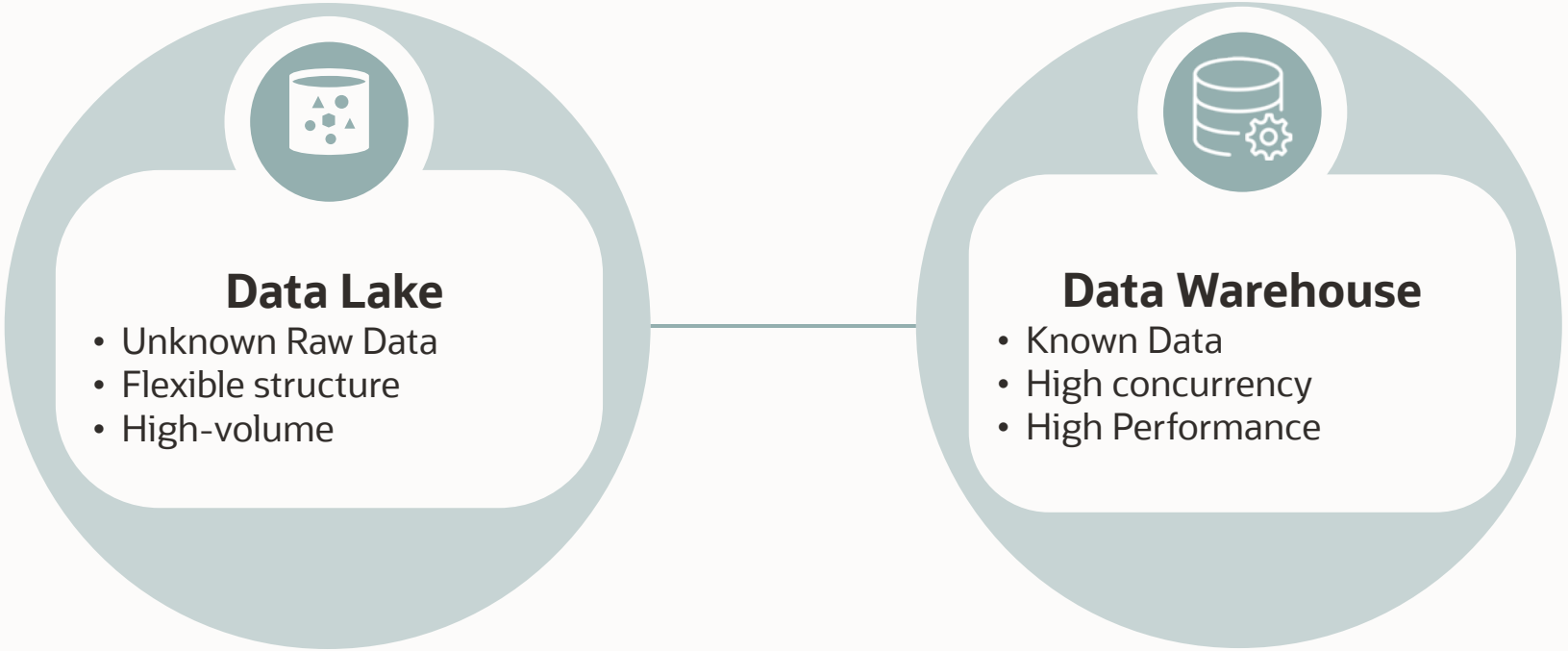
Immediately extend data warehouse analytics to new sources

- Opportunity to easily experiment
- Expand breadth – new attributes that potentially help understand or predict behavior
- Add depth – detailed events that are not yet of known value

Apply Oracle Database processing applied to shared data lake sources

- Another processing engine – like Spark

Data persistence and processing



Hadoop benefits and challenges

Benefits

- Capture all data
- Enables data democracy
- Process and analyze data at scale

Challenges

- Management of solution
- Governance - potential for data swamp
- Over-provisioning

Oracle Cloud Data Platform

Benefits

- Capture all data
- Enables data democracy
- Process and analyze data at scale
- **Fully managed services**
- **Unified metadata management**
- **Fully elastic compute – pay for use**

Challenges

- ~~Management of solution~~
- ~~Governance – potential for data swamp~~
- ~~Over-provisioning~~

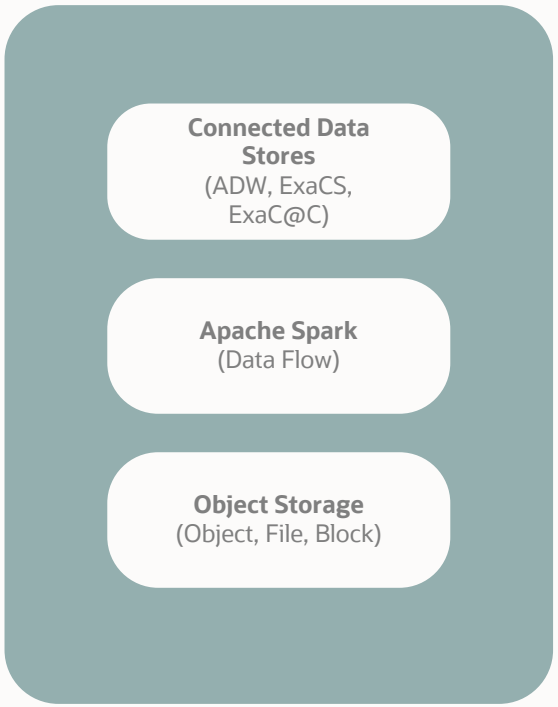
Cloud-native data lake & data warehouse

Transform, curate and analyze

Object storage-based persistence

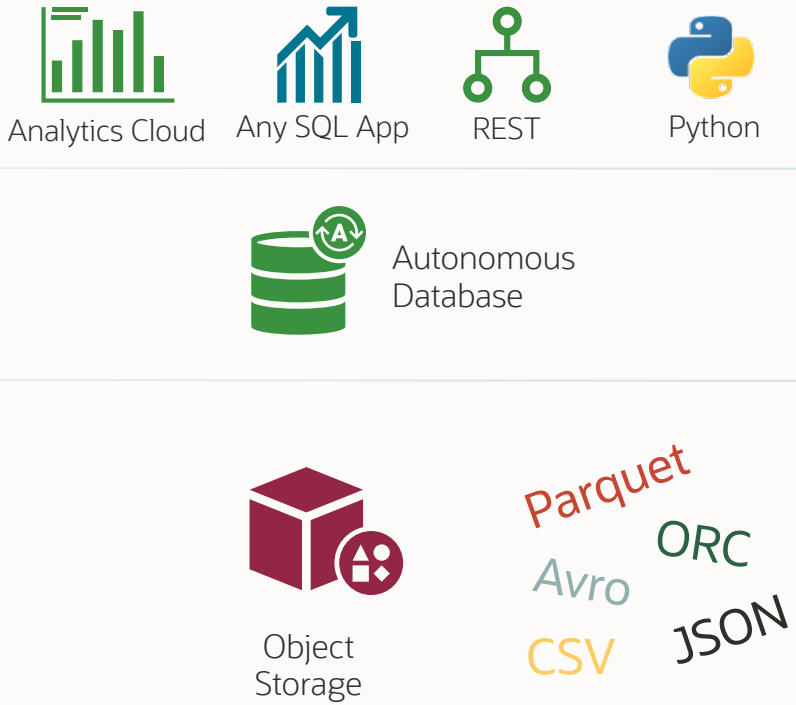
Spark to process the data

Integrated access and query with data warehouse



Analyzing information across data sources

Powerful SQL access to complex data



Easily load data into ADB from object storage

- Oracle Object Storage, Amazon S3, Azure

Optimized access to data lake objects

- External tables with optimized drivers for Parquet, ORC and Avro files
- Automatically derive schemas
- Highly flexible text processing
- Access both simple and complex data types
- Powerful SQL extensions for JSON and XML

Why a powerful SQL language is so important

Simple SQL access

Applications challenged to access complex sources

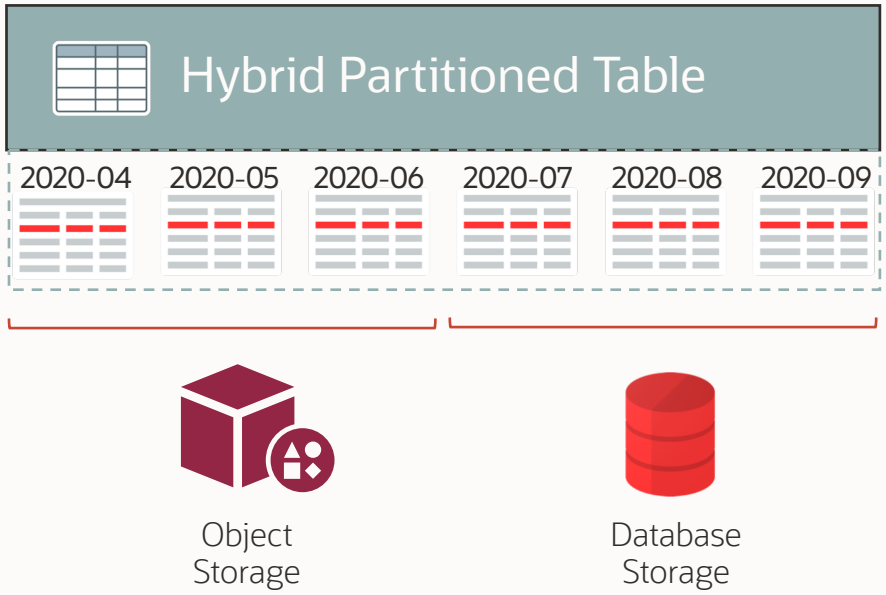
Autonomous Database

External table encapsulate complexity


```
{"data":{"stations":[{"lat":40.76727216,"eightd_station_services":[],"external_id":"66db237e-0aca-11e7-
82f6-3863bb44ef7c","electric_bike_surcharge_waiver":false,"rental_methods":
["KEY","CREDITCARD"],"short_name":"6926.01","has_kiosk":true,"rental_url":"http://app.citibikenyc.com/S6L
r/IBV092JufD?station_id=72","eightd_has_key_dispenser":false,"legacy_id":"72","name":"W 52 St & 11
Ave","region_id":"71","station_type":"classic","capacity":0,"station_id":"72","lon":-73.99392888},
{"lat":40.71911552,"eightd_station_services":[],"external_id":"66db269c-0aca-11e7-82f6-
3863bb44ef7c","electric_bike_surcharge_waiver":false,"rental_methods":
["KEY","CREDITCARD"],"short_name":"5430.08","has_kiosk":true,"rental_url":"http://app.citibikenyc.com/S6L
r/IBV092JufD?station_id=79","eightd_has_key_dispenser":false,"legacy_id":"79","name":"Franklin St & W
Broadway","region_id":"71","station_type":"classic","capacity":33,"station_id":"79","lon":-74.00666661},
{"lat":40.71117416,"eightd_station_services":[],"external_id":"66db277a-0aca-11e7-82f6-
3863bb44ef7c","electric_bike_surcharge_waiver":false,"rental_methods":
["KEY","CREDITCARD"],"short_name":"5167.06","has_kiosk":true,"rental_url":"http://app.citibikenyc.com/S6L
r/IBV092JufD?station_id=82","eightd_has_key_dispenser":false,"legacy_id":"82","name":"St James Pl & Pearl
St","region_id":"71","station_type":"classic","capacity":27,"station_id":"82","lon":-74.00016545},
{"lat":40.68382604,"eightd_station_services":[],"external_id":"66db281e-0aca-11e7-82f6-
3863bb44ef7c","electric_bike_surcharge_waiver":false,"rental_methods":
["KEY","CREDITCARD"],"short_name":"4354.07","has_kiosk":true,"rental_url":"http://app.citibikenyc.com/S6L
r/IBV092JufD?station_id=83","eightd_has_key_dispenser":false,"legacy_id":"83","name":"Atlantic Ave & Fort
Greene Pl","region_id":"71","station_type":"classic","capacity":62,"station_id":"83","lon":-73.97632328},
{"lat":40.74177603,"eightd_station_services":[],"external_id":"66db28b5-0aca-11e7-82f6-
3863bb44ef7c","electric_bike_surcharge_waiver":false,"rental_methods":
["KEY","CREDITCARD"],"short_name":"6148.02","has_kiosk":true,"rental_url":"http://app.citibikenyc.com/S6L
r/IBV092JufD?station_id=116","eightd_has_key_dispenser":false,"legacy_id":"116","name":"W 17 St & 8
Ave","region_id":"71","station_type":"classic","capacity":50,"station_id":"116","lon":-74.00149746},
```


- Parquet, Avro or ORC with columns containing arrays, structs, and more
- Complex JSON files

Simplified data management
Hybrid Partitioned Tables



Single table captures both internal and external data

- Share data with data lake object storage
- Example: Database storage for frequently accessed data, Parquet for cold data

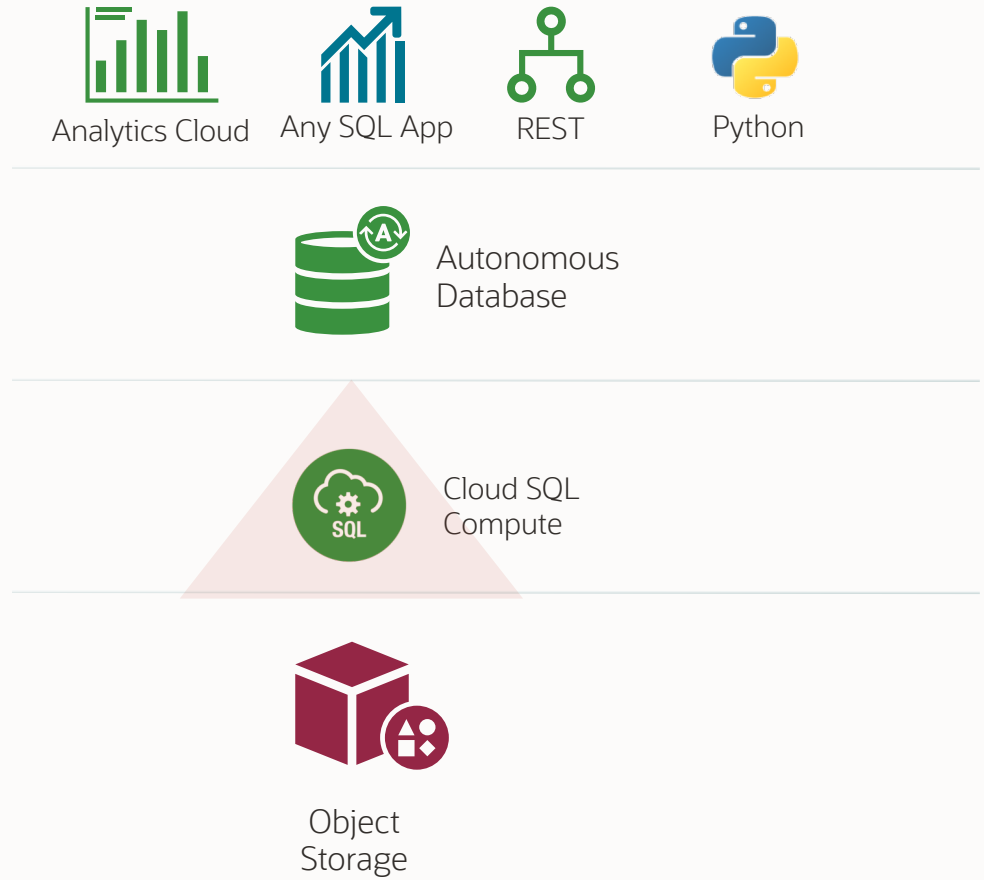
Efficiently access data across storage tiers

- Partition pruning
- Enables order-of-magnitudes faster query performance

Cost-effective solution

Add Cloud SQL scale-out compute tier

Object Store queries just get better



Scale out queries against Object Stores using Cloud SQL compute

- Offload query processing to separate, fully managed compute tier
- Filters, processes and aggregates data

Automatic and transparent

- ADB uses Cloud SQL only when needed
- Enabled with ADB auto scaling

Short term roadmap

OCI Data Catalog Vision

Self-service Data Discovery and Governance Solution

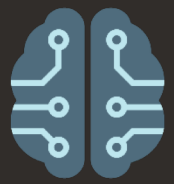
Search & Discovery

- Metadata harvesting from on-prem and cloud systems
- Semantic search, data profiling, lineage and impact analysis, data relationships

Metadata Curation

- Enterprise Business Glossary, Approval Workflows
- Tagging, User annotations, social collaboration, ratings/comments, **associations and links**

Data Intelligence



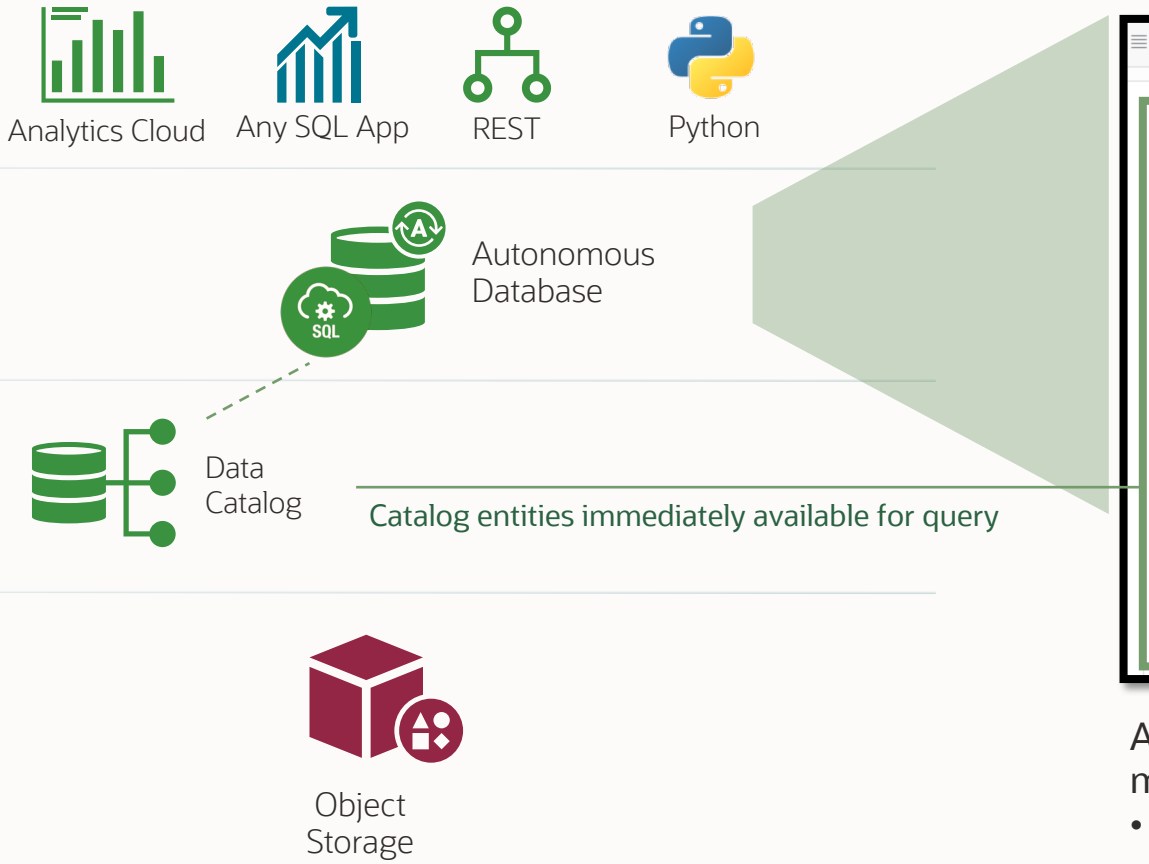
- AI/ML based Recommendations
- Auto-tagging, Auto-discovery, Auto-classification, Auto-association, Data Similarity

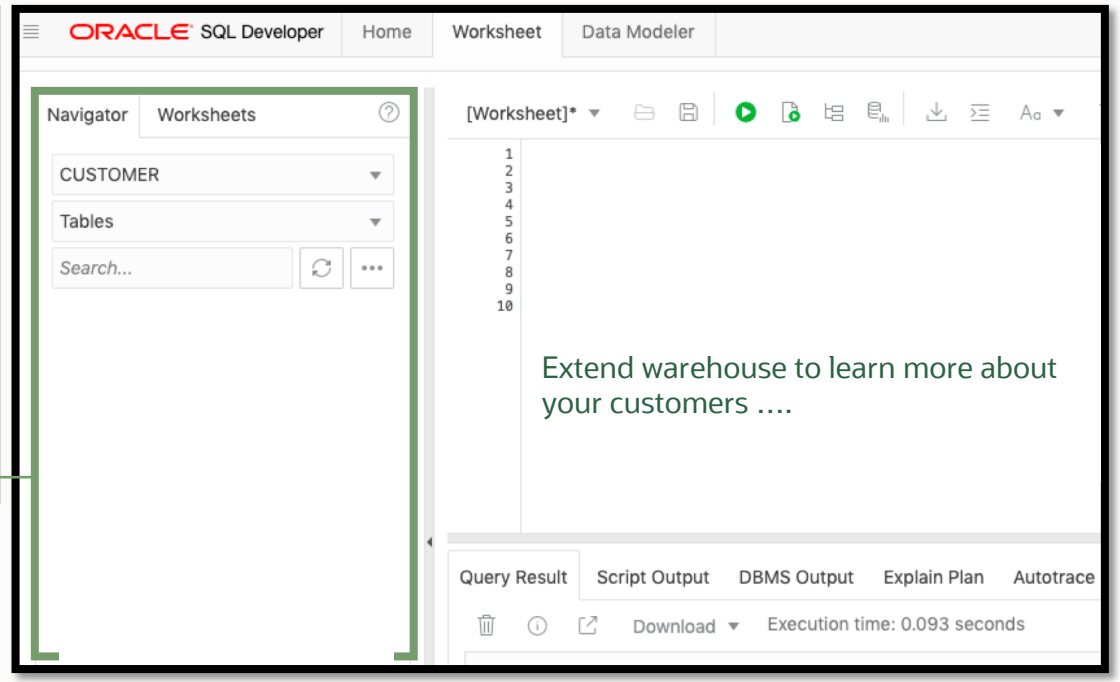
Enterprise Class

- OCI Native, REST APIs
- Hive Metastore for OCI Obj Store, ADW etc.
- Value added Integration with other OCI services

Integrate Data Catalog Assets with Autonomous Database

Blend Data Lake with Autonomous Database – Short term roadmap



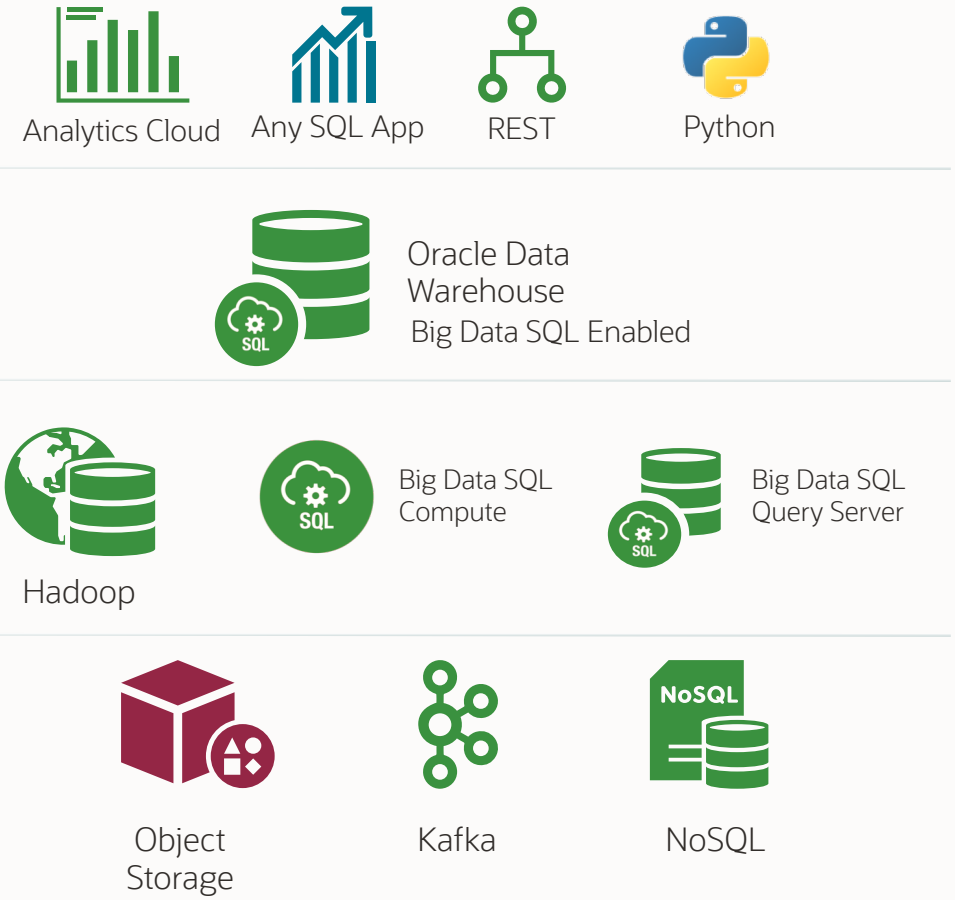


Autonomous Database automatically uses shared platform metadata

- Schemas extended with Data Catalog object storage assets
- External tables defined over sources
- Join data in warehouse and data lake

On-premise - Analyzing information across data sources

Oracle Big Data SQL



Easily query across sources

- One Oracle SQL query correlates information from multiple data stores
- Supports HDFS, Hive, Object Stores, Kafka and NoSQL
- No change to Oracle-based applications

Fast, scale-out processing using Smart Scan

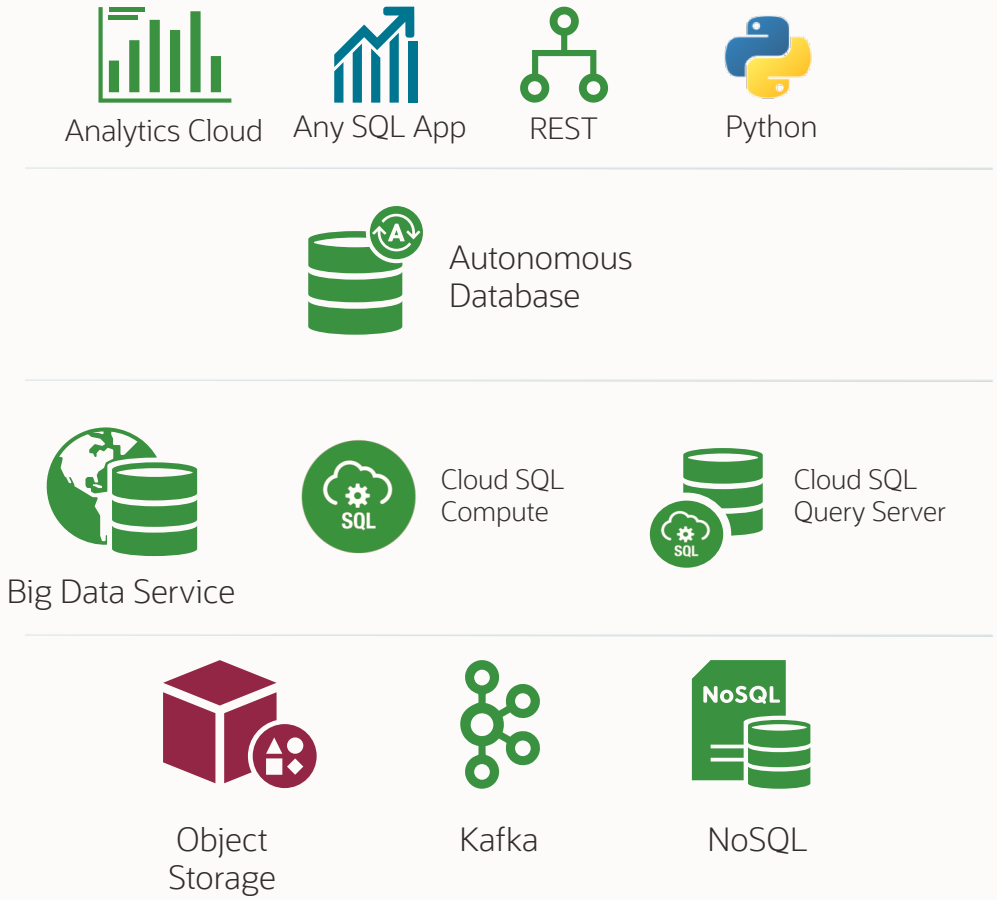
Oracle advanced security over all data

Optionally use Oracle query engine for ”SQL on Hadoop”

- Leverages existing Hive metadata and security

Transitioning workloads to cloud

Oracle Big Data Service with Oracle Cloud SQL



Oracle Big Data Service

- Comprehensive data lake platform
- Based on Cloudera Enterprise 6.3 - enabling workload portability
- Perfect for dev, test
- Integrated into Oracle Cloud

Easily deploy secure, HA clusters

Right-size solution

- Elastic - scale from small VM-based clusters to high performance bare metal

Cloud SQL Query Server delivers “SQL on Hadoop”

Summary

- Autonomous Database allows you to focus on gaining insights from all your data – leave the infrastructure to us
- Take advantage of cloud scale and economics to explore and understand all your data
- Get started today!

ORACLE