AI-school

Data Science - Big Data Analytics

marcusRB

Presentación Programa

AI-school
Data Science - Big Data Analytics

marcusRB

Data Pipeline

AI-school
Data Science - Big Data Analytics

Tecnología: BI & Data Warehouse

Big Data vs Data Warehouse

Big Data vs Data Warehouse

Factor	DWH	Big Data
Fuente de datos	Sistemas corporativos y transaccionales	Fuentes no tradicionales, como sensores, logs, vídeos, etc.
Volumen	Hasta 100 Terabytes	A partir de 100 Terabytes
Velocidad	Batch o procesos que no requieren respuesta inmediata	Respuesta inmediata
Variedad	Principalmente estructurada	De todo tipo
Veracidad	Organizada y de calidad	De calidad variable
Valor	BI y analítica	Machine learning, deep learning y anteriores
Objetivo	Toma de decisiones	Múltiple, pero destaca la creación de productos y servicios de datos

Big Data vs Data Warehouse

Comparativas On-Premise vs Cloud

Big Data en Cloud

On-Premises

AWS

GCP

Azure

1 © 2019 Gartner, Inc. and/or its affiliates. All rights reserved. Gartner is a registered trademark of Gartner, Inc. and its affiliates.

Data Warehouse en AWS

Data Warehouse en Azure

Data Warehouse en GCP

Traditional Data Warehouse	Complex ETL	Restricted to only a few users	Optimized for legacy BI	Optimized for batch data	Needs continuous patching	Needs an army of DBAs to operate
BigQuery	Automate data delivery	Make insights accessible	Build the foundation for AI	Tee up real-time insights	Protect business data	Simplify data operations

Big Data vs Data Warehouse

databricks

Big Data vs Data Warehouse

Big Data vs Data Warehouse

Big Data vs Data Warehouse

Comparativas On-Premise vs Cloud

This is one of the first electronic computers ever made. The first computer, created by Charles Babbage in 1822.

Comparativas On-Premise vs Cloud

the Manchester baby nicknamed Baby, was the world's first stored program computer. It was built at the Victoria University of Manchester in England on June 1948.

Comparativas On-Premise vs Cloud

1980s: Simple Desktop-like server machine

Traditionally, the bare metals were simple desktop which were placed in a small room inside the organization. It had the configuration of 256Mhz cpu and 2GB of disk.

1990s: Rack Mountable Web Server

Gradually, when there were increase in demand for web hosting services, rack server replaced the desktop-like machine. It had Intel P2 Xeon 450Mhz, 256Mb RAM, and a 24X CDROM player.

1998: First Google server was introduced with 256MB RAM and dual 200 Mhz CPUs.

2000s: Blade servers

2001-2002 saw the introduction of blade servers, a palm sized machine. In 2008, virtualization became popular and server were based on distributed computing technology. And with the expansion of virtualization technique, Cloud computing had its emergence.

Comparativas On-Premise vs Cloud

Comparativas On-Premise vs Cloud

Comparativas On-Premise vs Cloud

marcusRB

Comparativas On-Premise vs Cloud

Comparativas On-Premise vs Cloud

AI-school
Data Science - Big Data Analytics

Comparativas On-Premise vs Cloud

AI-school
Data Science - Big Data Analytics

Comparativas On-Premise vs Cloud

Comparativas On-Premise vs Cloud

https://www.youtube.com/watch?v=XZmGGAbHqa0

Comparativas On-Premise vs Cloud

AI-school
Data Science - Big Data Analytics

Comparativas On-Premise vs Cloud

marcusRB

On-Premises	IaaS Infrastructure as a Service	PaaS Platform as a Service	SaaS Software as a Service
Applications	Applications	Applications	Applications
Data	Data	Data	Data
Runtime	Runtime	Runtime	Runtime
Middleware	Middleware	Middleware	Middleware
O/S	O/S	O/S	O/S
Virtualization	Virtualization	Virtualization	Virtualization
Servers	Servers	Servers	Servers
Storage	Storage	Storage	Storage
Networking	Networking	Networking	Networking

You Manage | Other Manages

AI-school

Comparativas On-Premise vs Cloud

Comparativas On-Premise vs Cloud

Comparativas On-Premise vs Cloud

AWS

The AWS Platform
Enterprise Applications
Virtual Desktops
Collaboration and Sharing
Platform Services
Analytics
Hadoop
Real-time Streaming Data
Data Warehouse
Data Pipelines
App Services
Queuing & Notifications
Workflow
App streaming
Transcoding
Email
Search
Deployment & Management
One-click web app deployment
Dev/ops resource management
Resource Templates
Mobile Services
Identity
Sync
Mobile Analytics
Push Notifications
Administration & Security
Identity Management
Access Control
Usage Auditing
Key Storage
Monitoring and Logs
Core Services
Compute
(VMs, Auto-scaling and Load Balancing)
Storage
(Object, Block and Archival)
CDN
Databases
(Relational, NoSQL, Caching)
Networking
(VPC, DX, DNS)
Infrastructure
Regions
Availability Zones
Points of Presence

Microsoft Azure

GCP

marcusRB

Google Cloud

DEVELOPER'S CHEAT SHEET

v2019.7.17

Created by the Google Developer Relations Team

Maintained at https://github.com/gregsramblings/google-cloud-4-words

Feedback? @gregsramblings

COMPUTE

App Engine	Managed app platform
Cloud Functions	Event-driven serverless functions
Cloud Run	Serverless for containerized applications
Compute Engine	VMs, GPUs, TPUs, Disks
Kubernetes Engine (GKE)	Managed Kubernetes/containers
Anthos	Enterprise hybrid/multi-cloud platform

STORAGE

Cloud Storage	Object storage and serving
Nearline	Archival occasional access storage
Coldline	Archival rare access storage
Persistent Disk	VM-attached disks
Cloud Filestore	Managed NFS server

DATABASES

Cloud Bigtable	Petabyte-scale, low-latency, non-relational
Cloud Datastore	Horizontally scalable document DB
Cloud Firestore	Strongly-consistent serverless document DB
Cloud Memorystore	Managed Redis
Cloud Spanner	Horizontally scalable relational DB
Cloud SQL	Managed MySQL and PostgreSQL

DATA AND ANALYTICS

BigQuery	Data warehouse/analytics
BigQuery BI Engine	In-memory analytics engine
BigQuery ML	BigQuery model training/serving
Cloud Composer	Managed workflow orchestration service
Cloud Data Fusion	Graphically manage data pipelines
Cloud Dataflow	Stream/batch data processing
Cloud Datalab	Managed Jupyter notebook
Cloud Dataprep	Visual data wrangling
Cloud Dataproc	Managed Spark and Hadoop
Cloud Pub/Sub	Global real-time messaging
Data Catalog	Metadata management service
Data Studio	Collaborative data exploration/dashboarding
Genomics	Managed genomics platform

AI/ML

AI Hub	Hosted AI component sharing
AI Platform	Managed platform for ML
AI Platform Data Labeling	Data labeling by humans
AI Platform Deep Learning VMs	Preconfigured VMs for deep learning
AI Platform Notebooks	Managed JupyterLab notebook instances
AI Platform Training	Parallel and distributed training
AI Platform Predictions	Autoscaled model serving
AutoML Natural Language	Custom text models
AutoML Tables	Custom structured data models
AutoML Translation	Custom domain-specific translation
AutoML Video Intelligence	Custom video annotation models
AutoML Vision	Custom image models
Cloud AI Building Blocks	Hosted AI component repository
Cloud Natural Language API	Text parsing and analysis
Cloud Speech-To-Text API	Convert audio to text
Cloud Talent Solutions API	Job search with ML
Cloud Text-To-Speech API	Convert text to audio
Cloud Translation API	Language detection and translation
Cloud Video Intelligence API	Scene-level video annotation
Cloud Vision API	Image recognition and classification
Cloud TPU	Hardware acceleration for ML
Dialogflow Enterprise Edition	Create conversational interfaces
Document Understanding AI	Analyze, classify, search documents
Recommendations AI	Create custom recommendations
Vision Product Search	Visual search for products

NETWORKING

Carrier Peering	Peer through a carrier
Direct Peering	Peer with GCP
Dedicated Interconnect	Dedicated private network connection
Partner Interconnect	Connect on-prem network to VPC
Cloud Armor	DDoS protection and WAF
Cloud CDN	Content delivery network
Cloud DNS	Programmable DNS serving
Cloud Load Balancing	Multi-region load distribution
Cloud NAT	Network address translation service
IPsec VPN	Virtual private network connection
Network Service Tiers	Price vs performance tiering
Network Telemetry	Network telemetry service
Traffic Director	Service mesh traffic management
Google Cloud Service Mesh	Service-aware network management
Virtual Private Cloud	Software defined networking

INTERNET OF THINGS (IoT)

Cloud IoT Core	Device management and ingest data

IDENTITY AND SECURITY

Access Transparency	Audit cloud provider access
Binary Authorization	Kubernetes deploy-time security
Cloud Audit Logs	Audit trails for GCP
Cloud Data Loss Prevention API	Classify and redact sensitive data
Cloud HSM	Hardware security module service
Cloud IAM	Resource access control
Cloud Identity	Manage users, devices & apps
Cloud Identity-Aware Proxy	Identity-based app sign in
Cloud Key Management Service	Hosted key management service
Cloud Resource Manager	Cloud project metadata management
Cloud Security Scanner	App engine security scanner
Cloud Security Command Center	Asset inventory, discovery, search, management
Context-aware Access	End-user attribute-based access control
Event Threat Detection	Scans for suspicious activity
Security Key Enforcement	Two-step key verification
Shielded VMs	Hardened VMs
Titan Security Key	Two-factor authentication (2FA) device
VPC Service Controls	VPC constrain data

MANAGEMENT TOOLS

Cloud APIs	APIs for cloud services
Cloud Billing	Billing and cost management tools
Cloud Billing API	Programmatically manage GCP billing
Cloud Console	Web-based management console
Cloud Deployment Manager	Templated infrastructure deployment
Cloud Mobile App	iOS/Android GCP manager app
Cloud Shell	Browser-based terminal/CLI
Stackdriver Debugger	Live production debugging
Stackdriver Error Reporting	App error reporting
Stackdriver Logging	Centralized logging
Stackdriver Monitoring	Infrastructure and application monitoring
Stackdriver Profiler	CPU and heap profiling
Stackdriver Transparent SLIs	Monitor GCP services
Stackdriver Trace	App performance insights

DEVELOPER TOOLS

Cloud SDK	CLI for GCP
Cloud Build	Continuous integration/delivery platform
Cloud Code	Cloud native IDE extensions
Cloud Source Repositories	Hosted private git repos
Cloud Scheduler	Managed cron job service
Cloud Tasks	Asynchronous task execution
Cloud Tools for IntelliJ	IntelliJ GCP tool
Cloud Tools for PowerShell	PowerShell GCP tools
Cloud Tools for Visual Studio	Visual Studio GCP tools
Cloud Tools for Eclipse	Eclipse GCP tools
Container Registry	Private container registry/storage
Gradle App Engine Plugin	Gradle App Engine plugin
Maven App Engine Plugin	Maven App Engine plugin

MIGRATION TO GCP

Cloud Data Transfer	Data migration tools/CLI
Google Transfer Appliance	Rentable data transport box
Cloud Storage Transfer Service	Cloud to cloud transfers
BigQuery Data Transfer Service	Bulk import analytics data
Migrate from Amazon Redshift	Migrate from Redshift to BigQuery
Migrate from Teradata	Migrate from Teradata to BigQuery
Migrate for Anthos	Migrate VMs to GKE containers
Migrate for Compute Engine	Compute Engine migration tools
VM Migration	VM migration tools

API PLATFORM AND ECOSYSTEMS

API Analytics	API metrics
API Monetization	Monetize APIs
Apigee API Platform	Develop, secure, monitor APIs
Apigee Sense	API protection from attacks
Apigee Hybrid	Manage hybrid/multi-cloud API environments
Cloud Endpoints	Cloud API gateway
Cloud Healthcare API	Healthcare system GCP interoperability
Developer Portal	API managment portal
GCP Marketplace	Partner & open source marketplace

GOOGLE MAPS PLATFORM

Directions API	Get directions between locations
Distance Matrix API	Calculate travel times
Geocoding API	Convert address to/from coordinates
Geolocation API	Derive location without GPS
Maps Embed API	Web embedded maps
Maps JavaScript API	Dynamic web maps
Maps SDK for Android	Maps SDK for Android
Maps SDK for iOS	Maps SDK for iOS
Maps Static API	Web static maps
Maps Unity SDK	Unity SDK for games
Maps URLs	URL scheme for maps
Places API	Metadata about places (REST)
Places Library, Maps JS API	Metadata about places (JavaScript)
Places SDK for Android	Places SDK for Android
Places SDK for iOS	Places SDK for iOS
Roads API	Metadata about roads
Street View Static API	Static street view images
Street View Service	Interactive street view images
Time Zone API	Convert coordinates to timezone

G SUITE PLATFORM

App Maker	Assistive app building
Apps Script	Extend and automate everything
Editor Add-ons	Extend Docs, Sheets, Slides
Gmail Add-ons	Contextual apps in Gmail
Hangouts Chat Bots	Conversational bots in chat
Calendar API	Create and manage calendars
Classroom API	Provision and manage classrooms
Docs API	Create and edit documents
Drive API	Read and write files
Gmail API	Enhance Gmail
Sheets API	Read and write spreadsheets
Slides API	Create and edit presentations
Drive Picker	Drive file selection widget
Cloud Search	Unified search for enterprise
Admin SDK	Manage G Suite resources
Email Markup	Interactive email using schema.org
G Suite Marketplace	Storefront for integrated applications
Other G Suite APIs/SDKs	Contacts, Google+, Tasks, Vault...

MOBILE (FIREBASE)

Cloud Firestore	Document store and sync
Cloud Functions for Firebase	Event-driven serverless applications
Cloud Storage for Firebase	Object storage and serving
Crashlytics	Crash reporting and analytics
Firebase A/B Testing	Create A/B test experiments
Firebase App Indexing	App/Google search integration
Firebase Authentication	Drop-in authentication
Firebase Cloud Messaging	Send device notifications
Firebase Dynamic Links	Link to app content
Firebase Hosting	Web hosting with CDN/SSL
Firebase In-App Messaging	Send in-app contextual messages
Firebase Performance Monitoring	App performance monitoring
Firebase Predictions	Predict user targeting
Firebase Realtime Database	Real-time data synchronization
Firebase Remote Config	Remotely configure installed apps
Firebase Test Lab	Mobile testing device farm
Google Analytics for Firebase	Mobile app analytics
ML Kit for Firebase	ML APIs for mobile

GCP FOUNDATIONAL OPEN SOURCE PROJECTS

Apache Beam	Batch/streaming data processing
gRPC	RPC framework
gVisor	Secure container runtime
Istio	Connect and secure services
Knative	Serverless framework for Kubernetes
Kubeflow	ML Tool kit for Kubernetes
Kubernetes	Management of containerized applications
OpenCensus	Cloud native observability framework
TensorFlow	ML framework

ADDITIONAL RESOURCES

Google Cloud Home Page	cloud.google.com
Google Cloud Blog	cloud.google.com/blog
GCP Medium Publication	medium.com/google-cloud
Apigee Blog	apigee.com/about/blog
Firebase Blog	firebase.googleblog.com
G Suite Developers Blog	gsuite-developers.googleblog.com
Google Cloud Certifications	cloud.google.com/certification
Google Cloud System Status	status.cloud.google.com
Google Cloud Training	cloud.google.com/training
Google Developers Blog	developers.googleblog.com
Google Maps Platform Blog	mapsplatform.googleblog.com
Google Open Source Blog	opensource.googleblog.com
Google Security Blog	security.googleblog.com
Kaggle Home Page	www.kaggle.com
Kubernetes Blog	kubernetes.io/blog
Regions and Network Map	cloud.google.com/about/locations

AI-school

Entornos: Virtual Machine

KVM
Xen
oVirt
VAGRANT

Entornos: containers

Machine Virtualization

Containers

Entornos: containers

Entornos: serverless

Entornos: example serverless

AI-school
Data Science - Big Data Analytics

marcusRB

¿Dudas - Preguntas?

AI-school
Data Science - Big Data Analytics