



OPTIMUS PRIME

Transformando y evolucionando una
plataforma de datos Digital

Antonio Galindo

Data Fabric

OLTP

OLAP

Big Data

Data lake

Data exchanges

Data mesh

Data contracts

Lake House

Reverse ETL

Datawarehouse

ETL

ELT

Near Real Time

Batch

Pipeline

Estructurado

NoSQL

Data rivers

Semi Estructurado

No Estructurado

Eventos

Data marts

Grafos

Real Time

Documentos

Data domains

State of Data Engineering 2022 map

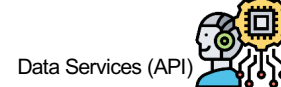
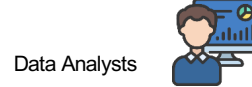
Presented by  lakeFS



<h3>Ingest SaaS</h3> <ul style="list-style-type: none">StitchAirbyteFivetranSegmentRiveryAzure Data FactoryDataddodatacoralMATILLIONSNOWFLOW	<h3>Object Storage</h3> <ul style="list-style-type: none">amazon S3Microsoft Azure Blob StorageGoogle Cloud StorageORACLE CLOUDIBM Cloud Object StorageAlibaba CloudhadoopMINIOzadaraSEAGATEcloudfilesfilebasewasabiCrowdStorageSwiftStackcephLYVEAzure Data Lake Storage Gen2PURESTORAGE	<h3>Metastore</h3> <ul style="list-style-type: none">hiveCloud DataprocAzure PurviewAmazon EMRCLUSTERFAIRCLUSTERFAIRdatabricks Unity <h3>Git for Data</h3> <ul style="list-style-type: none">lakeFSProject Nessie <h3>Open Table Formats</h3> <ul style="list-style-type: none">ONEHOUSEhudidatabricksDELTA LAKETabularICEBERGorc	<h3>Compute</h3> <ul style="list-style-type: none">databrickssparkCloudwickRAYamazon EMRAmazon SageMakerakkatrinoDASKASCEND.IOmapboxbodaiCLUSTERFAIR <h3>Analytics Engine</h3> <ul style="list-style-type: none">snowflakeamazon ATHENAamazon REDSHIFTGoogle Big QuerySynapsedatabricksclouderadremiodruidpinotstar+treeClickHouseStarburstpentahoQubaleimplyROCKSETFIREBOLT	<h3>Orchestration</h3> <ul style="list-style-type: none">AirflowFlyteAzure Data FactoryPREFECTLuigidagsterASTRONJMER <h3>Observability</h3> <ul style="list-style-type: none">MONTE CARLOlightupDatafoldKENSUgreat_expectationsDatabandHoloCleanBigeyeunravelWHYLABSgriffinMetaplaneawslabs/deequredataSODAelementaryacceldatatimeseekerAI	<h3>MLOps End-to-End</h3> <ul style="list-style-type: none">colabOctoMLABACUS.AIMETAFLOWDidatVertacnvrg.iomlflowsnorkelSELDONHugging FaceDataRobotW&BHOPSWORKSKubeflowGoogle Data StudioCLEAR MLValohaidominodata ikuAmazon SageMakerFLOYDHUBiguaziohydrosphere.ioZenMLMichelangeliKedroQwakcometH2O.ai <h3>Data Centric AI/ML</h3> <ul style="list-style-type: none">DAGsHubDVCactiveloopPachydermGRAVITILIGHTLY <h3>Feature Stores</h3> <ul style="list-style-type: none">RASGOTACTONFEASTKASKADAMOLECULAHOPSWORKSscribbleDataredis <h3>ML Observability</h3> <ul style="list-style-type: none">deepchecksmonaSuperwisefiddlerApresarizeWHYLABSArthurgalileoROBUST INTELLIGENCEGANTRYtrueraaporia <h3>Notebooks</h3> <ul style="list-style-type: none">DeepnoteHEXdataformdbtjupytercountnoteabledatabricksQuerybook	<h3>Discovery & Governance</h3> <ul style="list-style-type: none">MARQUEZboomiQwakCollibraBigIDmagdaApache AtlasOpen MetadatackanOKERAAlationAmundsendata.worldIBM WatsonIMMUTADataHubatlanprivaceraPlatformAcryl DataSELECT STAR
---	---	--	---	---	---	---

Arquitectura de datos

Data Consumption



Data Processing & Analysis

Reporting Platforms

Analytics Engines

Exploration and Discovery

Data Science Workbench

Specific Industry Solutions

Marketplace

Master Data Sets

Trusted Data Sets

Semantic Layer

Data Catalog

Exploratory Data Sets

MDM & Data Quality

Data Catalog

Data Hub



Analytics



Landing Hub



Data Lake



Data Ingestion

Replication

ELT/ETL

Streaming

Data Sources

1st Party Data

2nd Party Data

3rd Party Data



Governance



Security & Compliance



Infrastructure



¿Qué es **OPTIMUS PRIME** y por qué el nombre?

Todos tienen una arquitectura
de datos



Todos hacen BI o Analítica
Descriptiva

Todos tienen herramientas de
ingesta



Pocos automatizan

¿Necesidades o premisas para el diseños de la plataforma de datos?



OnPremise / Cloud / Multi-Cloud



Escalabilidad



Casos de uso



Multinegocios



Agilidad/Crecimiento Exponencial



Monetizar

¿Consideraciones al plantear la estrategia de la plataforma de datos?



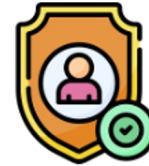
Data Fabric



Democratización de datos



Gobierno de Datos



Datos como un activo



Enfoque en el cliente



Data products



¿Cómo fue evolucionando tu plataforma de datos?



Disminución en la complejidad de la arquitectura



Reto con las herramientas de nicho



Aprovechamiento capacidad Open Source



Automatización de desarrollo de pipelines, despliegues (IaC, DevOps)

¿Cuáles son los principales beneficios hacia el negocio y al mercado?

Disminuir el time to insight



Aprovechar el talento



Estandarización y agilidad en procesos



Incrementar la calidad de datos

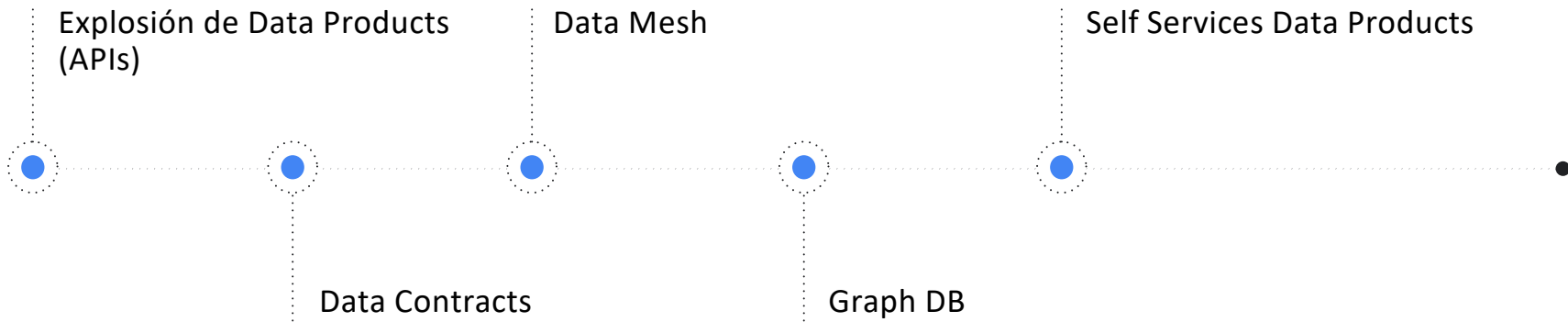


Automatizar



Mejoras en la Seguridad de la Información





Muchas gracias!

