

**CLOUDERA**

  
**Mastersystem**  
ICT Solutions

**PROPOSAL TEKNIS  
PROJECT REFRESHMENT, RENEWAL DAN EXPANSION  
PLATFORM BIG DATA**

**PT. BANK RAKYAT INDONESIA (PERSERO), TBK**

**MASTERSYSTEM INFOTAMA  
DENGAN  
CLOUDERA**



**9 MEI 2023**



**Silahkan Hubungi:**

**Jimmy Revalino**

Mastersystem Infotama

Sudirman 7.8 Tower I Lantai 25

Jl. Jend. Sudirman kav.7-8 Jakarta 10220

Phone: +6221 39731111 ext 8837

Mobile: +62 811 858 471

eMail: jimmy.revalino@mastersystem.co.id

<http://www.mastersystem.co.id>

**Adissa Alifia**

Mastersystem Infotama

Sudirman 7.8 Tower I Lantai 25

Jl. Jend. Sudirman kav.7-8 Jakarta 10220

Phone: +6221 39731111

Mobile: +62 813 8206 2797

eMail: adissa.alifia@mastersystem.co.id

<http://www.mastersystem.co.id>

***Disclaimer***

Dokumen ini digolongkan sebagai dokumen rahasia bisnis dan tidak diijinkan untuk di perbanyak sebagian ataupun seluruh isinya atau di berikan kepada siapapun tanpa ijin tertulis dari PT Mastersystem Infotama

Segala hal yang berhubungan dengan sifat dan gambaran teknis, prosedur dan rancangan, termasuk namun tidak terbatas, semua hal yang berhubungan dengan metodologi implementasi, teknik perencanaan dan pengaturan proyek, pengujian dan perancangan sistem digolongkan sebagai informasi yang bersifat rahasia dan tidak boleh diteruskan atau dibuka kepada pihak manapun tanpa ijin tertulis dari PT Mastersystem Infotama.

Segala perangkat lunak yang ada dalam dokumen ini digunakan dibawah ijin hak cipta dan hanya digunakan atau diperbanyak menurut aturan yang mengikat kepada ijin hak cipta perangkat lunak tersebut.

Keterangan yang tercantum dalam dokumen ini diberikan hanya untuk kepentingan informasi, dan segala perubahan mungkin saja terjadi, dan segala perubahan yang mungkin terjadi merupakan bentuk komitmen dari PT Mastersystem Infotama.

Hak Cipta © PT Mastersystem Infotama, 2023.

## DAFTAR ISI

<b>1 PENDAHULUAN.....</b>	<b>5</b>
1.1 <i>LATAR BELAKANG .....</i>	5
1.2 <i>TUJUAN .....</i>	5
<b>2 PROFIL PERUSAHAAN.....</b>	<b>6</b>
2.1 <i>MASTERSYSTEM INFOTAMA .....</i>	6
2.1.1 <i>Tentang Mastersystem.....</i>	6
2.1.2 <i>Kerjasama dengan Principal.....</i>	7
2.2 <i>CLOUDERA.....</i>	7
2.2.1 <i>Tentang Cloudera.....</i>	7
2.2.2 <i>Klien Cloudera.....</i>	9
2.2.3 <i>Operasi Global Cloudera.....</i>	9
<b>3 RUANG LINGKUP.....</b>	<b>10</b>
3.1 <i>MODEL LISENSI.....</i>	10
3.2 <i>RUANG LINGKUP YANG DIAJUKAN.....</i>	15
3.2.1 <i>Teknis .....</i>	15
3.2.2 <i>Aplikasi.....</i>	15
3.2.3 <i>Hardware.....</i>	15
3.2.4 <i>Implementasi.....</i>	15
3.2.5 <i>Transfer Knowledge .....</i>	18
3.3 <i>WARRANTY DAN SUPPORT.....</i>	19
3.4 <i>GARANSI/WARRANTY.....</i>	19
3.4.1 <i>Post Go-Live Support .....</i>	19
3.4.2 <i>Preventive Maintenance.....</i>	19
3.4.3 <i>Corrective Maintenance .....</i>	19
3.4.4 <i>Service Level Agreement (SLA).....</i>	20
3.4.5 <i>Support lainnya.....</i>	20
3.5 <i>DOKUMENTASI.....</i>	21
<b>4 METODOLOGI .....</b>	<b>22</b>
4.1 <i>MANAJEMEN PROYEK.....</i>	22
4.1.1 <i>Project Communication Plan .....</i>	22
4.1.2 <i>Prosedur Eskalasi .....</i>	23
4.1.3 <i>Bank BRI Steering Committee.....</i>	23
4.1.4 <i>Proses Penanganan Intellectual Property.....</i>	24
4.2 <i>MANAJEMEN RESIKO PROYEK.....</i>	24
4.2.1 <i>Risk Management .....</i>	24
4.2.2 <i>Sumber Utama Resiko .....</i>	24
4.2.3 <i>Tindakan Mitigasi yang Umum.....</i>	25
4.2.4 <i>Proses Resiko yang digunakan dalam Proyek ini .....</i>	25
4.3 <i>INFRASTRUKTUR .....</i>	25
4.4 <i>QUALITY ASSURANCE.....</i>	25
4.4.1 <i>Kebutuhan Umum dan Teknis Quality Assurance .....</i>	25
4.4.2 <i>Tindakan Mitigasi yang Umum.....</i>	26
4.5 <i>PROJECT CHANGE CONTROL PLAN .....</i>	27

---

4.5.1	<i>Penjelasan Awal</i> .....	27
4.5.2	<i>Tanggung Jawab Change Control</i> .....	27
4.5.3	<i>Change Request Process Flow</i> .....	28
4.5.4	<i>Rangkuman</i> .....	30
4.5.5	<i>Change Control History Register</i> .....	31
<b>5</b>	<b>ASUMSI DAN OUT OF SCOPE</b> .....	<b>32</b>
5.1	<i>OUT OF SCOPE</i> .....	32
<b>6</b>	<b>JADWAL PELAKSANAAN</b> .....	<b>33</b>
6.1	<i>IMPLEMENTASI DATALAKE DC</i> .....	33
6.2	<i>IMPLEMENTASI HBASE DC</i> .....	36
<b>LAMPIRAN 1: SPESIFIKASI TEKNIK – KEBUTUHAN FUNGSIONAL</b> .....		<b>38</b>
<b>LAMPIRAN 2: SPESIFIKASI TEKNIK – KEBUTUHAN TEKNIKAL DAN SECURITY</b> .....		<b>40</b>
<b>LAMPIRAN 3: CV TENAGA AHLI</b> .....		<b>41</b>
PROJECT MANAGER: NURMA SUMAR SIDIK.....		41
ENGINEER 1: HENDRI.....		42
ENGINEER 2: ARIS TITO SUGIHARTO .....		45
TECHNICAL WRITER: ADAM SULTHON.....		49
<b>LAMPIRAN 4: URAIAN PEKERJAAN SEJENIS</b> .....		<b>51</b>
<b>LAMPIRAN 5: MACHINE LEARNING ON CDP OVERVIEW</b> .....		<b>53</b>
<b>LAMPIRAN 6: CLOUDERA DATA WAREHOUSE</b> .....		<b>71</b>
<b>LAMPIRAN 7: APACHE OZONE</b> .....		<b>78</b>

---

## 1 PENDAHULUAN

---

### 1.1 Latar Belakang

Teknologi Big Data di Bank Rakyat Indonesia telah diimplementasikan sejak tahun 2017 dan dari tahun ke tahun utilisasi terus tumbuh, seiring dengan peningkatan jumlah data yang masuk, pengembangan aplikasi dalam bentuk Data Engineering, Data Analyst hingga Machine Learning sudah berjalan di dalam cluster Big Data. Saat ini BRI sudah mengimplementasikan Cloudera Data Platform, dimana merupakan Platform yang terbentuk dari ekosistem data yang digunakan untuk meningkatkan availability dan agility pengembangan di sisi Data.

Implementasi Big Data di BRI tidak hanya dalam bentuk Data Storage Architecture saja, melainkan sudah masuk ke dalam Operasional Data Store. Sesuai dengan desain implementasi Data Architecture yang kembangkan oleh Divisi EDM dan disahkan pada tahun 2022, Big Data merupakan pilihan pertama untuk meletakan data ke sebelum akhirnya didistribusikan ke beberapa aplikasi yang membutuhkan seperti Pelaporan, Data Warehouse hingga Data Visualisasi dalam bentuk Dashboard maupun Data Services.

Saat ini Big Data di BRI diimplementasikan ke dalam 4 buah cluster, dimana 2 cluster berperan sebagai Datalake dan 2 cluster berperan sebagai ODS. Dimana seluruh environment ini sudah masuk ke dalam perhitungan capacity planning yang dilakukan pada tahun 2022, dengan mempertimbangkan peningkatan workload dan SLA terhadap ecosystem aplikasi yang membutuhkan akses data maka akan dilakukan kegiatan refreshment, ekspansi dan migrasi Cluster Cloudera sesuai dengan hasil rekomendasi.

Terdapat kebutuhan perpanjangan lisensi dari Cloudera Datalake DR dan Cloudera HBase DC & DR menggunakan menggunakan pola metode perhitungan baru yang sebelumnya dalam jumlah node menjadi CCU (Cloudera Compute Unit).

Sehubungan dengan kegiatan tersebut maka, BRI memerlukan penambahan lisensi sehubungan dengan penambahan jumlah node, implementasi Cloudera Data Services dan Ozone serta migrasi Cluster ODS berdasarkan hasil rekomendasi dari pihak Cloudera.

### 1.2 Tujuan

Adapun yang menjadi tujuan dari pengadaan ini adalah sebagai berikut:

- a. Refreshment, Renewal dan Expansion Lisensi Cloudera untuk Cluster (Datalake DR, Hbase DC, Hbase DR dan Datalake DC)
- b. Lisensi dan Implementasi Cloudera Data Services di dalam Cluster Cloudera.
- c. Lisensi dan Implementasi Ozone di dalam Cluster Cloudera.
- d. Penyediaan jasa Profesional Services Cloudera untuk melakukan assessment, desain, implementasi dan review deployment sehubungan dengan kegiatan refreshment, ekspansi dan migrasi seluruh cluster di BRI.
- e. Penyediaan jasa Profesional Services Cloudera untuk mengimplementasikan Cloudera Data Services.
- f. Penyediaan jasa Local Support Partner yang akan melakukan support kegiatan assessment, desain, implementasi dan review deployment implementasi refreshment dan migrasi seluruh cluster di BRI.

---

## 2 PROFIL PERUSAHAAN

---

### 2.1 Mastersystem Infotama



#### 2.1.1 Tentang Mastersystem

PT Mastersystem Infotama (Mastersystem) adalah perusahaan Teknologi dan Informasi yang terkemuka di Indonesia. Didirikan pada tahun 1994. PT Mastersystem Infotama adalah perusahaan yang bergerak sebagai penyedia produk dan layanan untuk infrastuktur teknologi informasi, produk dan layanan untuk solusi bisnis serta dijalankan oleh tim manajemen dengan pengalaman 20 tahun lebih dalam bidangnya.

Setelah ulang tahun ke-15, Mastersystem membentuk kerjasama mutual berupa partnership dengan Microsoft, Cisco, Hewlett Packard, Hitachi, EMC Documentum, dan EMC Captiva, Teradata untuk menyediakan produk dan layanan terhadap proyek-proyek yang ada di Indonesia. Empat tahun terakhir Mastersystem fokus kepada penyediaan solusi untuk sektor bank dan finansial lainnya, juga telco dan juga untuk perusahaan berbasis *oil and gas*.

Komitmen Mastersystem adalah menyediakan layanan dan solusi yang terbaik untuk meraih kepuasan konsumen yang tinggi dengan dukungan lebih dari 200 karyawan.

**Kantor Pusat:**

Sudirman 7.8 Tower 1, Lantai 25  
Jl. Jend. Sudirman Kav. 7-8  
Jakarta 10220  
Indonesia

**Kantor Cabang:**

Darmo Galeria Center B-3  
Jl. Mayjen Sungkono No. 75  
Surabaya 60225  
Indonesia

### 2.1.2 Kerjasama dengan Principal

Solution	Category	Clients
 <b>Broadridge</b>	Enterprise Reconciliation, Confirmation / Treasury Matching	                  
 <b>CLOUDERA</b>	Big Data Platform	    
 <b>teradata.</b>	Enterprise Data Warehouse	   
 <b>axway</b>	Managed File Transfer API Management	   
 <b>TEMENOS</b>	Collection System	
 <b>AT SCALE</b>	Virtual Data Warehouse	
 <b>apiGee</b> Google Cloud Platform	API Management	    

Gambar 2.1 Mastersystem Partnership

## 2.2 Cloudera

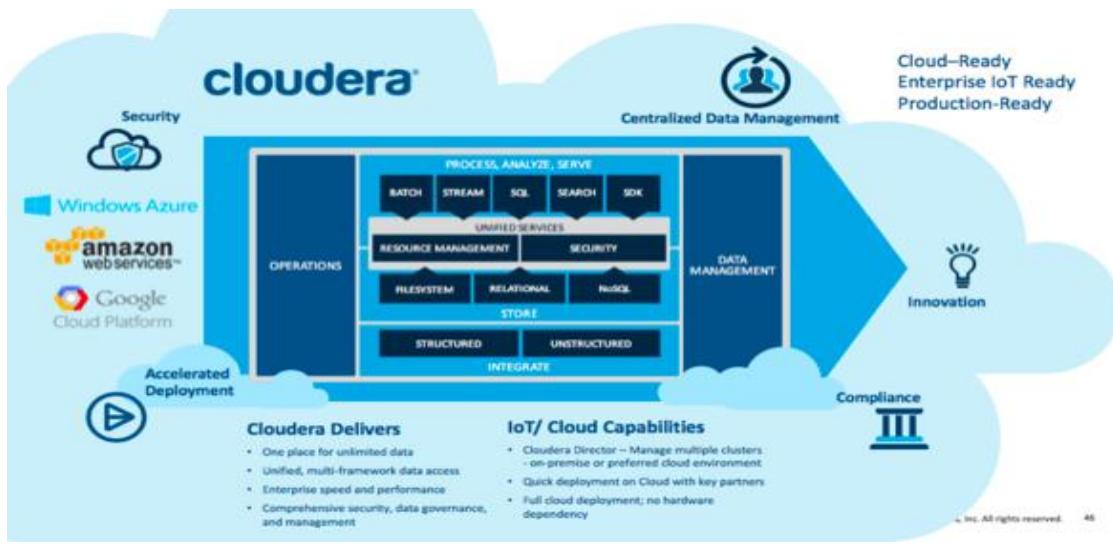
# CLUDERA

### 2.2.1 Tentang Cloudera

Cloudera merevolusi enterprise data management dengan menawarkan Platform yang pertama untuk Big Data, sebuah enterprise data hub yang dibangun pada Apache Hadoop™. Misi Cloudera adalah untuk membawa Hadoop ke Enterprise. Tagline Cloudera adalah untuk "Ask Bigger Questions" dan kami percaya "Data can make what is impossible today, possible tomorrow."

Cloudera memiliki lebih dari 1000+ subscription customers dengan melebihi 90% renewal rate. Customer terbesar kami melebihi 40,000 nodes under management. Cloudera telah mendapatkan lebih banyak saham daripada competitor gabungan lainnya. Cloudera akan terus memimpin dan berinovasi dalam data management space dengan membangun atau memperoleh produk-produk inovatif yang membantu customer kami untuk mengatasi tantangan bisnis yang kompleks – lebih cepat, lebih baik & cost-efficient.

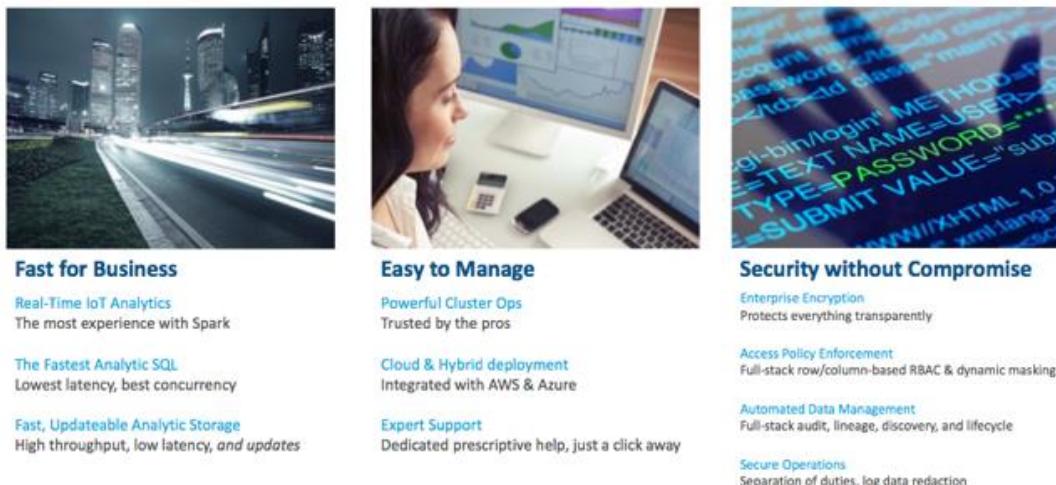
Agar Hadoop bisa memberikan solusi "Enterprise Class" Cloudera yakin bahwa environment perlu di-manage, secure, dan diatur dengan baik. Di dunia open source, sangat penting bahwa penyedia Hadoop berkontribusi pada masyarakat dan memberikan inovasi, selain lines of code. Ini adalah inovasi terdepan dari Cloudera yang telah menciptakan solusi Hadoop yang paling siap production untuk Enterprise Class customers dengan integrasi management, security, data governance, back up dan disaster recovery untuk memberi nama beberapa kekuatan market leading dari Enterprise Data Hub (EDH) Cloudera.



**Gambar 2.2 Kapabilitas Cloudera**

Cloudera dikenal sebagai kunci dari Hadoop innovator di pasaran. Chief Architect kami, Doug Cutting mengembangkan Hadoop pada tahun 2005. Cloudera berdiri pada tahun 2008, yang pertama kali memasarkan Hadoop Enterprise. Cloudera adalah contributor terbesar di ekosistem dari banyak komponen open source. Solusi Hadoop open source merupakan pendekatan solusi “arsitektur horizontal”.

Biaya dan potential lock dari “pendekatan vertical stack” terlihat jelas pada hardware maupun software yang mencoba dan mengecilkan atau meminimalkan kemampuan Hadoop sekarang dan pentingnya inovasi komunitas open source di masa depan.



**Gambar 2.3 Faktor Pembeda Cloudera**

## 2.2.2 Klien Cloudera



Gambar 2.4 Contoh Client Cloudera

Untuk referensi dan informasi lebih detail, silakan merujuk ke:

<https://www.cloudera.com/about/customers.html>

## 2.2.3 Operasi Global Cloudera

Cloudera memiliki banyak office di seluruh dunia. Kantor pusat di California, US, dan regional terdekat di Singapore. Berikut lokasi kantor global kami:

- London, England,
- Cork, Ireland,
- Munich, Germany,
- Paris, France,
- Dubai, UAE,
- Budapest, Hungary
- Melbourne, Australia,
- Beijing China.

Daftar lengkap lokasi kantor kami juga bisa dilihat di link berikut:

<https://www.cloudera.com/about/locations.html>

### Kantor Pusat:

Cloudera, Inc.  
2550 Great America Way,  
Santa Clara, CA 95054

### Kantor di Indonesia:

PT Cloudera Indonesia  
The Plaza, Lantai 7  
JI. M. H. Thamrin No. 28 – 30  
Jakarta Pusat, 10350

### 3 RUANG LINGKUP

#### 3.1 Model Lisensi

Berikut model lisensi dari solusi yang kami ajukan.

No	Product Name	Product Code	Description	Qty	
<b>1</b>	<b>Cloudera License Y1 (30 April 2023 - 29 April 2024)</b>				
<b>1.1</b>	<b>Datalake DR / Hbase DR / Hbase DC (30 April 2023 - 29 April 2024)</b>			<b>56</b>	<b>Unit</b>
	CDP Private Cloud Base Edition - Business	CDP-PVC-BASE-BUS	Cloudera Data Platform Private Cloud Base Edition - Annual Subscription per Node for up to 16 CCU for compute and up to 48 TB for storage. Business-Level Support.	56	Unit
	Cloudera Compute	COMPUTE	COMPUTE: price per CCU per year for compute in excess of 16 CCU per Node	0	Unit
	Cloudera HDFS Storage	STORAGE	HDFS STORAGE: price per TB per year.	0	Unit
<b>1.2</b>	<b>Primary / Datalake DC (28 Maret 2024 - 29 April 2024)</b>			<b>44</b>	<b>Unit</b>
	CDP Private Cloud Base Edition - Business	CDP-PVC-BASE-BUS	Cloudera Data Platform Private Cloud Base Edition - Annual Subscription per Node for up to 16 CCU for compute and up to 48 TB for storage. Business-Level Support.	44	Unit
	Cloudera Compute	COMPUTE	COMPUTE: price per CCU per year for compute in excess of 16 CCU per Node	2112	Unit
	Cloudera HDFS Storage	STORAGE	HDFS STORAGE: price per TB per year.	1632	Unit
	Cloudera Data Science Workbench, 10-Pack	CE-DSW	Subscription for a block of 10 unique identifiers (users) for the Cloudera Data Science Workbench product. Cloudera Data Science Workbench must be ordered with a subscription to the Data Hub Edition or Data Engineering Edition of the Cloudera Enterprise platform, and Customer must agree to Cloudera's CDSW T&Cs.	1	Unit
<b>1.3</b>	<b>Expansion Datalake DC ( 30 November 2023 - 29 April 2024)</b>			<b>26</b>	<b>Unit</b>
	CDP Private Cloud Base Edition - Business	CDP-PVC-BASE-BUS	Cloudera Data Platform Private Cloud Base Edition - Annual Subscription per Node for up to 16 CCU for compute and up to 48 TB for storage. Business-Level Support.	26	Unit
	Cloudera Compute	COMPUTE	COMPUTE: price per CCU per year for compute in excess of 16 CCU per Node	352	Unit
	Cloudera HDFS Storage	STORAGE	HDFS STORAGE: price per TB per year.	864	Unit
	Ozone or any 3rd Party Storage	STORAGE-OS3	Ozone or 3rd Party Storage price per TB per year.	1152	Unit
<b>1.4</b>	<b>Expansion Hbase DC ( 30 November 2023 - 29 April 2024)</b>			<b>10</b>	<b>Unit</b>
	CDP Private Cloud Base Edition - Business	CDP-PVC-BASE-BUS	Cloudera Data Platform Private Cloud Base Edition - Annual Subscription per Node for up to 16 CCU for compute and up to 48 TB for storage. Business-Level Support.	10	Unit
	Cloudera Compute	COMPUTE	COMPUTE: price per CCU per year for compute in excess of 16 CCU per Node	96	Unit
	Cloudera HDFS Storage	STORAGE	HDFS STORAGE: price per TB per year.	0	Unit

No	Product Name	Product Code	Description	Qty
<b>1.5</b>	<b>Data Services Datalake DC ( 30 November 2023 - 29 April 2024)</b>			<b>704</b> <b>Unit</b>
	CDP Private Cloud Data Services per CCU - Business	CDP-PVC-DTSC-CCU-BUS	Cloudera Data Platform Private Cloud Data Services: Annual Subscription per CCU. Business-Level Support. AGPL License.	704 Unit
<b>1.6</b>	<b>CDSW for DR (30 April 2023 - 29 April 2024)</b>			<b>1</b> <b>Unit</b>
	Cloudera Data Science Workbench, 10 User Pack - Business	CDP-DSW-10-BUS	Cloudera Data Science Workbench Applications: Annual Subscription for a block of 10 unique identified Users. Business Support.	1 Unit
<b>1.7</b>	<b>NiFi Expansion (30 April 2023 - 29 April 2024)</b>			<b>32</b> <b>Unit</b>
	Cloudera Data Platform Flow Management Edition Subscription, 4 Cores - Business Support	CDP-CFM-4-BUS	Cloudera Data Platform Flow Management Edition - Annual Subscription, 4 Cores - Business Support. AGPL License.	32 Unit
<b>2</b>	<b>Cloudera License Y2 (30 April 2024 - 29 April 2025)</b>			
<b>1.1</b>	<b>Datalake DR / Hbase DR / Hbase DC (30 April 2024 - 29 April 2025)</b>			<b>56</b> <b>Unit</b>
	CDP Private Cloud Base Edition - Business	CDP-PVC-BASE-BUS	Cloudera Data Platform Private Cloud Base Edition - Annual Subscription per Node for up to 16 CCU for compute and up to 48 TB for storage. Business-Level Support.	56 Unit
	Cloudera Compute	COMPUTE	COMPUTE: price per CCU per year for compute in excess of 16 CCU per Node	0 Unit
	Cloudera HDFS Storage	STORAGE	HDFS STORAGE: price per TB per year.	0 Unit
<b>1.2</b>	<b>Primary / Datalake DC (30 April 2024 - 29 April 2025)</b>			<b>44</b> <b>Unit</b>
	CDP Private Cloud Base Edition - Business	CDP-PVC-BASE-BUS	Cloudera Data Platform Private Cloud Base Edition - Annual Subscription per Node for up to 16 CCU for compute and up to 48 TB for storage. Business-Level Support.	44 Unit
	Cloudera Compute	COMPUTE	COMPUTE: price per CCU per year for compute in excess of 16 CCU per Node	2112 Unit
	Cloudera HDFS Storage	STORAGE	HDFS STORAGE: price per TB per year.	1632 Unit
	Cloudera Data Science Workbench, 10-Pack	CE-DSW	Subscription for a block of 10 unique identifiers (users) for the Cloudera Data Science Workbench product. Cloudera Data Science Workbench must be ordered with a subscription to the Data Hub Edition or Data Engineering Edition of the Cloudera Enterprise platform, and Customer must agree to Cloudera's CDSW T&Cs.	1 Unit
<b>1.3</b>	<b>Expansion Datalake DC (30 April 2024 - 29 April 2025)</b>			<b>26</b> <b>Unit</b>

No	Product Name	Product Code	Description	Qty	
	CDP Private Cloud Base Edition - Business	CDP-PVC-BASE-BUS	Cloudera Data Platform Private Cloud Base Edition - Annual Subscription per Node for up to 16 CCU for compute and up to 48 TB for storage. Business-Level Support.	26	Unit
	Cloudera Compute	COMPUTE	COMPUTE: price per CCU per year for compute in excess of 16 CCU per Node	352	Unit
	Cloudera HDFS Storage	STORAGE	HDFS STORAGE: price per TB per year.	864	Unit
	Ozone or any 3rd Party Storage	STORAGE-OS3	Ozone or 3rd Party Storage price per TB per year.	1152	Unit
<b>1.4</b>	<b>Expansion Hbase DC (30 April 2024 - 29 April 2025)</b>			<b>10</b>	<b>Unit</b>
	CDP Private Cloud Base Edition - Business	CDP-PVC-BASE-BUS	Cloudera Data Platform Private Cloud Base Edition - Annual Subscription per Node for up to 16 CCU for compute and up to 48 TB for storage. Business-Level Support.	10	Unit
	Cloudera Compute	COMPUTE	COMPUTE: price per CCU per year for compute in excess of 16 CCU per Node	96	Unit
	Cloudera HDFS Storage	STORAGE	HDFS STORAGE: price per TB per year.	0	Unit
<b>1.5</b>	<b>Data Services Datalake DC (30 April 2024 - 29 April 2025)</b>			<b>704</b>	<b>Unit</b>
	CDP Private Cloud Data Services per CCU - Business	CDP-PVC-DTSC-CCU-BUS	Cloudera Data Platform Private Cloud Data Services: Annual Subscription per CCU. Business-Level Support. AGPL License	704	Unit
<b>1.6</b>	<b>CDSW for DR (30 April 2024 - 29 April 2025)</b>			<b>1</b>	<b>Unit</b>
	Cloudera Data Science Workbench, 10 User Pack - Business	CDP-DSW-10-BUS	Cloudera Data Science Workbench Applications: Annual Subscription for a block of 10 unique identified Users. Business Support.	1	Unit
<b>1.7</b>	<b>NiFi Expansion (30 April 2024 - 29 April 2025)</b>			<b>32</b>	<b>Unit</b>
	Cloudera Data Platform Flow Management Edition Subscription, 4 Cores - Business Support	CDP-CFM-4-BUS	Cloudera Data Platform Flow Management Edition - Annual Subscription, 4 Cores - Business Support. AGPL License.	32	Unit
<b>3</b>	<b>Cloudera License Y3 (30 April 2025 - 29 April 2026)</b>				
<b>1.1</b>	<b>Datalake DR / Hbase DR / Hbase DC (30 April 2025 - 29 April 2026)</b>			<b>56</b>	<b>Unit</b>
	CDP Private Cloud Base Edition - Business	CDP-PVC-BASE-BUS	Cloudera Data Platform Private Cloud Base Edition - Annual Subscription per Node for up to 16 CCU for compute and up to 48 TB for storage. Business-Level Support.	56	Unit
	Cloudera Compute	COMPUTE	COMPUTE: price per CCU per year for compute in excess of 16 CCU per Node	0	Unit
	Cloudera HDFS Storage	STORAGE	HDFS STORAGE: price per TB per year.	0	Unit
<b>1.2</b>	<b>Primary / Datalake DC (30 April 2025 - 29 April 2026)</b>			<b>44</b>	<b>Unit</b>

No	Product Name	Product Code	Description	Qty	
	CDP Private Cloud Base Edition - Business	CDP-PVC-BASE-BUS	Cloudera Data Platform Private Cloud Base Edition - Annual Subscription per Node for up to 16 CCU for compute and up to 48 TB for storage. Business-Level Support.	44	Unit
	Cloudera Compute	COMPUTE	COMPUTE: price per CCU per year for compute in excess of 16 CCU per Node	2112	Unit
	Cloudera HDFS Storage	STORAGE	HDFS STORAGE: price per TB per year.	1632	Unit
	Cloudera Data Science Workbench, 10-Pack	CE-DSW	Subscription for a block of 10 unique identifiers (users) for the Cloudera Data Science Workbench product. Cloudera Data Science Workbench must be ordered with a subscription to the Data Hub Edition or Data Engineering Edition of the Cloudera Enterprise platform, and Customer must agree to Cloudera's CDSW T&Cs.	1	Unit
<b>1.3</b>	<b>Expansion Datalake DC (30 April 2025 - 29 April 2026)</b>			<b>26</b>	<b>Unit</b>
	CDP Private Cloud Base Edition - Business	CDP-PVC-BASE-BUS	Cloudera Data Platform Private Cloud Base Edition - Annual Subscription per Node for up to 16 CCU for compute and up to 48 TB for storage. Business-Level Support.	26	Unit
	Cloudera Compute	COMPUTE	COMPUTE: price per CCU per year for compute in excess of 16 CCU per Node	352	Unit
	Cloudera HDFS Storage	STORAGE	HDFS STORAGE: price per TB per year.	864	Unit
	Ozone or any 3rd Party Storage	STORAGE-OS3	Ozone or 3rd Party Storage price per TB per year.	1152	Unit
<b>1.4</b>	<b>Expansion Hbase DC (30 April 2025 - 29 April 2026)</b>			<b>10</b>	<b>Unit</b>
	CDP Private Cloud Base Edition - Business	CDP-PVC-BASE-BUS	Cloudera Data Platform Private Cloud Base Edition - Annual Subscription per Node for up to 16 CCU for compute and up to 48 TB for storage. Business-Level Support.	10	Unit
	Cloudera Compute	COMPUTE	COMPUTE: price per CCU per year for compute in excess of 16 CCU per Node	96	Unit
	Cloudera HDFS Storage	STORAGE	HDFS STORAGE: price per TB per year.	0	Unit
<b>1.5</b>	<b>Data Services Datalake DC (30 April 2025 - 29 April 2026)</b>			<b>704</b>	<b>Unit</b>
	CDP Private Cloud Data Services per CCU - Business	CDP-PVC-DTSC-CCU-BUS	Cloudera Data Platform Private Cloud Data Services: Annual Subscription per CCU. Business-Level Support. AGPL License.	704	Unit
<b>1.6</b>	<b>CDSW for DR (30 April 2025 - 29 April 2026)</b>			<b>1</b>	<b>Unit</b>
	Cloudera Data Science Workbench, 10 User Pack - Business	CDP-DSW-10-BUS	Cloudera Data Science Workbench Applications: Annual Subscription for a block of 10 unique identified Users. Business Support.	1	Unit
<b>1.7</b>	<b>NiFi Expansion (30 April 2025 - 29 April 2026)</b>			<b>32</b>	<b>Unit</b>

No	Product Name	Product Code	Description	Qty	
	Cloudera Data Platform Flow Management Edition Subscription, 4 Cores - Business Support	CDP-CFM-4-BUS	Cloudera Data Platform Flow Management Edition - Annual Subscription, 4 Cores - Business Support. AGPL License.	32	Unit

## 3.2 Ruang Lingkup yang Diajukan

Proposal ini mencakup assessment, desain, implementasi dan review deployment sehubungan dengan Cluster Cloudera di BRI. Adapun uraian pekerjaan dan ruang lingkup pengadaan yang dibutuhkan BRI adalah:

### 3.2.1 Teknis

Berikut adalah informasi teknis berdasarkan kebutuhan Bank BRI:

- Penggunaan Framework Project Management di dalam pelaksanaan implementasi project hingga project dapat dilaksanakan secara tepat waktu, tepat anggaran dan tepat target.
- Mampu melaksanakan assessment sehubungan dengan tahap persiapan instalasi, migrasi, dan Implementasi Cluster Cloudera BRI.
- Melakukan proses assessment sehubungan dengan infrastruktur dan deployment cluster Big Data di BRI.
- Memberikan alternatif atau *workaround* yang akan dibahas dengan BRI dan menjadi kesepakatan di dalam implementasi Cluster Cloudera di BRI.
- Melakukan integrasi kepada keseluruhan *platform existing* BRI di dalam solusi yang ditawarkan hingga mampu mengakomodir *workload environment* dengan *surrounding application*.

### 3.2.2 Aplikasi

Terkait aplikasi, akan berfokus pada kegiatan assessment, desain, implementasi dan review deployment Cluster Cloudera bank BRI yang memiliki minimum fungsi namun tidak terbatas pada:

- Arsitektur Implementasi dengan mekanisme *Distributed* atau *Clustering*, dimana mempertimbangkan kepada roadmap implementasi platform data di BRI.
- Arsitektur implementasi sesuai dengan best practice berdasarkan rekomendasi Principal Cloudera.
- Strategi Migrasi Workload Existing ke lingkungan pengembangan Baru.

### 3.2.3 Hardware

Kegiatan assessment, desain, implementasi dan review deployment Cluster Cloudera di BRI memiliki minimum kapabilitas namun tidak terbatas pada:

- Perangkat Utama:  
Mastersystem memberikan rincian kebutuhan infrastruktur sesuai kebutuhan dan spesifikasi yang dipersyaratkan dalam dokumen ini dan aplikasi pendukung yang dibutuhkan.
- Assessment:  
Assessment sehubungan dengan ketersediaan perangkat wajib mengikuti strategi best practice yang ditawarkan oleh Cloudera sehingga mampu memberikan akses kepada maksimum kapabilitas Cloudera dalam menyediakan teknologi perangkat.

### 3.2.4 Implementasi

Mastersystem wajib melaksanakan implementasi perencanaan implementasi sesuai dengan jangka waktu, ruang lingkup dan informasi dibawah ini:

- Melakukan proses assessment sehubungan dengan infrastruktur dan deployment cluster Big Data di BRI. Adapun proses assessment adalah sebagai berikut:

- a. Melakukan review terhadap kondisi existing Cluster sebelum dilakukan migrasi server dan penambahan worker node.
- b. Melakukan review sehubungan dengan arsitektur server Baremetal secara compute, storage, network dan lain-lain.
- c. Melakukan review sehubungan dengan workload yang berjalan di eksisting cluster baik untuk Datalake dan ODS.
- Melakukan proses assessment, desain, implementasi dan review deployment sehubungan dengan kegiatan yang telah ditentukan sebagai berikut:
  - a. Melakukan proses Migration Planning sehubungan dengan roadmap refreshment server pada Cluster Cloudera Datalake DC BRI.
  - b. Melakukan proses Migration Planning sehubungan dengan roadmap peningkatan kapabilitas performance Cluster Hbase DC BRI.
  - c. Melakukan proses Implementation & Integration Planning sehubungan dengan ekspansi kapasitas Cluster Cloudera Datalake DC beserta dengan implementasi Ozone untuk keperluan archiving data, berikut dengan data pipeline untuk mekanisme proses archiving data.
  - d. Melakukan proses Implementation & Integration Planning sehubungan dengan implementasi produk Cloudera Data Services pada Cluster Cloudera Datalake DC BRI.
  - e. Melakukan kegiatan migrasi server untuk 44 node Cluster Datalake DC (Old Server) ke server Baremetal (Refreshment) berikut dengan data dan aplikasi di dalam Cluster tersebut. Berikut juga dengan kegiatan konfigurasi secure dan optimizing cluster.
  - f. Melakukan kegiatan migrasi server untuk Cluster Hbase DC dari Infrastruktur Openstack ke Server Baremetal berikut dengan data dan aplikasi di dalam Cluster tersebut. Berikut juga dengan kegiatan konfigurasi secure dan optimizing cluster.
  - g. Melakukan kegiatan penambahan worker node sebanyak 26 server untuk HDFS Cluster Datalake DC berikut dengan pemasangan lisensi.
  - h. Melakukan kegiatan implementasi sehubungan dengan pembaharuan lisensi untuk Cluster Datalake ODC, Hbase DC dan ODC.
  - i. Melakukan kegiatan implementasi produk Cloudera Data Services dan Ozone berikut dengan instalasi, pemasangan lisensi, functional test, dan non-functional test. Berikut juga dengan kegiatan konfigurasi secure dan optimizing cluster.
  - j. Melakukan pembuktian dengan cara kegiatan piloting penggunaan Cloudera Data Services, sekurang-kurangnya sehubungan dengan kegiatan
    - Pengembangan Model dan Deployment Model dengan menggunakan Cloudera Machine Learning
    - Pengembangan Datalake house dengan menggunakan Cloudera Data Warehouse yang terintegrasi prosesnya dengan platform Integration Tools Existing seperti Informatica dan NiFi.
- Memahami kebutuhan, kondisi, dan pekerjaan dari user pihak BRI. Dalam hal ini melakukan pendampingan dan analisa sehubungan dengan:
  - a. Tahapan Analisa sehubungan dengan proses persiapan implementasi use case.
  - b. Tahapan Pengujian Persiapan dari hasil analisa sehubungan dengan proses persiapan implementasi use case.
  - c. Tahapan implementasi use case beserta dengan mitigasi risikonya.

d. Tahapan pendampingan dan pengujian setelah proses implementasi use case pada lingkungan yang berinteraksi dengan Cluster seperti:

- Informatica DEI dan EDC
- CDSW
- AtScale
- Confluent
- Spark Job
- Nifi
- Proses pembacaan dan penulisan data dari dan ke lingkungan database sekurang-kurangnya MySQL, SQL Server, Hbase, Postgresql, DB2, Cassandra SQL. Termasuk aplikasi surrounding yang dikembangkan internal BRI dalam bentuk Microservices.
- Berkolaborasi di dalam proses dokumentasi sehubungan dengan administrasi yang ada di BRI.
- Melakukan penyusunan dan implementasi strategi change management, communication, training (CCMT) serta business deployment.
- Melakukan penyusunan dan implementasi strategi untuk functional & technical testing.
- Melakukan proses functional & technical testing untuk Cluster Cloudera yang telah diimplementasikan.
- Mendokumentasikan seluruh proses implementasi dalam bentuk sekurang-kurangnya dokumen SDLC sesuai dengan standar serta terintegrasi pada Confluence dan JIRA di BRI.
- Tim harus melakukan demo progress pekerjaan setiap satu minggu sekali di hari dan waktu yang ditentukan kemudian dan sudah terintegrasi pada Confluence & JIRA di BRI.
- Berkaitan jika salah satu anggota tim tidak bisa hadir, vendor berkewajiban untuk menyediakan pengganti;
- Biaya konsumsi, kesehatan, alat komunikasi, laptop/pc, akomodasi, transportasi, software (di luar yang sudah disediakan oleh BRI) dan biaya-biaya lain di luar biaya pembelian mandays adalah di luar tanggung jawab BRI;
- Dapat bekerjasama dan berkomunikasi dengan pekerja lain dalam tim dan bertanggung jawab terhadap kewajiban masing-masing dengan baik.
- Memiliki integritas, profesionalisme, dan patuh terhadap peraturan yang ada di Bank Rakyat Indonesia.
- Seluruh hasil pekerjaan dalam dokumen yang dihasilkan adalah merupakan hak milik BRI dan tidak diperkenankan untuk menyimpan, mengelola, menyebarluaskan dan menggunakan pada project lain di luar BRI.
- Seluruh informasi dan data milik BRI tidak boleh dan dilarang keras untuk disebarluaskan tanpa seijin pihak BRI.

### 3.2.5 Transfer Knowledge

Mastersystem melakukan proses transfer knowledge terhadap hasil pekerjaan kepada tim BRI. Rinciannya sebagai berikut:

1. Menyediakan transfer knowledge dengan materi Administrator platform Cloudera yang baru diimplementasikan di BRI:
  - o Pemaparan architecture implementasi platform Cloudera yang telah diimplementasikan.
  - o Proses instalasi, upgrade dan konfigurasi platform Cloudera yang telah diimplementasikan.
  - o Proses monitoring platform Cloudera yang telah diimplementasikan dengan integrasi ke surrounding application.
  - o Proses backup, restore dan snapshot platform Cloudera.
  - o Proses eskalasi issue apabila terjadi problem dari komunikasi kepada Local Support hingga Support Portal platform Cloudera yang telah diimplementasikan.
  - o Proses eskalasi issue ke BRI apabila terjadi masalah diluar platform Cloudera yang berimpact terhadap platform Cloudera.
2. Menyediakan Training dengan materi Developer dan Administrator platform Cloudera yang dapat diakses secara on-demand dengan tujuan untuk:
  - o Mempelajari seluruh Platform Cloudera yang diimplementasikan di BRI.
  - o Mempelajari basic troubleshoot yang sering terjadi di dalam platform Cloudera.
  - o Mempelajari best practice yang digunakan dalam penggunaan Ozone sebagai data archive tools Big Data Cluster.
  - o Cara penggunaan Platform Cloudera Data Services dan komponennya secara menyeluruh.
  - o Cara melakukan migrasi dari Cloudera Data Science Workbench ke Cloudera Machine Learning.
  - o Membangun end to end use cases dengan memanfaatkan surrounding platform di BRI.

Pelaksanaan training/transfer knowledge dilakukan secara hybrid (onsite dan/atau online) menyesuaikan dengan kondisi dan kebutuhan pelatihan yang sesuai.

### 3.3 Warranty dan Support

Mastersystem memberikan warranty dan support sesuai jangka waktu, layanan/support yang terdiri dari:

- Menyediakan Premium Support termasuk support feature dari Principal penyedia platform Cloudera.
- Project Manager secara berkala melakukan diskusi sehubungan dengan implementasi platform Cloudera baik untuk dukungan Production maupun Development.
- Menyediakan jasa dukungan dan roadmap kebutuhan human resources local support dengan resources yang telah menjalani training, memiliki pengalaman dan tersertifikasi yang diberikan oleh principal platform Cloudera.
- Melakukan monitoring sesuai operational model yang disepakati selama jangka waktu managed service dan menyediakan sistem/tool monitoring yang dapat diakses secara online oleh BRI dan/atau terintegrasi dengan sistem monitoring yang ada di BRI.
- Menyediakan Warranty Support terhadap hasil implementasi platform Cloudera.
- Melakukan proses transfer knowledge terhadap hasil pekerjaan kepada tim BRI.

### 3.4 Garansi/Warranty

Mastersystem memberikan jaminan dan dukungan sesuai dengan jangka waktu, maintenance dan jaminan layanan (SLA). Jangka waktu jaminan garansi selama 3 (Tiga) bulan setelah go-live.

#### 3.4.1 Post Go-Live Support

Memberikan jaminan perbaikan pasca implementasi (setelah tanggal go-live) selama 12 (dua belas) bulan. Layanan yang diberikan minimum namun tidak terbatas pada bugs fix, update patch, penambahan plugin dan perbaikan lainnya untuk memastikan solusi berjalan dengan baik dan sesuai dengan kebutuhan. Pemenuhan layanan dilaksanakan sesuai dengan SLA yang terdapat dalam dokumen ini.

#### 3.4.2 Preventive Maintenance

Melakukan pemeliharaan untuk menjaga performansi aplikasi dan perangkat agar beroperasi secara optimal. Layanan preventive maintenance minimum namun tidak terbatas pada (update patch, penambahan version/plugin, load testing untuk aplikasi dan pengecekan fisik maupun system untuk perangkat). Jadwal pengecekan dilakukan minimal 4 kali dalam 1 tahun secara on site dan memberikan dokumentasi kegiatan preventive yang telah dilaksanakan.

#### 3.4.3 Corrective Maintenance

Dilakukan selama jangka waktu managed service. Memberikan layanan perbaikan dan pemulihan ketika terjadi kendala dengan ketentuan sebagai berikut:

- a. Memiliki manajemen respon (termasuk metode) dan sumber daya yang memadai untuk mencapai SLA yang ditetapkan.
- b. Memberikan layanan helpdesk 24 jam yang dapat diakses melalui media telefon, email dan social media seperti whatsapp (optional).
- c. Melakukan perbaikan dan pemulihan sistem ketika terjadi kendala atau aplikasi tidak dapat digunakan sesuai dengan SLA yang terdapat dalam dokumen ini.
- d. Pelaksana pekerjaan harus dapat menyediakan SDM yang on site dan on call ketika terjadi kendala atas permintaan BRI.

### 3.4.4 Service Level Agreement (SLA)

Klasifikasi Persoalan	Definisi	Waktu Respon	Waktu Resolusi
<b>EMERGENCY (Severity 1)</b>	Persoalan yang menyebabkan fungsi-fungsi utama tidak bekerja sama sekali sehingga menyebabkan sistem gagal beroperasi secara keseluruhan.	Sejak pemberitahuan, maka penugasan diberikan dalam waktu maksimum 30 menit, dan support akan segera diberikan onsite.	Support akan diberikan dalam waktu maksimum 4 jam, dan support akan segera diberikan online (On-call basis).
<b>HIGH (Severity 2)</b>	Persoalan yang menyebabkan penurunan yang sangat tajam dari kinerja fungsi-fungsi utama yang menyebabkan fungsi-fungsi tertentu tidak berfungsi.	Sejak pemberitahuan, maka penugasan dan respon akan segera diberikan secepatnya dalam waktu 1 jam.	Support akan diberikan dalam waktu maksimum 12 jam, dan support akan segera diberikan online (On-call basis).
<b>LOW (Severity 3)</b>	Persoalan yang tidak terlalu mempengaruhi sistem secara keseluruhan. Contohnya: <ul style="list-style-type: none"> <li>• Penurunan fungsi-fungsi administrasi rutin.</li> <li>• <i>User Interface</i> yang membingungkan. Misal: Letak menu, tombol, atau langkah-langkah yang harus dilakukan.</li> <li>• Kesalahan dokumentasi.</li> </ul>	Sejak pemberitahuan, maka penugasan dan respon akan segera diberikan secepatnya dalam waktu 1 jam.	Support akan diberikan dalam waktu maksimum 48 jam (2 Hari), dan support akan segera diberikan online (On-call basis).

### 3.4.5 Support lainnya

- On-Site 1 Engineer 7x24 jam dan on call diluar jam kerja dimana bertugas melakukan pemeriksaan secara berkala dan berdiskusi dengan team BRI sehubungan dengan laporan mingguan.
- Solution Architecture (Ad-Hoc) yang bertugas memberikan pendampingan sehubungan dengan perencanaan penggunaan platform Cloudera untuk multiple use case yang akan diimplementasikan di BRI.

### 3.5 Dokumentasi

Dokumentasi yang dibutuhkan di dalam proses transfer knowledge dan sosialisasi sekurang-kurangnya menyampaikan dan membahas sehubungan dengan:

1. Dokumentasi hasil review dari eksisting arsitektur dan target arsitektur.
2. Dokumen analisa kebutuhan dan tahapan persiapan proses assessment, desain, implementasi dan review deployment sehubungan dengan kegiatan yang telah ditentukan.
3. Dokumen Normal Change Management berkolaborasi dengan Pihak BRI.
4. Dokumen Emergency Change Management apabila dibutuhkan berkolaborasi dengan pihak BRI.
5. Dokumentasi hasil implementasi platform Cloudera, termasuk dengan tahapan implementasi, kendala, penyelesaiannya.
6. Dokumentasi functional test dan non-functional test dari implementasi platform Cloudera.
7. Dokumentasi rekomendasi yang diperoleh dari hasil review dan diskusi kebutuhan dari BRI.
8. Dokumen Standard Operational Procedure (SOP) untuk penggunaan produk Ozone dan Cloudera Data Services, termasuk monitoring, troubleshoot, dan user management.
9. Dokumentasi petunjuk dan mitigasi resiko ketika mengimplementasikan platform Cloudera.

## 4 METODOLOGI

### 4.1 Manajemen Proyek

Bab ini menjelaskan metodologi yang umumnya kami gunakan. Karena pekerjaan yang kami ajukan merupakan bagian dari solusi overall:

- Anggota tim Project dan metodologi akan disesuaikan dengan scope pekerjaan yang diajukan.
- Project Manager adalah Project Manager internal untuk menangani scope pekerjaan yang diajukan, bukan Project Manager overall sebagai system integrator.

#### 4.1.1 Project Communication Plan

Penting untuk menjaga komunikasi efektif dan efisien selama durasi proyek. Komunikasi proyek perlu diatur dengan standar proses dengan contoh sebagai berikut.

		Mastersystem				Bank BRI			
Periode	Meeting dan Komunikasi	PM	SA	TL	TM	PM	SA	TL	TM
Harian	Status defect & issue dan rencana perbaikan ketika periode SIT dan UAT	M	O	O	O	M	O	O	O
Mingguan	Mastersystem Status Meeting	M	M	M	M				
	Project Status Review Meeting	M	O	M	O	M	O	M	O
	Bank BRI Internal					M	M	M	M
Bulanan	Project Progress Review (Steering Committee) Meeting	M		O		M		O	
	Project Health Check / Governance Meeting	M							
Ad hoc, sesuai kebutuhan	Acceptance Review Meeting	M	O	M	O	M	O	M	O
	Change Control Board Meeting	M	O	M	O	M	O	M	O
	Project Schedule / Updates	M	O	M	O	M	O	M	O
	Technical Meeting	M	O	M	O	M	O	M	O
	Deliverables review meeting	M	O	M	O	M	O	M	O

PM – Project Manager, SA – Solution Architect, TL – Technical Lead, TM – Team Member

M – Mandatory, O – Optional

#### 4.1.2 Prosedur Eskalasi

Project Manager BRI dan Mastersystem akan bertanggung jawab untuk identifikasi dan manajemen issue proyek. Issue akan dicatat dan diassign ke PIC yang tepat untuk disolusikan. Issue yang mengakibatkan perubahan pada Project Plan/Timeline dan juga Scope of Works akan didokumentasi secara lengkap dan disetujui/tanda tangan bersama oleh Project Steering Committee. Kemudian proses akan mengikuti prosedur Change Control.

Eskalasi issue akan mengikuti contoh penjelasan berikut:

Committee	Komposisi	Proses Pengambilan Keputusan	Eskalasi
Joint Steering Committee	BRI: - Steering Committee - Program Business Sponsor - Program Manager  Mastersystem: - Steering Committee - Relationship/Account Manager - Program Manager	Steering Committee akan berdiskusi dan memutuskan. Jika masih belum dapat diambil keputusan, maka akan dieskalasikan lebih jauh ke Senior Management untuk memberikan arahan kepada tim operasional.	
Program Working Committee	BRI: - Program Manager - Program Enterprise Architect - Project managers – functional, technical and implementation  Mastersystem: - Program Manager - Project Manager - Technical Leads	Pengambilan keputusan oleh komite akan diambil berdasarkan diskusi. Jika tidak tercapai keputusan, omette akan merekomendasikan saran dalam bentuk opsi, alasan, dan resiko kepada chairman. Chairman kemudian akan memilih rekomendasi opsi. Issue yang masih belum terpecahkan akan dieskalasi ke Joint Steering Committee untuk keputusan final.	Joint Steering Committee

#### 4.1.3 Bank BRI Steering Committee

Steering Committee berhubungan dengan keputusan strategis yang mempengaruhi proyek, dan juga mungkin dibutuhkan untuk bertindak taktis. Ini memastikan progress tetap berlanjut, perubahan dilacak dan diatur, dan issue yang tidak dapat dipecahkan oleh tim Proyek dapat disolusikan di level Steering Committee.

Mastersystem akan hadir sesuai dan jika direquest oleh Project Owner Bank BRI untuk mempresentasikan point tertentu atau menjawab pertanyaan spesifik Mastersystem dari sponsor bisnis. Mastersystem akan diberikan pemberitahuan lebih awal untuk mengidentifikasi peserta yang tepat untuk meeting ini dan mengundangnya.

#### **4.1.4 Proses Penanganan Intellectual Property**

Tim Bank BRI dan Mastersystem harus setuju untuk menghormati kerahasiaan dari material masing-masing. Untuk memastikan IP terlindungi, setiap anggota tim akan perlu mengikuti aturan berikut:

BRI akan menyediakan sistem tersentral untuk menyimpan materi IP. Hanya orang-orang dengan peran aktif dalam proyek ini yang akan diberikan akses, dan untuk setiap pemberian akses akan diingatkan dengan agreement kerahasiaan. Adalah tanggung jawab Project Manager Bank BRI untuk memastikan bahwa tidak ada pihak lain yang dapat mengakses direktori proyek tersebut.

### **4.2 Manajemen Resiko Proyek**

Sub bab berikut menjelaskan Risk Management Process, yang akan dilakukan sejalan Proyek.

#### **4.2.1 Risk Management**

Proyek direncanakan untuk mendapatkan hasil tertentu, dan ada kemungkinan bahwa hasil tersebut dapat dicapai dikarenakan banyak faktor ("risk event") yang dapat terjadi yang membuat tujuan awal sulit atau tidak ekonomis untuk dilakukan. Manajemen Proyek yang baik menangani faktor-faktor ini untuk menghindari gagalnya suatu proyek. Proses Risk Management memberikan Project Manager cara yang terstruktur untuk mempersiapkan event-event seperti itu.

#### **4.2.2 Sumber Utama Resiko**

Dari pengalaman kami, resiko utama yang terjadi pada project biasanya terkait dengan:

- Analisa detail yang dibutuhkan, terkait banyak data, definisi, dan kepemilikan data.
- Menganggap remeh impact dari issue data quality.
- Delay memberikan feedback review, fasilitas, informasi dan testing.
- Kebutuhan memberikan hasil yang cepat untuk mendemonstrasikan nilai kepada sponsor bisnis.
- Definisi persetujuan kriteria sukses.
- Pengukuran kriteria sukses.
- Kepatuhan terhadap kriteria sukses yang disetujui.
- Ekspektasi yang tinggi dari semua personil yang terlibat.
- Kebutuhan dan ekspektasi klien yang tidak jelas atau tidak diekspresikan (scope creep).
- Spesifikasi yang tidak akurat.
- Kebutuhan baru (misal: perubahan regulasi pemerintah, fokus bisnis baru).
- Perubahan yang diperlukan pada organisasi klien tidak ditangani secara memadai.
- Masalah atau kegagalan teknis.
- Reorganisasi major pada organisasi yang terkait.
- Kegagalan mengantar hardware atau software sesuai waktu.

Namun, berapapun faktor yang mempengaruhi kesuksesan proyek, selama resiko dapat diidentifikasi dan dimengerti, tindakan perlindungan bisa dilakukan dengan membuat strategi mitigasi. Oleh karena itu praktik manajemen proyek yang baik memiliki proses manajemen resiko untuk mengidentifikasi, mengkuantifikasi dan mitigasi risk secara periodik reguler.

Proses manajemen resiko terdiri dari sub proses berikut:

- Identifikasi resiko.

- Kuantifikasi risiko.
- Pengembangan respon terhadap resiko.
- Kendali respon terhadap resiko.
- Respon terhadap resiko.

#### **4.2.3 Tindakan Mitigasi yang Umum**

Risk dapat ditangani dengan salah satu cara berikut:

- Menghindari resiko (avoid): menghilangkan ancaman dari resiko, biasanya dengan mengeliminasi penyebabnya. Sebagai contoh, ini dapat dilakukan dengan memilih pendekatan yang berbeda, sehingga aktivitas tertentu dapat dihilangkan dari Project Plan atau menggantikan mereka dengan pendekatan yang tidak memberikan resiko.
- Mitigasi resiko: mengurangi kemungkinan terjadinya resiko atau mengurangi impact dari resiko. Sebagai contoh, meminimalkan impact dari PIC kunci meninggalkan proyek dengan cara mengadakan backup untuk setiap posisi kunci.
- Menerima risk (accept): mengambil keputusan secara sadar untuk menerima resiko, misal karena biaya untuk mitigasi terlalu tinggi dibandingkan dengan impact nya, atau jika tidak ada mitigasi atau menghindar tidak dapat dilakukan.
- Transfer risk: untuk menghilangkan resiko dari tim proyek dan mentransfer nya ke tim lain. Sebagai contoh, membuat tim lain bertanggung jawab terhadap aktivitas tertentu dapat menjadi salah satu cara mentransfer resiko.

#### **4.2.4 Proses Resiko yang digunakan dalam Proyek ini**

Mastersystem akan mengidentifikasi dan kuantifikasi list risk event. Hal ini akan dibahas ketika meeting proyek mingguan. Bahasan yang dilakukan adalah sesi brainstorming untuk mengevaluasi impact atau probabilitas dari event risk spesifik, mengkuantifikasi impact dari setiap risk event dan merancang strategi mitigasi resiko. Hasil dari meeting akan didokumentasikan dalam Risk Register. Project Manager Mastersystem akan menentukan baseline dari Risk Register pada akhir proses untuk direview dan disetujui oleh PM Bank BRI.

Hasil dari proses ini akan dimasukkan ke dalam Project Schedule dalam bentuk task tambahan atau alternatif. Ini akan diajukan sebagai Change Request, dan membutuhkan approval serupa Change Request.

### **4.3 Infrastruktur**

Sesuai penjelasan pada TOR dan rapat penjelasan, infrastruktur akan disiapkan oleh tim Bank BRI.

## **4.4 Quality Assurance**

### **4.4.1 Kebutuhan Umum dan Teknis Quality Assurance**

Menyetujui Quality Management Plan, termasuk semua aktivitas yang menjelaskan keseluruhan proses QA sesuai dengan Project Plan. Memastikan bahwa solusi yang dibuat sesuai dengan kontrak yang disetujuk, memenuhi atau melebihi standar kualitas yang telah didiskusikan dan memenuhi SDLC.

Proses QA termasuk beberapa aktivitas seperti:

1. Peer review

Review rekan adalah cara untuk menghilangkan cacat pada produk lebih awal dan efisien. Dalam proses ini, rekan kerja dari pembuat produk memeriksa produk untuk mengidentifikasi cacat pada produk dan memperbaiki kekurangan.

## 2. Formal Technical Review

Review Formal Teknis dilakukan untuk deliverable yang dijadwalkan pada project schedule. Review ini memeriksa bahwa deliverable benar-benar dihasilkan di setiap fase. Ini termasuk mengevaluasi baik proses untuk menghasilkan deliverable dan juga kualitas dari deliverable tersebut. Review ini adalah bagian terintegrasi pada proses QA. Review ini adalah tanggung jawab dari Project Manager Mastersystem, Project Manager Teknis Bank BRI, dan Project Manager Bisnis Bank BRI.

## 3. Requirement Traceability Matrix

Matrix ini digunakan untuk memonitor pemenuhan kebutuhan/requirement di berbagai stage pada project, seperti software design, coding, unit testing, dan acceptance.

### **4.4.2 Tindakan Mitigasi yang Umum**

Quality Record digunakan untuk memastikan ada pendekatan terkendali untuk project ini. Ini memastikan repositori lengkap dari deliverable kunci dapat diakses oleh auditor ataupun PMO jika dibutuhkan.

Quality Assurance dalam scope Mastersystem di project ini berarti proses untuk memfasilitasi kendali dari progress dan performa proyek.

List dokumen berikut merepresentasikan quality record dari proyek. Project Manager seharusnya menjaga beberapa dokumen terkait kualitas berikut:

- Project Progress Reports/Project Status Report
- Kontrak antara Mastersystem dan Bank BRI
- Project Schedule
- Change Requests Forms dan Change Control Log
- Sign-off forms signed off Client.
- Meeting Minutes
- Issue Log
- Risk Register

## 4.5 Project Change Control Plan

### 4.5.1 Penjelasan Awal

Project change control memastikan proyek dapat merespon ke perubahan kebutuhan sementara tetap menjaga kendali dan delivery yang efisien. Perubahan pada project scope, kebutuhan/requirement, termin kommersial akan secara formal diatur sesuai rencana ini. Change Order yang diapprove melalui proses ini akan mengubah Statement of Work dan juga project plan yang disetujui bersama.

Proses ini memastikan komunikasi yang jelas dan pengambilan keputusan, dan semua stakeholder mengerti komitmen saat ini dan status dari perubahan yang diajukan. Project Manager Mastersystem dan Project Manager Bank BRI akan secara bersama mengatur proses change control ini.

### 4.5.2 Tanggung Jawab Change Control

Individual yang diassign untuk peran ini akan diidentifikasi di halaman tanda tangan. Mereka harus memiliki otorisasi penuh untuk menjalankan tugas yang diidentifikasi. Satu individual dapat melakukan lebih dari satu peran.

- Project Manager Mastersystem
  - Semua proses perubahan dikomunikasikan hanya melalui PM Mastersystem.
  - Semua komunikasi tentang status change order adalah dengan PM.
  - PM Mastersystem akan koordinasi semua aktivitas internal Mastersystem untuk mengevaluasi dan merespon permintaan perubahan dari Bank BRI.
- Project Manager Bank BRI
  - Semua request change order akan disubmit oleh PM Bank BRI atau kontak change control yang disetujui bersama.
  - Semua komunikasi tentang status change order adalah dengan PM Bank BRI atau kontak change control tersebut.
  - PM Bank BRI akan koordinasi semua aktivitas internal Bank BRI dan untuk mereview untuk mendukung proses ini.
- Mastersystem Approver
  - PIC ini terotorisasi untuk meng-approve change order.
  - PIC ini mungkin mendeklegasikan sebagian approval ke PM Mastersystem untuk case by case basis.
- Bank BRI Approver
  - PIC ini terotorisasi untuk meng-approve perubahan proyek ini terkait teknis, pricing, dan juga term pada kontrak.
  - PIC ini akan melakukan koordinasi internal approval untuk mendukung proses ini.
- Project Steering Committee (PSC)
  - PSC akan memberikan solusi issue yang dieskalasi baik oleh BRI ataupun Mastersystem jika kedua belah pihak tidak setuju change order yang diajukan.
  - PSC akan meng-approve term dan change request dengan impact melebihi otoritas dari approver normal Bank BRI dan Mastersystem.

#### 4.5.3 Change Request Process Flow

##### Generating Change Request

Ketika kebutuhan untuk perubahan Agreement diidentifikasi, change request akan dibuat berisi beberapa detail berikut:

- Title dari perubahan.
- Originator dan tanggal request atau rekomendasi perubahan.
- Alasan perubahan.
- Detail perubahan, termasuk spesifikasi sistem.

Jika Change Request terkait kustomisasi fungsional sistem tambahan, interface baru atau perubahan pada interface yang sudah ada, report baru atau perubahan pada report yang sudah ada, Bank BRI perlu mendeskripsikan detail dari perubahan yang dibutuhkan. Detail tersebut harus berisi:

Untuk kustomisasi fungsional	Untuk interface baru atau perubahan interface yang sudah ada	Untuk report baru atau perubahan report yang sudah ada
<ul style="list-style-type: none"> <li>▪ Input data</li> <li>▪ Processing, workflow</li> <li>▪ Formulas for calculation</li> <li>▪ Required outcome</li> </ul>	<ul style="list-style-type: none"> <li>▪ Batch/on-line</li> <li>▪ Interface structure</li> <li>▪ File fields business description</li> <li>▪ Fields attributes</li> <li>▪ Frequency</li> <li>▪ File sample</li> </ul>	<ul style="list-style-type: none"> <li>▪ On-line/upon-request or batch/EOD report</li> <li>▪ Report structure</li> <li>▪ Reports fields business description</li> <li>▪ Fields attributes</li> <li>▪ Report layout</li> </ul>
<ul style="list-style-type: none"> <li>• Proposed implementation date or revised project plan, as applicable.</li> <li>• The price and payment schedule of the proposal.</li> </ul>		

##### Merespon Change Request

Dalam merespon change request dari Bank BRI atau termasuk dalam change request yang diajukan oleh Mastersystem, maka Mastersystem akan mendetailkan:

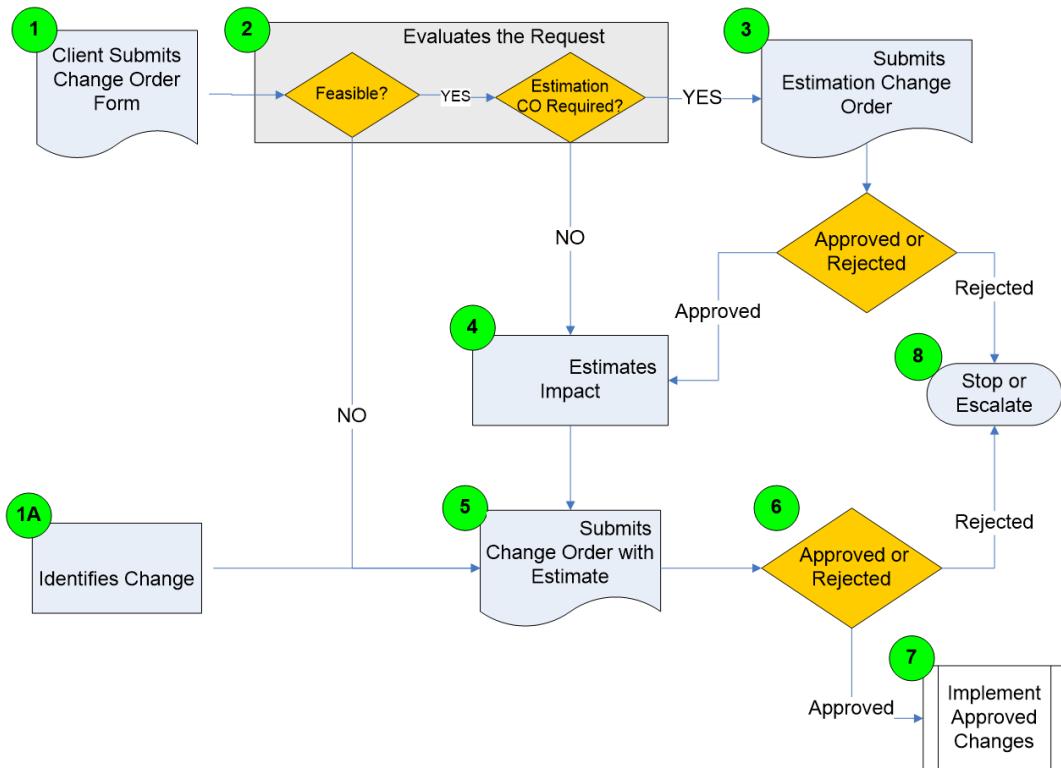
- Harga dan charge akibat perubahan, jika ada.
- Detail dari service, service maintenance, kebutuhan spesifikasi produk/sistem, effort yang dibutuhkan, dan assessment impact sebagaimana berlaku.
- Detail environment dan lokasi sebagaimana berlaku.
- Implikasi dari mekanisme support post go live.
- Timetable implementasi.

Di mana memungkinkan, Mastersystem akan membuat proposal change request ke Bank BRI dalam waktu 20 hari setelah menerima request. Bank BRI kemudian akan membuat respon formal terhadap change proposal dari Mastersystem dalam 5 hari.

Setelah menerima respon Mastersystem, Bank BRI dapat memilih di antara:

- Menerima proposal, dan kemudian mengamandemen kontrak sesuai dengan respon Mastersystem.
- Atau menarik request yang diajukan, di mana SOW akan berlanjut tanpa perubahan.

Detail dari change control process flow adalah sebagai berikut.



### **Penjelasan diagram**

#### **1. Bank BRI submit Change Order Form**

PM Bank BRI dapat submit change request kapan pun, atau di aspek manapun pada proyek. Change order harus didokumentasikan dalam bentuk Change Order Form. PM Mastersystem akan mengupdate form tuntuk konfirmasi tanggal penerimaan secara tertulis, dan memberikan jawaban perkiraan waktu yang dibutuhkan untuk merespon.

#### **2. Mastersystem mengidentifikasi perubahan/Change yang direquest**

Mastersystem akan mengeluarkan Change Order Form, yang sudah termasuk perkiraan impact terhadap project schedule, biaya, scope ataupun termin lainnya. PM Bank BRI akan melakukan konfirmasi penerimaan secara tertulis dan akan merespon request dalam waktu 5 hari.

#### **3. Mastersystem mengevaluasi request**

Mastersystem akan mengevaluasi change order dari Bank BRI untuk menentukan disposisi awal. Respon akan dilakukan dalam 5 hari, atau sesuai yang dispesifikasikan ketika penerimaan Change Order. Ada 3 kemungkinan hasil dari evaluasi ini:

- a. Mastersystem menerima request untuk estimasi. Mastersystem akan melanjutkan dengan evaluasi dan estimasi impact (Step 4).
- b. Mastersystem menolak request sebagai tidak mungkin mempertimbangkan proyek saat ini. Mastersystem akan mengupdate Change Order Form dengan penjelasan dan mengembalikannya ke PM Bank BRI (Step 5).
- c. Mastersystem merespon dengan Estimation Change Order, karena evaluasi menyeluruh mungkin membutuhkan waktu atau biaya lebih daripada yang dimungkinkan dalam rencana proyek saat ini. Mastersystem akan mengupdate Change Order form dengan penjelasan (Step 3).

#### **4. Mastersystem submit respon “Estimation Change Order”**

Step ini dijalankan hanya jika di Step 2, Mastersystem menentukan bahwa analisa full dari change order tidak dapat diakomodasi dalam project schedule atau budget proyek saat ini. PM Mastersystem akan merespon dengan “Estimation Change Order” menunda request untuk dianalisa lebih lanjut dan menspesifikasikan waktu atau cost tambahan untuk analisa menyeluruh. Jika Bank BRI menerima, Mastersystem akan mulai bekerja untuk mengestimasi impact (Step 4). Jika Bank BRI menolak, request awal akan “ditutup” tanpa aktivitas lebih lanjut.

#### **5. Mastersystem mengestimasi impact**

Mastersystem akan menganalisa request dan menyiapkan respon. Analisa ini akan menangani impact kebutuhan, service, deliverable, biaya, project schedule, dan term teknis atau komersial lainnya.

#### **6. Mastersystem submit Change Order dengan Estimate/Periraan**

PM Mastersystem akan secara formal submit Change Order Form yang telah diupdate ke PM Bank BRI. Respon ini adalah proposal untuk mengimplementasikan perubahan. Dokumen ini akan mendeskripsikan perubahan, impact, dan tambahan biaya ataupun waktu yang dibutuhkan untuk implementasi.

#### **7. Client dan Mastersystem Approve atau Reject Change Order**

PM Bank BRI akan mengkoordinasikan internal review dan approval yang dibutuhkan. Secara default, evaluasi dan respon akan dilakukan dalam 5 hari kerja atau sesuai persetujuan kedua belah pihak. Change order dapat diapprove oleh kedua belah pihak ditandai dengan menandatangani Change Order Response form. Jika kedua belah pihak tidak mencapai kesepakatan, atau jika tidak ada respon sesuai waktu yang ditentukan, Change Order akan ditutup dan diabaikan.

#### **8. Mengimplementasikan Approved Changes**

Setelah diterima, para PM akan memodifikasi project plan dan mengimplementasikan perubahan. Detail step akan tergantung sifat perubahan, namun biasanya termasuk:

- Mengkomunikasikan keputusan yang diapprove.
- Trigger proses project Configuration Management.
- Mengupdate project schedule.
- Mengupdate requirements document dan traceability matrix
- Mengupdate test plan dan test script.

#### **9. Stop atau Eskalasi**

Jika Change Order ditolak (oleh Mastersystem atau oleh BRI), maka status ditandai sebagai selesai dan tidak ada pekerjaan lanjutan. Jika kedua pihak tidak berhasil mencapai kesepakatan, meeting Project Steering Committee dapat diajukan oleh salah satu pihak untuk mengeskalasi request keputusan/decision.

#### **4.5.4 Rangkuman**

Proses change control akan digunakan untuk mengatur semua perubahan terhadap project baseline yang sudah disetujui bersama. Ini memastikan transparansi dan persetujuan kedua belah pihak terhadap ekspektasi proyek saat ini. Ini juga akan membantu PM Mastersystem dan Bank BRI mengakomodasi perubahan sementara memastikan delivery proyek yang fokus dan efisien.

#### **4.5.5 Change Control History Register**

Histori log dari status semua change request yang diproses dalam proyek ini, progress/penolakan/eksekusi/implementasi dan impact nya akan dikelola oleh Project manager dan disebut sebagai Change Control Register, bisa berbentuk Excel ataupun format serupa, dapat dilaporkan dalam Progress Monitoring Report.

---

## 5 ASUMSI DAN OUT OF SCOPE

---

Berikut adalah rincian asumsi atau out of scope dalam project ini:

### 5.1 Out of Scope

- Integrasi ke channel Bank BRI atau system lain yang tidak berhubungan dengan Big Data.
- OS hardening dan penetration test.
- Pengembangan apapun di system existing Bank BRI.
- Migrasi surrounding system akan dilakukan oleh BRI.
- Tambahan use case.
- Penambahan scope selain yang telah disetujui di Proposal.
- Tambahan coverage produk selain yang telah disetujui di Proposal.
- Tambahan service selain yang telah disetujui di Proposal.

## 6 JADWAL PELAKSANAAN

Jadwal proyek ini adalah perkiraan dan perlu diskusi requirement lebih lanjut untuk finalisasi.

## 6.1 Implementasi Datalake DC

No	Task	Week	PIC	W 1	W 2	W 3	W 4	W 5	W 6	W 7	W 8	W 9	W 10	W 11	W 12	W 13	W 14	W 15	W 16	W 17	W 18	W 19	W 20	W 21	W 22	W 23	W 24	W 25	W 26	W 27	W 28	W 29	W 30	W 31	W 32	W 33	W 34	W 35
	Replace 44 Nodes New																																					
5	Migration for Master,Edge, and Utility	5	MSI																																			
6	Support surrounding application migration	3	MSI																																			
7	Installation and Configure Ozone	1	CLOUDERA /MSI																																			
8	Pilot Migration from HDFS to Ozone	1	CLOUDERA /MSI																																			
9	Post Cluster Refreshment support	1	CLOUDERA /MSI																																			
10	Set up prerequisites on 22 node (Data Services)	1	CLOUDERA /MSI																																			
11	Install CDP Data Services on 22 node	1	CLOUDERA /MSI																																			
12	Migrate CDSW to Cloudera Machine Learning	2	CLOUDERA /MSI																																			
13	Functional Test	1	CLOUDERA /MSI																																			
14	SIT	2	BRI/MSI																																			
15	UAT	2	BRI/MSI																																			
16	Documentation	5	MSI																																			
17	Handover - MSI	1	MSI																																			

No	Task	Week	PIC	W 1	W 2	W 3	W 4	W 5	W 6	W 7	W 8	W 9	W 10	W 11	W 12	W 13	W 14	W 15	W 16	W 17	W 18	W 19	W 20	W 21	W 22	W 23	W 24	W 25	W 26	W 27	W 28	W 29	W 30	W 31	W 32	W 33	W 34	W 35
18	Handover - CLOUDERA Review - Performance Test	1	CLOUDERA																						■													
19	Go Live	1																							■													
20	Warranty 3 months	12	MSI																						■	■	■	■	■	■	■	■	■	■	■			
	Total	51																																				

## 6.2 Implementasi HBase DC

No	Task	Week	PIC	W 1	W 2	W 3	W 4	W 5	W 6	W 7	W 8	W 9	W 10	W 11	W 12	W 13	W 14	W 15	W 16	W 17	W 18	W 19	W 20	W 21	W 22	W 23	W 24	W 25	W 26	W 27	W 28	W 29	W 30	W 31	
1	Setup prerequisites on 20 Node	1	MSI																																
2	Install CDP	1	MSI																																
3	Configure Security	2	MSI																																
4	Replicate data from Existing HBASE-DC	3	MSI																																
5	Migrate workload from existing HBASE-DC	3	MSI																																
6	Reconfigure Replication DRC	1	MSI																																
7	Setup Prerequisite Nifi 3 Node Existing + 4 New Nodes (DC)	1	MSI																																
8	Add Nifi Node to Cloudera Manager	1	MSI																																
9	Configure Security Nifi Node	1	MSI																																
10	Migrate Job from Existing Nifi	3	MSI																																
11	Setup Prerequisite Nifi 4 New Nodes (DR)	1	MSI																																
12	Add Nifi Node to Cloudera Manager	1	MSI																																
13	Configure Security Nifi Node	1	MSI																																

No	Task	Week	PIC	W 1	W 2	W 3	W 4	W 5	W 6	W 7	W 8	W 9	W 10	W 11	W 12	W 13	W 14	W 15	W 16	W 17	W 18	W 19	W 20	W 21	W 22	W 23	W 24	W 25	W 26	W 27	W 28	W 29	W 30	W 31
14	Documentation	5	MSI																															
15	Functional Test Review - Performance Test	1	MSI																															
16	SIT	2	BRI/MSI																															
17	UAT	2	BRI/MSI																															
18	Handover	1	MSI																															
19	Go Live	1																																
20	Warranty 3 months	12	MSI																															
	Total	44																																

## LAMPIRAN 1: SPESIFIKASI TEKNIS – KEBUTUHAN FUNGSIONAL

No	Requirement	FS/MM/LM/NV/NS	M/O	Vendor Response and Supporting Document
1	Melakukan instalasi license pada Cluster Server yang terdapat di BRI On Premises	FS	M	Hal ini akan kami sediakan
2	Merancang perencanaan implementasi dengan memperhatikan risiko dan dampak kepada surrounding platform Cloudera.	FS	M	Hal ini akan kami sediakan
3	Melakukan proses migrasi server untuk 44 node Cluster Datalake DC BRI (Old Server) ke server Baremetal (Refreshment).	FS	M	Hal ini akan kami sediakan
4	Melakukan proses ekspansi dengan penambahan worker node sebanyak 26 server untuk Cluster Datalake DC.	FS	M	Hal ini akan kami sediakan
5	Melakukan proses migrasi dari infrastruktur server Openstack ke server Baremetal untuk Cluster Hbase DC.	FS	M	Hal ini akan kami sediakan
6	Melakukan penyesuaian konfigurasi Cluster Hbase DC setelah migrasi dengan tuning configuration untuk performance Cluster Hbase DC.	FS	M	Hal ini akan kami sediakan
7	Melakukan proses ekspansi dengan penambahan node untuk Services NIFI.	FS	M	Hal ini akan kami sediakan
8	Melakukan proses pembaharuan lisensi untuk Cluster Datalake ODC, Hbase DC dan ODC.	FS	M	Hal ini akan kami sediakan
9	Melakukan proses instalasi, integrasi, dan konfigurasi Cloudera Data Services pada Cluster Datalake DC.	FS	M	Hal ini akan kami sediakan
10	Mengimplementasikan seluruh komponen platform Cloudera CDP Private Cloud, seperti Cloudera Data Warehouse, Cloudera Machine Learning, Cloudera Management Console, Cloudera Streaming Analytics, Cloudera Data Engineering, Cloudera Data Services, dll.	FS	M	Hal ini akan kami sediakan
11	Melakukan proses instalasi, integrasi, dan konfigurasi Cloudera Ozone pada Cluster Datalake DC.	FS	M	Hal ini akan kami sediakan
12	Melakukan pengujian secara best practice terkait dengan migrasi dari CDSW ke Cloudera Machine Learning (Script, Job, Model API, dll).	FS	M	Hal ini akan kami sediakan
13	Melakukan pengujian secara best practice terkait dengan penggunaan Ozone untuk archiving data dari Cluster Datalake DC.	FS	M	Hal ini akan kami sediakan
14	Melakukan functional dan non-functional test terhadap platform Cloudera yang akan diimplementasikan, termasuk migrasi, ekspansi, dan instalasi baru.	FS	M	Hal ini akan kami sediakan
15	Melakukan penambahan resource dan node server untuk Platform NIFI pada Cluster Hbase DC.	FS	M	Hal ini akan kami sediakan

No	Requirement	FS/MM/LM/NV/NS	M/O	Vendor Response and Supporting Document
16	Melakukan Implementasi Monitoring Cluster dengan menggunakan Platform External milik BRI, serta merancang tampilan monitoring pada Cloudera Manager sesuai dengan kebutuhan BRI.	FS	M	Hal ini akan kami sediakan

## LAMPIRAN 2: SPESIFIKASI TEKNIS – KEBUTUHAN TEKNIKAL DAN SECURITY

No	Requirement	FS/MM/ LM/NV/ NS	Solution Availability(3M /6M/1Y/M1)	M/O	Vendor Response and Supporting Document
1	Memiliki security layer untuk melakukan enkripsi seperti SSL dan Kerberos	FS	-	M	Hal ini akan kami sediakan
2	Mampu mengeluarkan alert untuk kondisi tertentu	FS	-	M	Hal ini akan kami sediakan
3	Memiliki kemampuan untuk berjalan pada architecture distributed atau single node	FS	-	M	Hal ini akan kami sediakan
4	Memiliki user dan roles access management	FS	-	M	Hal ini akan kami sediakan
5	Memiliki kemampuan data management dalam mengelola data sensitif	FS	-	M	Hal ini akan kami sediakan

## LAMPIRAN 3: CV TENAGA AHLI

### Project Manager: Nurma Sumar Sidik

<b>Personal</b>	Place/Date of Birth : Padang, 1 Maret 1996 Address : Gandaria Utara, Jakarta Formal Education : Bachelor of Science, IPB Phone/Email : 089617559541 / nurma.sumar@mastersystem.co.id
<b>Working Experience</b>	April 2020 – Now <b>Project Manager</b> PT. Mastersystem Infotama
<b>Project Experience</b>	<ol style="list-style-type: none"><li>1. Implementasi Big Data DRC PT. Bank Rakyat Indonesia (Persero), Tbk</li><li>2. Pengadaan Perangkat Server, Network dan pendukung lain untuk DRC BRI</li><li>3. Pengadaan Perangkat Server dan Network untuk ODC BRI</li><li>4. Implementasi Eliminasi SPOF BRI</li><li>5. Implementasi Switch Replikasi di DC dan DRC BRI</li><li>6. Implementasi Fortigate, Bluecoat dan Solarwinds di DC IIF</li><li>7. Relokasi DC PT. BUMA</li><li>8. Network Refreshment di DC Komatsu Undercarriage Indonesia</li><li>9. Pengadaan Infrastruktur dan HSM di Bank Raya</li><li>10. Pengadaan Perangkat Jaringan Transformasi Jakarta Smart City di Pemda Jakarta</li><li>11. Konsultan Proyek Relokasi Data Center sebagai Arsoitek Jaringan di IIF</li></ol>
<b>Professional Training</b>	<ol style="list-style-type: none"><li>1. Project Management Training (PMBOK 6)</li><li>2. Agile Project Management Training</li></ol>

## Engineer 1: Hendri



### Personal

Place/Date of Birth : Jakarta, 1996 13<sup>th</sup> Apr  
Address : Jl. Hanoman I No. 21  
Cengkareng, Jakarta Barat , 11740  
Formal Education : Bina Nusantara University – Computer System

### Work Experience

2022 – Now PT. Mastersystem Infotama  
**Data Engineer**

### Project Experience

1. Cloudera Upscale IOT Cluster at Astra Honda Motor (2022)

Job Role: Cloudera Administrator

- Implementation Cloudera Data Platform Private Cloud
- Implementation Cloudera Flow Management
- Implementation Cloudera Streaming Analytics
- Administer Cloudera Cluster

2. Cloudera Rekening Koran at BRI (2022)

Job Role: Cloudera Administrator

- Implementation Cloudera Data Platform Private Cloud

- Implementation Cloudera Flow Management (Apache Nifi)
- Building Streaming and Batch Pipeline for Operational Database HBASE
- Administer Cloudera Cluster

3. Big Data Analytics at Bank BCA (2021)

Job Role: Cloudera Administrator Support

- Troubleshooting Error on Cloudera Cluster
- Administer Cloudera Cluster

**Training Experience**

1. (Ondemand Cloudera) Cloudera Administrator Training
2. (Ondemand Cloudera) Cloudera Data Flow
3. (Ondemand Cloudera) Cloudera Essential for CDP
4. (Ondemand Cloudera) Apache Kafka Basic
5. (Ondemand Cloudera) Streaming Processing, Management and Analytics with CDF

**Certification**

1. Cloudera Data Platform Private Cloud Administrator (2023)

*Credly*

OFFICIAL TRANSCRIPT - ISSUED ON 16 FEBRUARY 2023

Hendri .

ID#: cb8cb482-7fcd-4872-ba95-44e87ac59de5

Birth Date: 13 April 1996

[hendri@datalabs.id](mailto:hendri@datalabs.id)

[www.credly.com/users/hendri.396a15de](http://www.credly.com/users/hendri.396a15de)



**CDP Certified Administrator - Private Cloud Base**

Issued by: Cloudera

Issued to: Hendri Hendri

Issued on: 11 February 2023

Expires on: 10 February 2025

**Description**

This badge indicates the certificant has passed the CDP Administrator - Private Cloud Base exam. This exam test the critical Cloudera Data Platform (CDP) skills and knowledge required to succeed as an administrator. Skills and knowledge that must be demonstrated include: Installation & Initial Configuration, Connecting & Managing Data Sources, Managing Users, Monitoring & Troubleshooting, and Managing Data Security & Governance.

END OF TRANSCRIPT

Page 1 of 1

## Engineer 2: Aris Tito Sugiharto

### Personal

Place/Date of Birth : Kuningan, 1996 06<sup>th</sup> Mar

Address : Jl. Depkes 2 No. 06

Bekasi, 17412

Formal Education : Gunadarma University – Computer System

### Work Experience

2020 – Now PT. Mastersystem Infotama

**Data Engineer**

### Project Experience

1. Big Data Analytics at Bank BRI (2021)

Job Role: Cloudera Administrator

- Install Cloudera Big Data with TLS and Kerberos for DRC Cluster
- Install Cloudera Data Science Workbench
- Administer Cloudera Cluster
- Install Cloudera Data Platform Environment DC dan DRC for Operational Database

2. Big Data Analytics at Bank BCA (2021)

Job Role: Cloudera Administrator

- Install Cloudera Big Data Cloudera Data Platform for Datalake without TLS and Kerberos for DRC Cluster
- Setup Replication to DRC Cluster
- Administer Cloudera Cluster

3. Cloudera Big Data at MACF (2021)

Job Role: Data Engineer & Cloudera Administrator

- Install Cloudera Big Data with TLS and Kerberos for Development Cluster and Production Cluster
- Administer Cloudera Cluster
- Design Loading Data for Initial and Daily Load using Cloudera Flow Management

#### 4. Cloudera Big Data Refresh at BNI (2022)

Job Role: Cloudera Administrator

- Install Cloudera Data Science Workbench
- Upgrade CDH to CDP
- Assisst Installation for CDP Datalake DC and DRC
- Administer Cloudera Cluster

#### 5. Cloudera Upgrade at BCA (2022)

Job Role: Cloudera Administrator

- Upgrade CDH to CDP for Datalake Cluster Production, Development and DRC
- Administer Cloudera Cluster

#### 6. Cloudera Upgrade at BRI (2022)

Job Role: Cloudera Administrator

- Upgrade CDH to CDP Datalake Cluster Production and DRC
- Upgrade Cloudera Data Science Workbench
- Administer Cloudera Cluster

#### 7. Cloudera Upgrade at Mandiri (2022)

Job Role: Cloudera Administrator

- Upgrade CDH to CDP
- Administer Cloudera Cluster

#### 8. Cloudera Upscale IOT Cluster at Astra Honda Motor (2022)

Job Role: Cloudera Administrator

- Implementation Cloudera Data Platform Private Cloud
- Implementation Cloudera Flow Management

- Implementation Cloudera Streaming Analytics
  - Administer Cloudera Cluster
9. Cloudera Data Science Migration at BCA (2022)
- Job Role: Cloudera Administrator
- Upgrade Cloudera Data Science Workbench
  - Migrating Cloudera Data Science Workbench Master to New Master
  - Administer Cloudera Cluster
- Training Experience**
1. (Ondemand Cloudera) Cloudera Administrator Training (2021)
  2. (Ondemand Cloudera) Cloudera Data Flow (2021)
  3. (Coursera) Google Cloud Platform Fundamentals: Core Infrastructure
  4. (Coursera) Elastic Google Cloud Infrastructure: Scaling and Automation
  5. (Coursera) Essential Google Cloud Infrastructure: Core Services
  6. (Coursera) Essential Google Cloud Infrastructure: Foundation
  7. (Ondemand Cloudera) Cloudera Essential for CDP
  8. (Ondemand Cloudera) Apache Kafka Basic
  9. (Ondemand Cloudera) Streaming Processing, Management and Analytics with CDF
- Certification**
1. CDP Certified Generalist (2022)
  2. CDP Certified CDP Administrator - Private Cloud Base (2022)

*Credly*

OFFICIAL TRANSCRIPT - ISSUED ON 18 OCTOBER 2022

ARIS TITO SUGIHARTO

ID#: c3c8ea60-b93a-44c0-bb78-d2553adce353

Birth Date: 6 March 1996

aristitosugharto@gmail.com

[www.credly.com/users/aristitosugharto](http://www.credly.com/users/aristitosugharto)



**CDP Certified Generalist**

Issued by: Cloudera

Issued to: ARIS TITO SUGIHARTO

Issued on: 4 June 2022

Expires on: 3 June 2024

**Description**

This badge indicates the certificant has broad, general knowledge of the Cloudera Data Platform (CDP). CDP combines the best of Hortonworks' and Cloudera's technologies to deliver the industry's first enterprise data cloud. CDP delivers powerful self-service analytics across hybrid and multi-cloud environments, along with sophisticated and granular security and governance policies that IT and data leaders demand.



**CDP Certified Administrator - Private Cloud Base**

Issued by: Cloudera

Issued to: ARIS TITO SUGIHARTO

Issued on: 12 June 2022

Expires on: 11 June 2024

**Description**

This badge indicates the certificant has passed the CDP Administrator - Private Cloud Base exam. This exam test the critical Cloudera Data Platform (CDP) skills and knowledge required to succeed as an administrator. Skills and knowledge that must be demonstrated include: Installation & Initial Configuration, Connecting & Managing Data Sources, Managing Users, Monitoring & Troubleshooting, and Managing Data Security & Governance.

END OF TRANSCRIPT

Page 1 of 1

## Technical Writer: Adam Sulthon



### Personal

Place/Date of Birth : Jakarta, 1998, 17<sup>th</sup> Feb  
Address : Jl. Pondok Kelapa Selatan 3C, No.43  
Jakarta Timur, 13450  
Formal Education : Gunadarma University – Computer System

### Work Experience

2022 – now : PT. Mastersystem Infotama  
**Data Engineer**

### Project Experience

1. Direktorat Jendral Pajak Indonesia Bigdata Administrator (2022)  
Job Role : Cloudera Administrator
  - Maintenance Cloudera
  - Troubleshooting Cloudera Services
  - Preventive and Corrective Maintenance
  
2. Mandiri Bigdata Administrator (2022)  
Job Role : Cloudera Administrator
  - Maintenance Cloudera
  - Troubleshooting Cloudera Services
  - Preventive and Corrective Maintenance

3. BRI Bigdata Administrator (2022)

Job Role : Cloudera Administrator

- Maintenance Cloudera
- Troubleshooting Cloudera Services
- Preventive dan Corrective Maintenance

**Training Experience**

1. (Ondemand Cloudera) Cloudera Data Platform
2. (Ondemand Cloudera) Cloudera Data Flow
3. (Ondemand Cloudera) Cloudera Data Engineering
4. (Ondemand Cloudera) Cloudera Data Werehouse
5. (Ondemand Cloudera) Cloudera Machine Learning

**Certification**

-

## LAMPIRAN 4: URAIAN PEKERJAAN SEJENIS

No.	Nama Perusahaan	Judul Proyek	SPK/ Kontrak / PO	Tanggal
1	PT. Bank Mandiri (Persero )	Pengadaan Big Data Analytics	SPK (OPS.SPC/SS1.1830/2016)	19 Desember 2016
2	PT. Sigma Cipta Caraka (untuk PT. Bank Rakyat Indonesia ( Persero))	Perjanjian Pengadaan Software Instalasi dan Maintenance Big Data Antara PT. Sigma Cipta Caraka dan PT. Mastersystem Infotama	Kontrak (052-PRC/SCC/MAIN/18)	28 Februari 2018
3	PT. Telekomunikasi Selular	Kontrak Payung Pengadaan Cloudera Enterprise License Platform Antara PT. Telekomunikasi Selular dan PT. Mastersystem Infotama	Kontrak (MPA003-1901-8172)	23 Januari 2019
4	PT. Telekomunikasi Selular	Kontrak Pengadaan Jasa Professional Cloudera Untuk Proyek Hadoop Expansion Antara PT. Telekomunikasi Selular dan PT. Mastersystem Infotama	Kontrak (HOC181435)	26 Februari 2019
5	PT Metranet (untuk PT. Bank Rakyat Indonesia (Persero))	Pengadaan Peningkatan Kapasitas Cluster Cloudera Big Data	PO (1010201900000035PR)	15 Maret 2019
6	PT Metranet (untuk PT. Bank Rakyat Indonesia (Persero))	Pengadaan License Cloudera untuk Kebutuhan Enterprise Data Warehouse (EDW)	PO (10102019000000138PR)	20 Maret 2019
7	PT. Bank Mandiri (Persero)	Pengadaan License Cloudera Hadoop untuk Project Big Data Analytics Expansion dan Project Fraud Risk Management System	SPK (TIB.SPC/SS1.1741/2019)	7 Agustus 2019
8	PT. Telekomunikasi Selular	Kontrak Pengadaan Jasa Professional Cloudera Phase II 2019 - 2020 Antara PT. Telekomunikasi Selular dan PT. Mastersystem Infotama	Kontrak (HOC191128)	4 Desember 2019
9	PT. Telekomunikasi Selular	Kontrak Induk Pengadaan SVOC Implementation With Cloudera Data Science Workbench Antara PT. Telekomunikasi Selular dan PT. Mastersystem Infotama	Kontrak (MPA003-1912-0244)	20 Desember 2019
10	PT. Bank Rakyat Indonesia (Persero)	Pengadaan Cluster Big Data untuk DRC	SPK (B.178.P-PBJ/PIT/02/2020)	12 Februari 2020
11	PT. Bank Mandiri (Persero)	Pengadaan Jasa Outcourcing Borongan Resident Solution Architect Big Data	Kontrak (KES.EDM/DOP/SPK/SPP.0001/2020)	11 Juni 2020
12	PT. Bank Mandiri (Persero)	Pengadaan Jasa Pemeliharaan Aplikasi Big Data Analytics	SPK (TBS.SPC/SS1.1052/2020)	28 Juli 2020
13	PT. Telekomunikasi Selular	Kontrak Pengadaan Jasa Professional Cloudera 2020 - 2021 Antara PT. Telekomunikasi Selular dan PT. Mastersystem Infotama	Kontrak (M100001643)	8 Agustus 2020
14	PT. Telekomunikasi Selular	Kontrak Payung Pengadaan Cloudera Enterprise License for Fit to Purpose Data Pipeline Antara PT. Telekomunikasi Selular dan PT. Mastersystem Infotama	Kontrak ( M100002158)	24 November 2020
16	PT. Bank Negara Indonesia (Persero) , Tbk	Surat Perintah Kerja Pengadaan Maintenance Big Data	SPK (PFA/3.2/168)	5 Mei 2021
17	PT. Telekomunikasi Selular	Purchase Order untuk Cloudera Professional Services Phase IV	PO (4200020933)	13 Juli 2021
18	PT. Bank Negara Indonesia (Persero) , Tbk	Penambahan Lisensi Software dan Peningkatan/ Upgrade versi Big Data BNI	SPK (PFA/3.2/320)	26 Agustus 2021
19	PT. Telekomunikasi Selular	Kontrak Pengadaan Cloudera Professional Services for Support Fit to Purpose Project	Kontrak (M100003509)	9 November 2021

No.	Nama Perusahaan	Judul Proyek	SPK/ Kontrak / PO	Tanggal
20	PT. Telekomunikasi Selular	Amandemen Pertama Atas Kontrak Payung Pengadaan Cloudera Enterprise License Platform Antara PT. Telekomunikasi Selular dan PT. Mastersystem Infotama	Kontrak (M100100957)	26 Januari 2022
21	PT. Bank Rakyat Indonesia (Persero)	Pengadaan Implementasi Reengineering Sistem Rekening Koran	SPK (4200001192-PLO/IGP/PTT/04/2022)	26 April 2022
22	PT. Bank Mandiri (Persero)	Pengadaan Subscription License Cloudera- Aplikasi Big Data Analytics	SPK (TIB.SPC/ITSS.166/2023)	08 Februari 2023
23	PT. Bank Mandiri (Persero)	Pengadaan Implementasi Big Data Cluster Production	SPK (TIB.SPC/ITSS.354/2023)	27 Maret 2023

## LAMPIRAN 5: MACHINE LEARNING ON CDP OVERVIEW

### Our Goal: SCALABLE ENTERPRISE MACHINE LEARNING



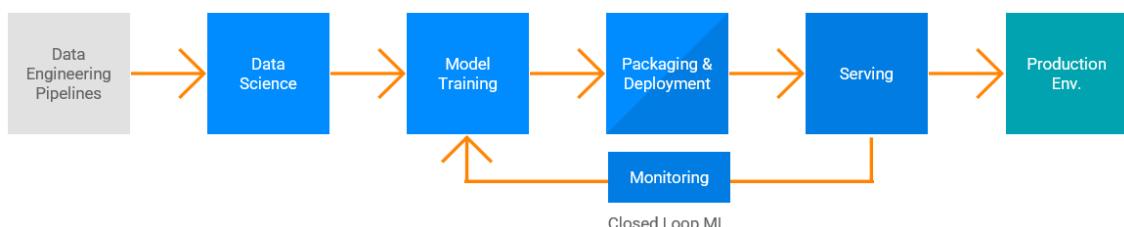
FROM THE LAB...

TO THE FACTORY...

TO THE BUSINESS

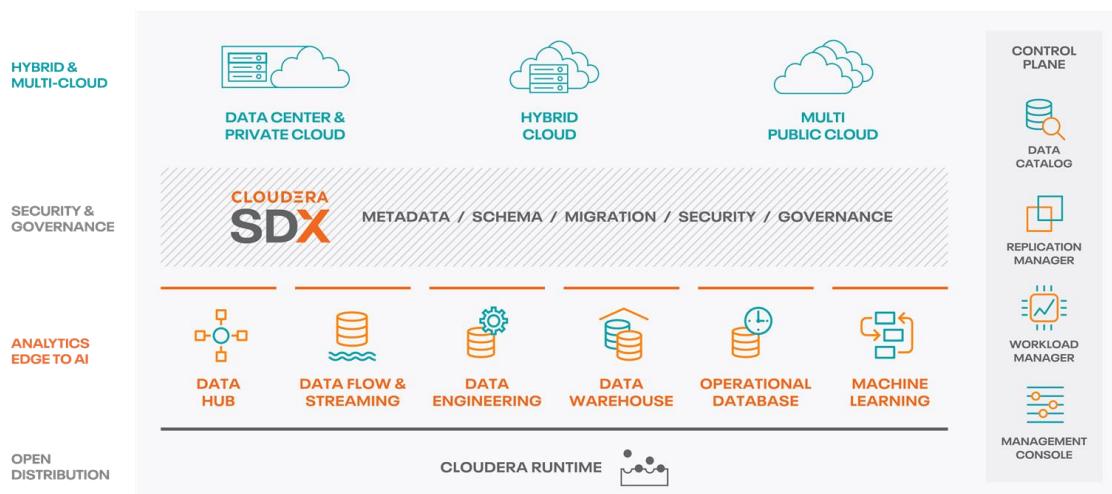
Moving from exploration to production at factory scale

- GET TO PRODUCTION, SCALE ML & AI USE CASES

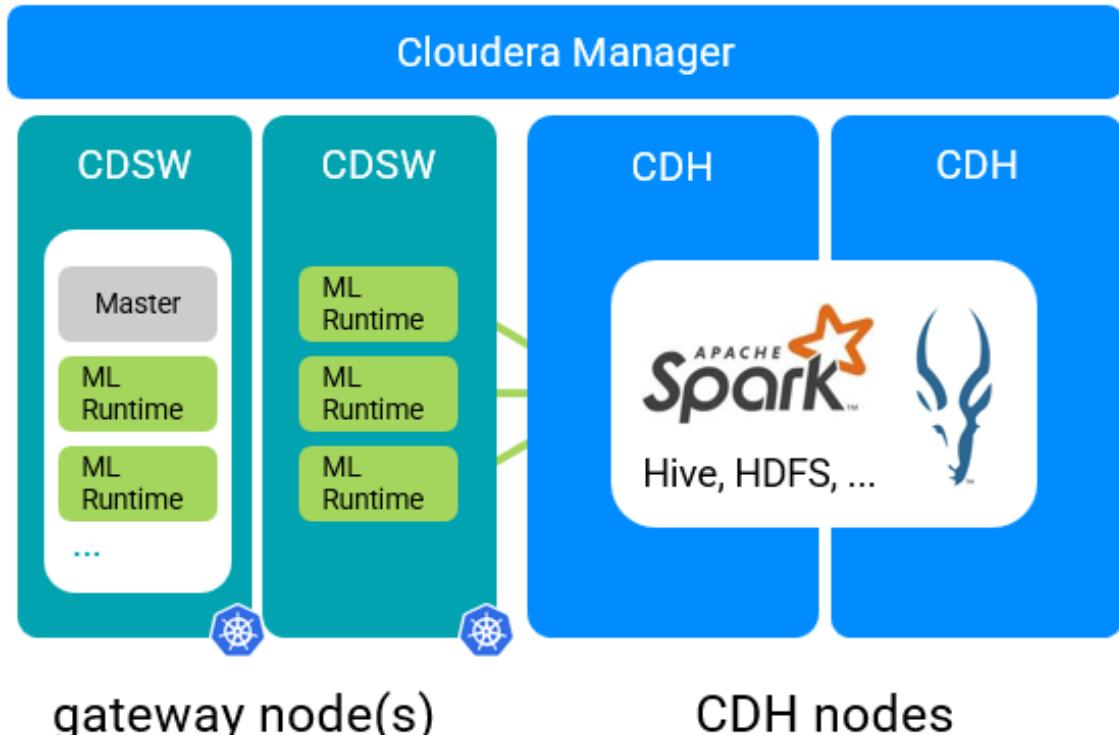


An integrated lifecycle is easier to use, manage and secure

- WORLD'S FIRST ENTERPRISE DATA CLOUD: CLOUDERA DATA PLATFORM (CDP)



- CDSW IS DESIGNED FOR THE DATACENTER



gateway node(s)

CDH nodes

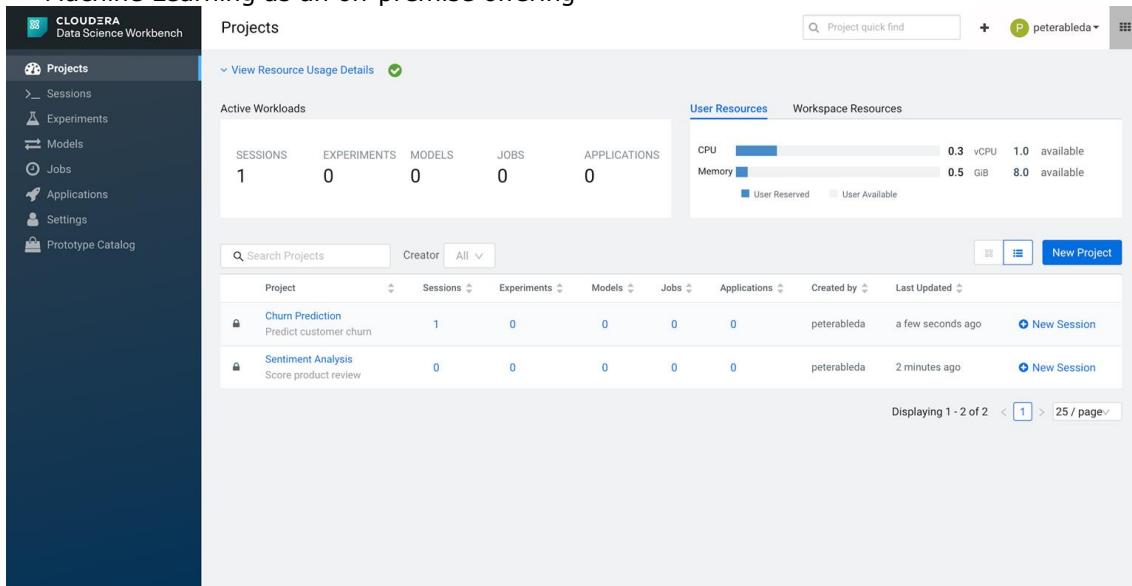
Containerized environments with scalable compute powered by Kubernetes and CDH

- Easy addition for self-service data science
- Seamless connectivity to existing clusters
- Flexible, isolated, reproducible user environments

- Each ML Workspace is an Elastic k8s cluster

The screenshot shows the configuration of an ML workspace. On the left, there's a form with "Advanced Options" turned on. It has sections for "CPU Settings" (with an "Instance Type" dropdown and a note to "Select an Environment to see options"), "Autoscale Range" (a slider from 0 to 10 with a value of 5), and "GPU Settings" (with an "Instance Type" dropdown and a note to "Select an Environment to see options"), which also has an "Autoscale Range" slider from 0 to 10 with a value of 10. On the right, there's a screenshot of the AWS CloudFormation console showing an environment named "mlx-dev-prod-env" created within a "cloud environments" stack. The stack status is "CREATE\_COMPLETE" with a green checkmark. The stack ID is "crn:cdp:environments:us-west-1:12a0079b-1591-4ca0-b721-...".

- Machine Learning as an on-premise offering

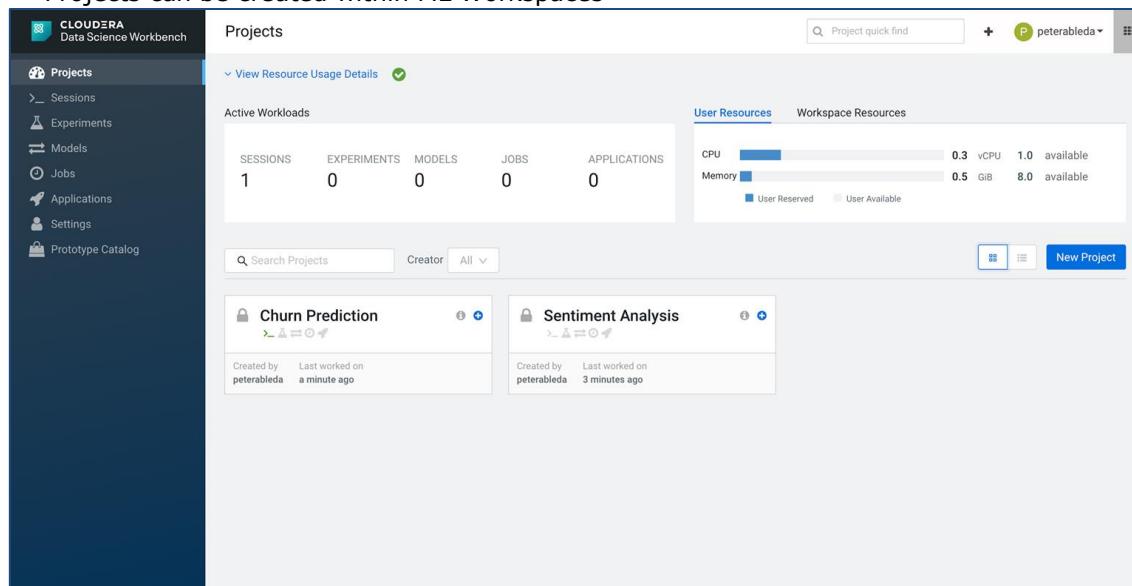


The screenshot shows the Cloudera Data Science Workbench interface. On the left is a dark sidebar with navigation links: Projects, Sessions, Experiments, Models, Jobs, Applications, Settings, and Prototype Catalog. The main area is titled 'Projects' and displays 'Active Workloads' with counts of 1 Session, 0 Experiments, 0 Models, 0 Jobs, and 0 Applications. Below this is a table of projects:

Project	Sessions	Experiments	Models	Jobs	Applications	Created by	Last Updated
Churn Prediction Predict customer churn	1	0	0	0	0	peterableda	a few seconds ago
Sentiment Analysis Score product review	0	0	0	0	0	peterableda	2 minutes ago

At the bottom right of the table, it says 'Displaying 1 - 2 of 2'. There are also buttons for 'New Project' and 'New Session'.

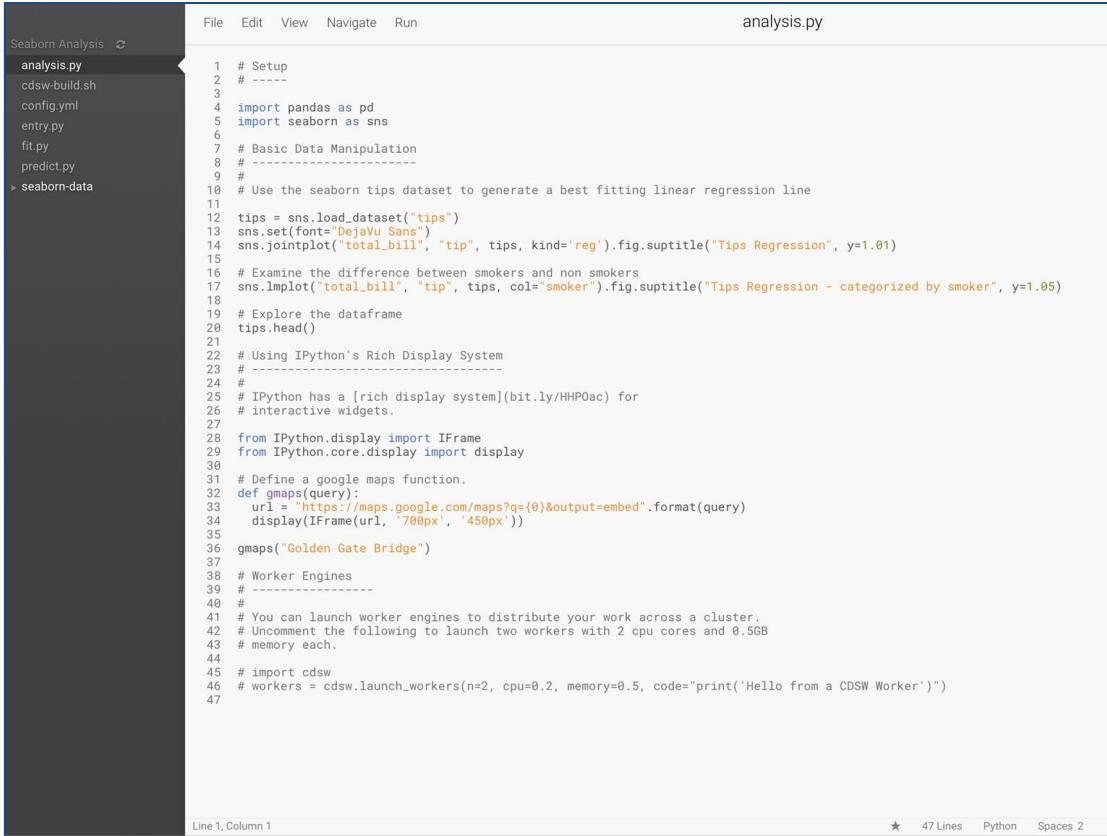
- Projects can be created within ML Workspaces



This screenshot shows the same Cloudera Data Science Workbench interface after new projects have been created. The 'Active Workloads' section now shows 1 Session, 0 Experiments, 0 Models, 0 Jobs, and 0 Applications. The project table now includes two entries:

Project	Sessions	Experiments	Models	Jobs	Applications	Created by	Last worked on
Churn Prediction	1	0	0	0	0	peterableda	a minute ago
Sentiment Analysis	0	0	0	0	0	peterableda	3 minutes ago

- A project is a directory structure containing code, data, and assets



The screenshot shows a Jupyter Notebook interface with a dark theme. On the left, a sidebar lists files: analysis.py, cdsw-build.sh, config.yml, entry.py, fit.py, predict.py, and a folder named 'seaborn-data'. The main area displays the contents of analysis.py. The code uses the Seaborn library to manipulate a tips dataset, create jointplots, and explore the data. It also demonstrates the IPython Rich Display System and Google Maps integration. The code is numbered from 1 to 47. At the bottom right of the code cell, there are icons for a star, 47 Lines, Python, and Spaces 2.

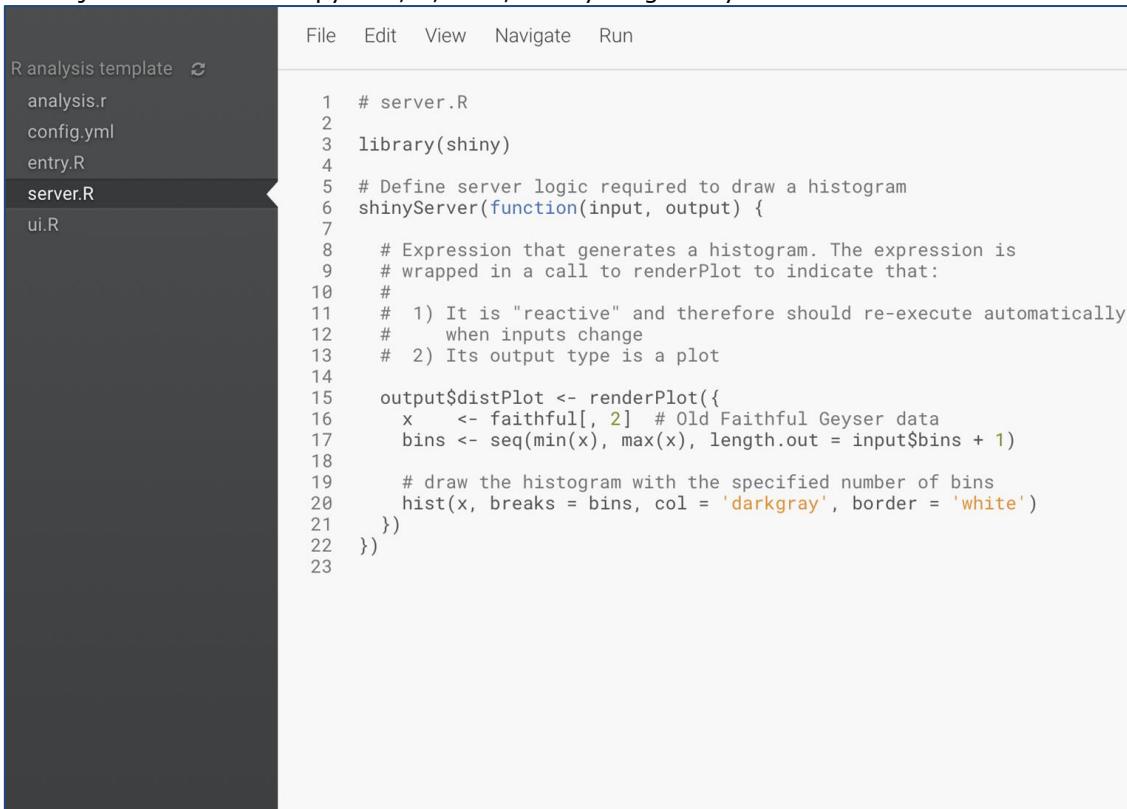
```

Seaborn Analysis ☀
analysis.py
cdsw-build.sh
config.yml
entry.py
fit.py
predict.py
» seaborn-data

1 # Setup
2 # -----
3
4 import pandas as pd
5 import seaborn as sns
6
7 # Basic Data Manipulation
8 # -----
9 #
10 # Use the seaborn tips dataset to generate a best fitting linear regression line
11
12 tips = sns.load_dataset("tips")
13 sns.set(font="DejaVu Sans")
14 sns.jointplot("total_bill", "tip", tips, kind='reg').fig.suptitle("Tips Regression", y=1.01)
15
16 # Examine the difference between smokers and non smokers
17 sns.lmplot("total_bill", "tip", tips, col="smoker").fig.suptitle("Tips Regression - categorized by smoker", y=1.05)
18
19 # Explore the dataframe
20 tips.head()
21
22 # Using IPython's Rich Display System
23 # -----
24 #
25 # IPython has a [rich display system](bit.ly/HHP0ac) for
26 # interactive widgets.
27
28 from IPython.display import IFrame
29 from IPython.core.display import display
30
31 # Define a google maps function.
32 def gmaps(query):
33     url = "https://maps.google.com/maps?q={0}&output=embed".format(query)
34     display(IFrame(url, '700px', '450px'))
35
36 gmaps("Golden Gate Bridge")
37
38 # Worker Engines
39 #
40 #
41 # You can launch worker engines to distribute your work across a cluster.
42 # Uncomment the following to launch two workers with 2 cpu cores and 0.5GB
43 # memory each.
44
45 # import cdsw
46 # workers = cdsw.launch_workers(n=2, cpu=0.2, memory=0.5, code="print('Hello from a CDSW Worker')")
```

Line 1, Column 1      ★ 47 Lines    Python    Spaces 2

- Project code can be in python, R, Java, or anything else you'd like



The screenshot shows a Jupyter Notebook interface with a dark theme. On the left, a sidebar lists files: analysis.r, config.yml, entry.R, server.R, and ui.R. The main area displays the contents of server.R. The code defines a shinyServer function that generates a histogram using the shiny and gridExtra packages. It specifies the input and output types, and uses renderPlot to create the histogram. The code is numbered from 1 to 23.

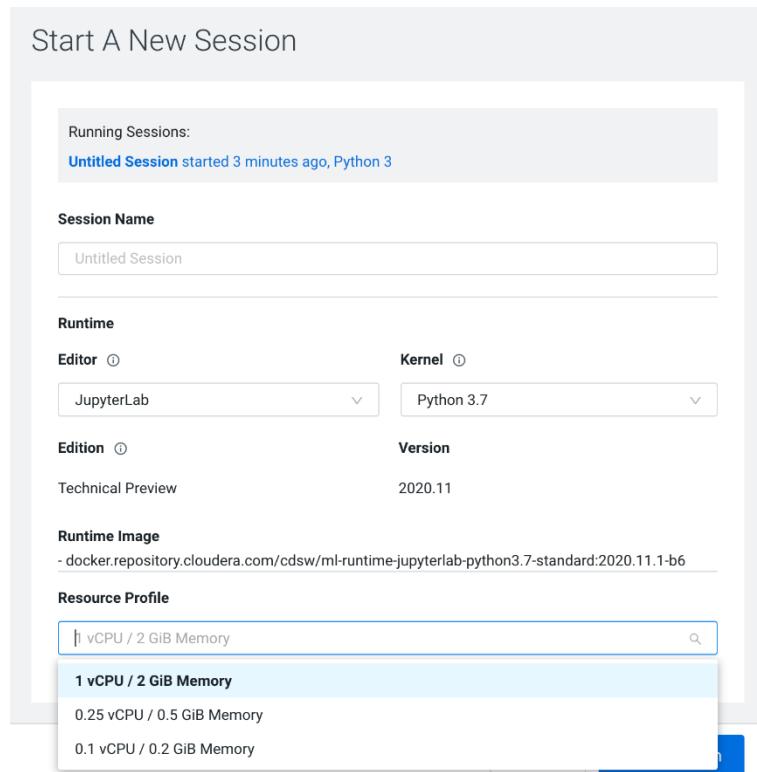
```

R analysis template ☀
analysis.r
config.yml
entry.R
server.R
ui.R

File Edit View Navigate Run

1 # server.R
2
3 library(shiny)
4
5 # Define server logic required to draw a histogram
6 shinyServer(function(input, output) {
7
8   # Expression that generates a histogram. The expression is
9   # wrapped in a call to renderPlot to indicate that:
10  #
11  # 1) It is "reactive" and therefore should re-execute automatically
12  #    when inputs change
13  # 2) Its output type is a plot
14
15  output$distPlot <- renderPlot({
16    x   <- faithful[, 2] # Old Faithful Geyser data
17    bins <- seq(min(x), max(x), length.out = input$bins + 1)
18
19    # draw the histogram with the specified number of bins
20    hist(x, breaks = bins, col = 'darkgray', border = 'white')
21  })
22 })
23
```

- Projects are worked on within isolated docker containers



Docker containers are used for project **development, model serving, applications, experiments** and scheduled **jobs**

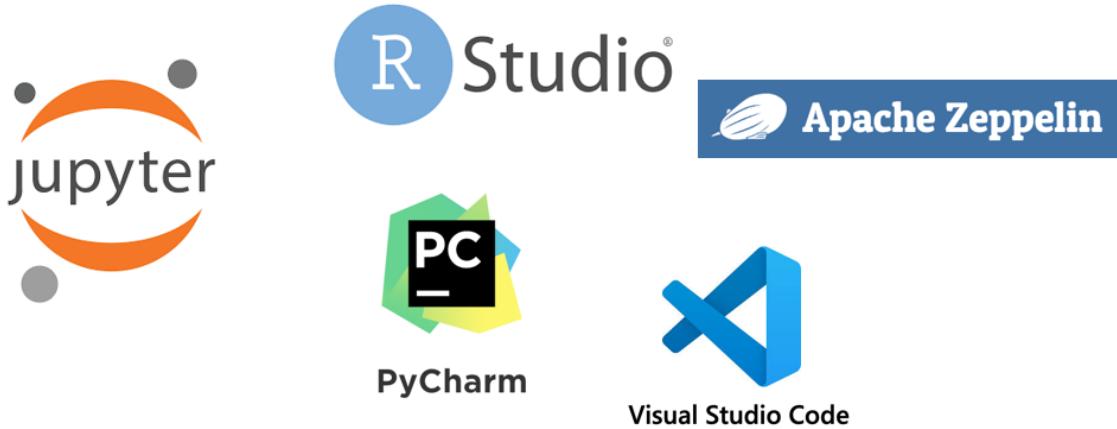
\*Note: Users select resources to use within the overall workspace constraints

- CML Supports Built-in Workbench and Notebooks



Jupyter Notebooks and Workbench are integrated with the engine to provide a **full IDE** capability. Editor preferences range broadly in the data science community.

- CML Also Supports Any Editor



Only 1 of 3 data scientists use notebooks - [Kaggle ML and DS Survey 2019](#)

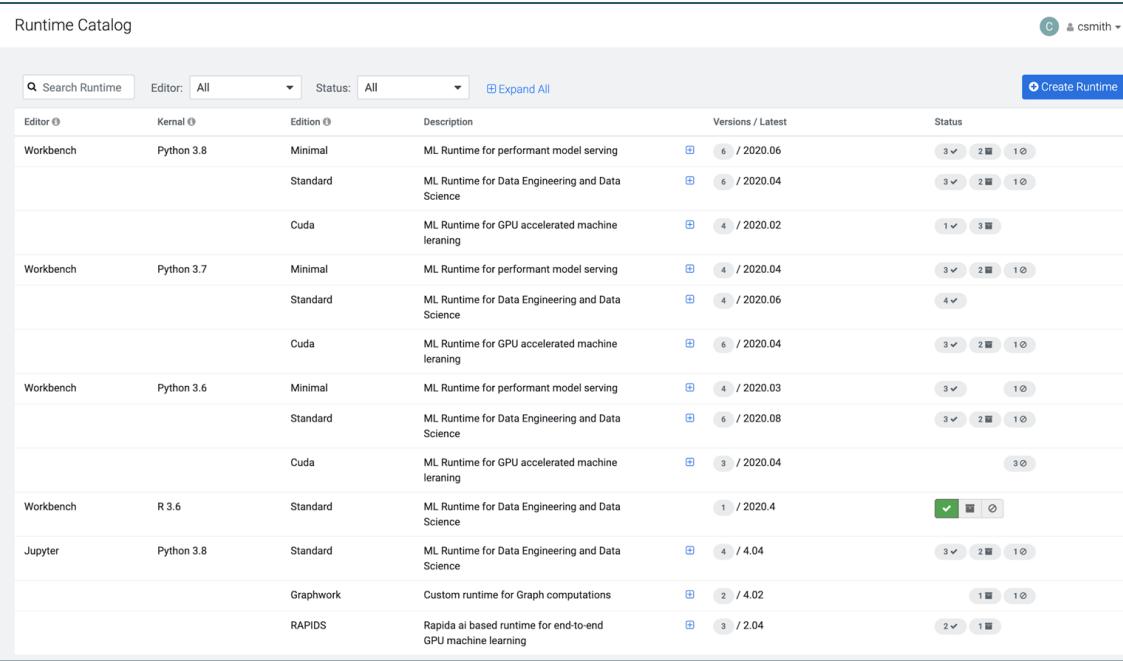
- Cloudera ML Runtimes & Engines



CML: Powered by NVIDIA RAPIDS for *Python, Pandas, Spark*

- RAPIDS Accelerated, GPU Ready | Dask, Python, R, Spark & More

Runtime Catalog

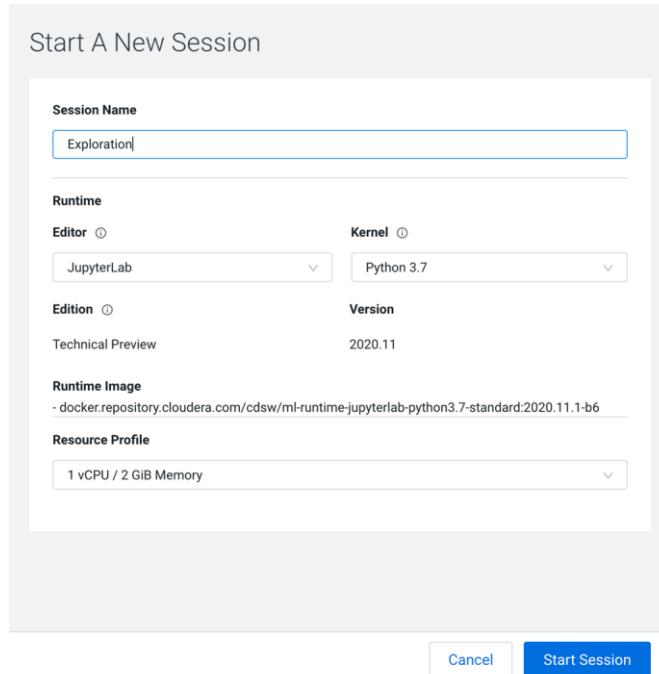


Editor ⓘ	Kernel ⓘ	Edition ⓘ	Description	Versions / Latest	Status
Workbench	Python 3.8	Minimal	ML Runtime for performant model serving	6 / 2020.06	3 2 1
		Standard	ML Runtime for Data Engineering and Data Science	6 / 2020.04	3 2 1
		Cuda	ML Runtime for GPU accelerated machine learning	4 / 2020.02	1 3
Workbench	Python 3.7	Minimal	ML Runtime for performant model serving	4 / 2020.04	3 2 1
		Standard	ML Runtime for Data Engineering and Data Science	4 / 2020.06	4
		Cuda	ML Runtime for GPU accelerated machine learning	6 / 2020.04	3 2 1
Workbench	Python 3.6	Minimal	ML Runtime for performant model serving	4 / 2020.03	3 2 1
		Standard	ML Runtime for Data Engineering and Data Science	6 / 2020.08	3 2 1
		Cuda	ML Runtime for GPU accelerated machine learning	3 / 2020.04	3
Workbench	R 3.6	Standard	ML Runtime for Data Engineering and Data Science	1 / 2020.4	✓ 0
Jupyter	Python 3.8	Standard	ML Runtime for Data Engineering and Data Science	4 / 4.04	3 2 1
		Graphwork	Custom runtime for Graph computations	2 / 4.02	1 1
		RAPIDS	Rapida ai based runtime for end-to-end GPU machine learning	3 / 2.04	2 1

We ship with several standard engines; new ones can be built and registered by users. Runtime Catalog is a Roadmap item but different runtimes are already available

- Runtime Environments Enable Flexibility in Customization

Start A New Session



Session Name  
Exploration

Runtime

Editor ⓘ	Kernel ⓘ
JupyterLab	Python 3.7

Edition ⓘ      Version

Technical Preview      2020.11

Runtime Image  
- docker.repository.cloudera.com/cdsw/ml-runtime-jupyterlab-python3.7-standard:2020.11.1-b6

Resource Profile  
1 vCPU / 2 GiB Memory

Cancel      Start Session

- **Editor**  
Development interface to write and execute code. Examples: Workbench, JupyterLab
- **Kernel**  
Execution engine for the session of work. Examples: Python 3.7, Python 3.8, R3.6
- **Edition**  
Set of capabilities (tools/libraries) available for the run. Examples: Standard, RAPIDS

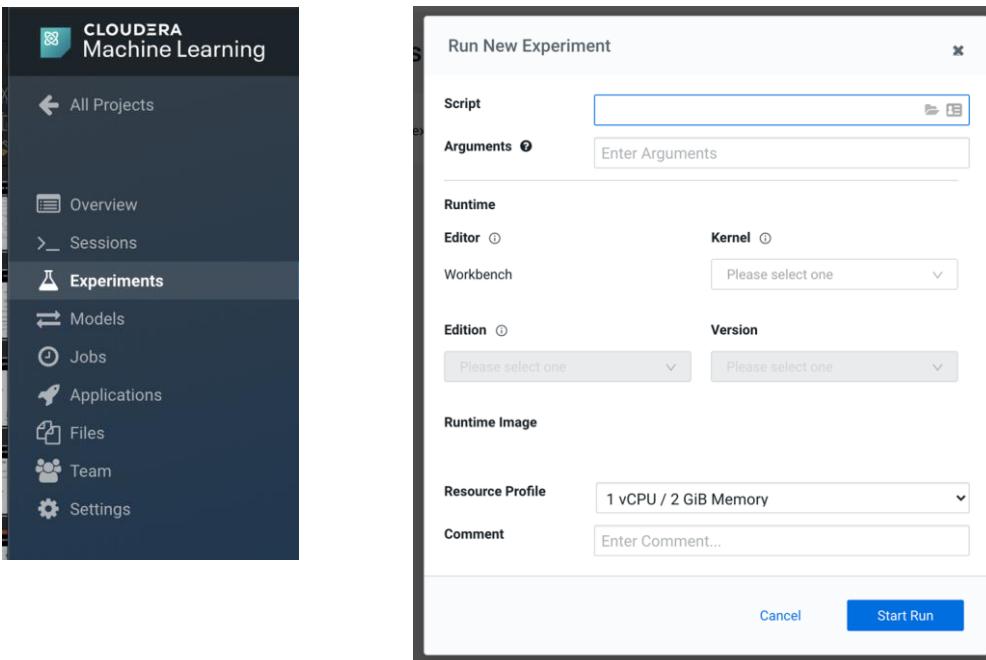
- **Version**
  - Major version of the runtime. Examples: 2020-04, 2020-06
- Additional packages can be installed and used as needed

## st / Seaborn Analysis



\*Note: These are **per-project** and picked by **developers**

- Experiments are used for multiple runs with different parameters



The screenshot shows the Cloudera Machine Learning interface. On the left, there is a sidebar with the following navigation options:

- All Projects
- Overview
- Sessions
- Experiments** (selected)
- Models
- Jobs
- Applications
- Files
- Team
- Settings

To the right of the sidebar, a modal window titled "Run New Experiment" is displayed. The modal contains the following fields:

- Script:** A file input field.
- Arguments:** An input field labeled "Enter Arguments".
- Runtime:**
  - Editor:** A dropdown menu with "Please select one".
  - Workbench:** A dropdown menu with "Please select one".
- Edition:** A dropdown menu with "Please select one".
- Version:** A dropdown menu with "Please select one".
- Runtime Image:** A dropdown menu with "1 vCPU / 2 GiB Memory".
- Resource Profile:** A dropdown menu with "1 vCPU / 2 GiB Memory".
- Comment:** An input field labeled "Enter Comment...".

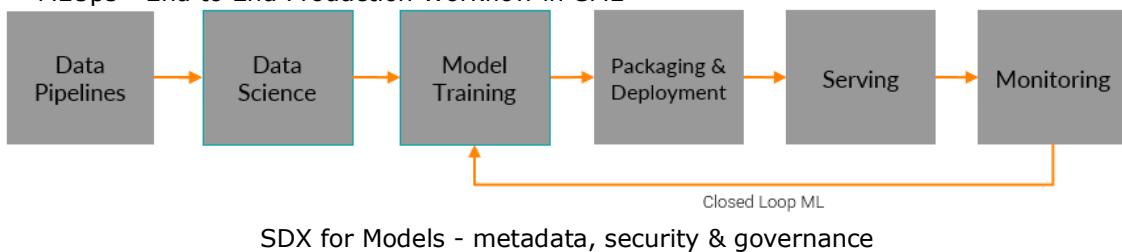
At the bottom of the modal are two buttons: "Cancel" and "Start Run".

Hyper-parameter tuning can be done by looking at the outputs of different runs

- App development can be done in the framework of your choice



- MLOps - End to End Production Workflow in CML



SDX for Models - metadata, security & governance

- Hidden ML Lifecycle Challenges

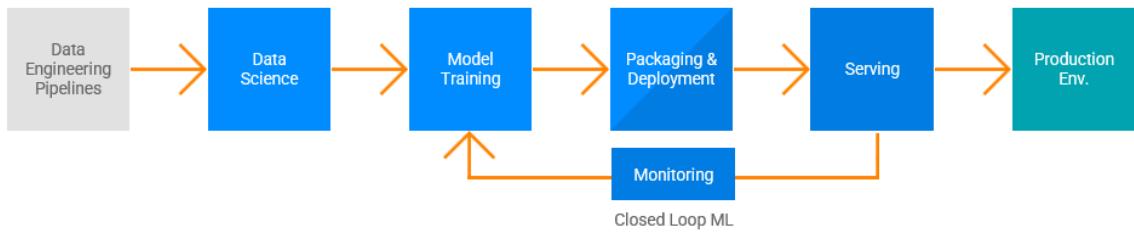


The technical debt of ML creates a barrier to production

Source: <https://papers.nips.cc/paper/5656-hidden-technical-debt-in-machine-learning-systems.pdf>

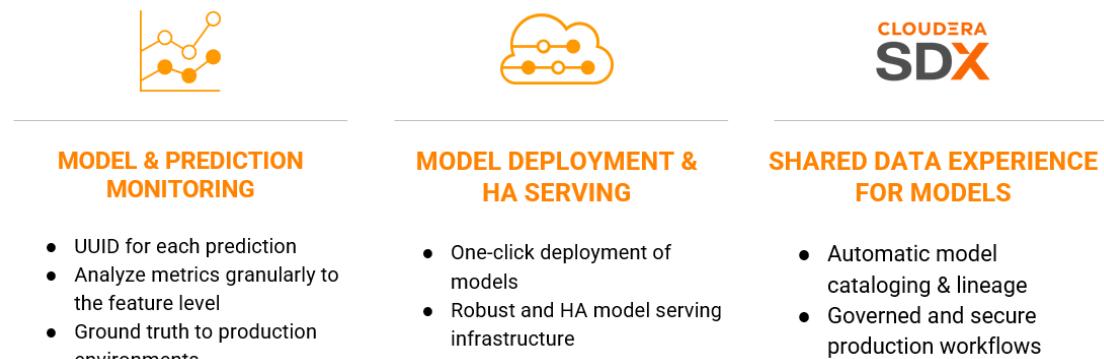
Only a small fraction of real-world ML systems is composed of the ML code, as shown by the small black box in the middle. The required surrounding infrastructure is vast and complex.

- Get To Production Faster, Scale AI Use Cases



An integrated lifecycle is easier to use, manage and secure

- MLOps in Cloudera Machine Learning



Enabling Production ML At Scale

- Models are Deployed Within a Project



The screenshot shows the Cloudera Machine Learning interface. The left sidebar has a dark blue background with white icons and text. The 'Files' icon is highlighted with a light blue background. The main area has a white background with a dark grey header bar. The header bar contains the Cloudera logo, the text 'CLOUDERA Machine Learning', and the breadcrumb navigation 'st / Seaborn Analysis'. Below the header, the breadcrumb 'Files / predict.py' is shown. The main content area displays Python code for a machine learning prediction script.

```
# Read the fitted model from the file model.pkl
# and define a function that uses the model to
# predict petal width from petal length

import pickle

model = pickle.load(open('model.pkl', 'rb'))

def predict(args):
    petal_length = float(args.get('petal_length'))
    result = model.predict([[petal_length]])
    return result[0][0]
```

Write the function that makes the prediction

### Create a Model

**General**

Name \*

Description \*

**Build**

File \*

Function \*

Example Input ⓘ

```
{
  "petal_length": 3.2
}
```

Then use CML to serve it up

- Models Can Be Scaled Up within CML

### Deployment

**Engine Profile**

**GPUs**

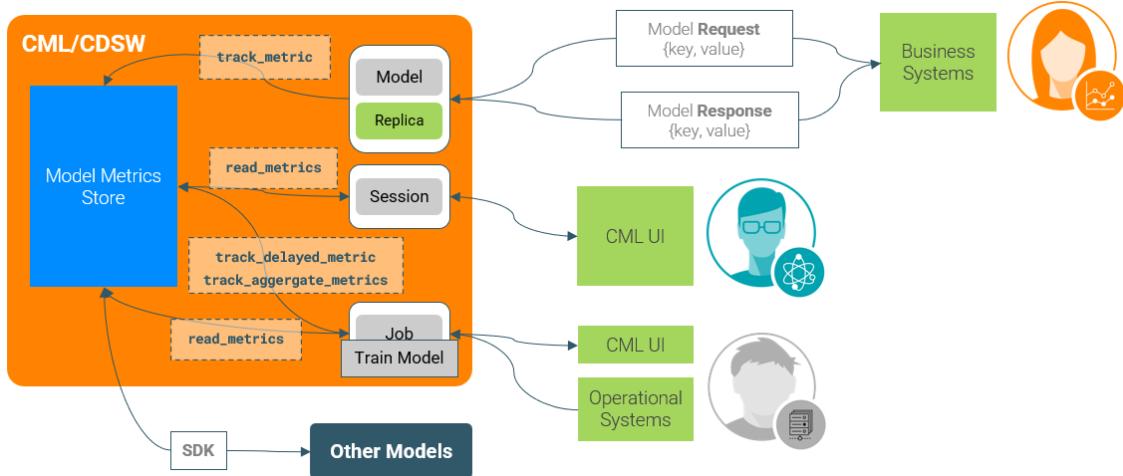
**Replicas**

[Set Environmental Variables](#)

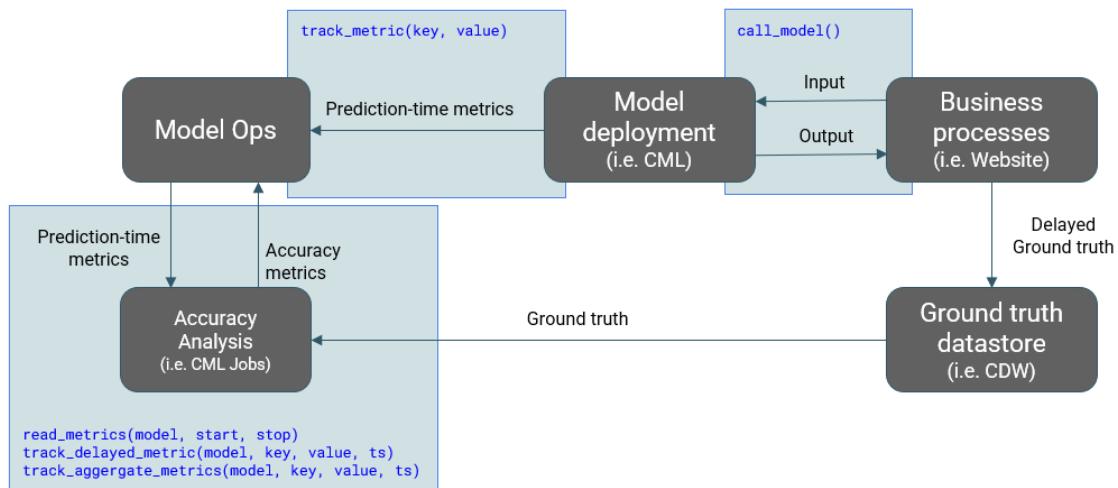
**Deploy Model**   [Cancel](#)

Use this to build **scalable FAAS** within the project. Project is snapshot-ed for a model so that later development changes don't impact production models.

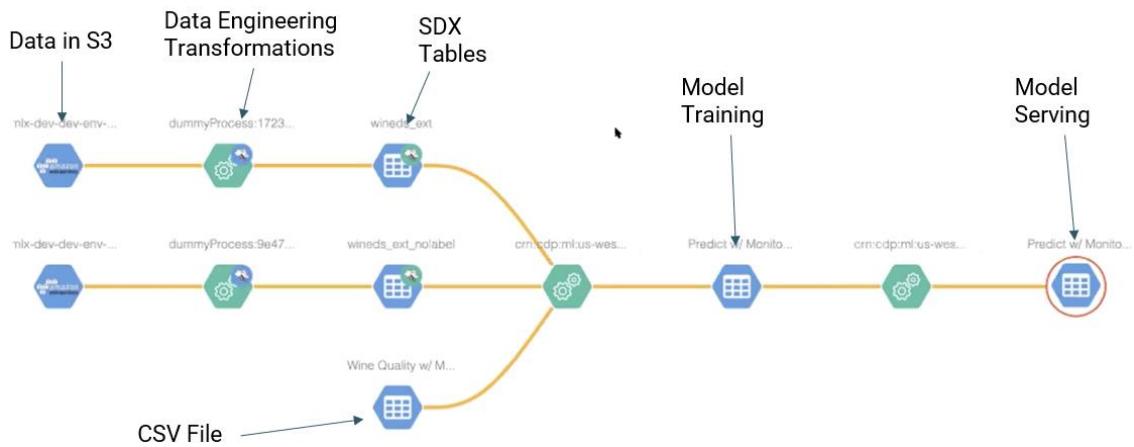
- ML OPS WORKFLOW



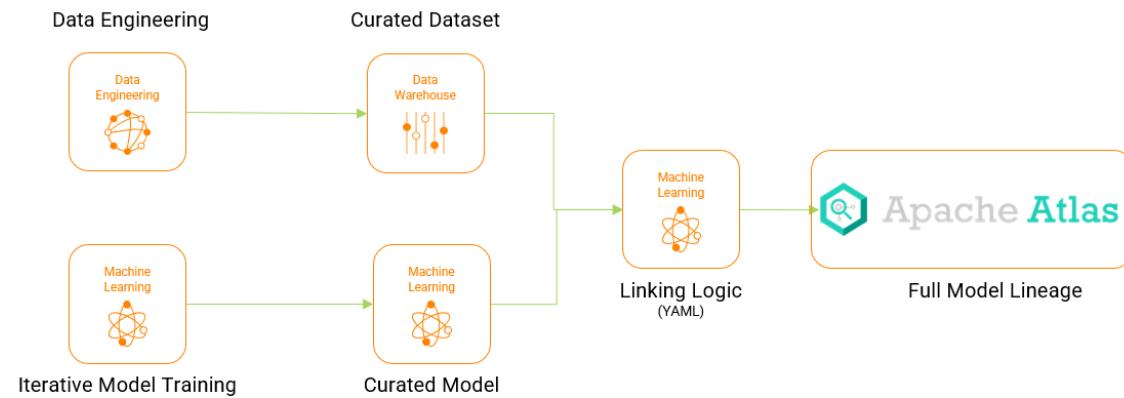
- MLOps SDK: APIs Allow Models to Track and Monitor Metrics**



- End-to-End Auditability for Production Models through SDX



- Model Lineage Tracking is done in Atlas



## AUGMENTED ANALYTICS

- Challenges from Users

ML IS REPETITIVE



Repetitive use cases  
Repetitive ML Patterns

HIGH INTEREST,  
LOW EXPERTISE



ML in the hands of  
Data Analysts

- Cloudera Data Science Workbench

The screenshot displays the Cloudera Data Science Workbench interface, specifically the 'Applied ML Prototypes Catalog' section. The left sidebar shows navigation links for Projects, Sessions, Experiments, Models, Applications, Settings, and Prototype Catalog, with 'Prototype Catalog' currently selected. The main area is titled 'Applied ML Prototypes Catalog' and features a grid of nine prototype cards:

- Churn Modeling with XGBoost**: EXPLAINABILITY\_XGBOOST
- Airline Delay Prediction**: XGBOOST
- Deep Learning for Image Analysis**: COMPUTER\_VISION IMAGES
- Neural Question Answering**: NLP
- Structural Time Series**: TIME\_SERIES
- Deep Learning for Anomaly Detection**: ANOMALY\_DETECTION DEEP\_NEURAL\_NETS
- Fraud Detection**: FRAUD\_DETECTION ANOMALY\_DETECTION
- Sentiment Analysis with SparkNLP and Tensorflow**: SENTIMENT\_ANALYSIS SPARK

At the bottom left, a warning message states: '⚠ License expires in 59 days <<'.

CLOUDERA Data Science Workbench

Projects Sessions Experiments Models Jobs Applications Settings Prototype Catalog Admin

License expires in 59 days <<

### Applied ML Prototypes Catalog

Search prototypes Tags: all

Churn Modeling with XGBoost

This project demonstrates how to build a classifier model (Logistic Regression, Xgboost) to predict the churn probability for a group of customers from a telecoms company. In addition, the model can then be interpreted using LIME. Both the classifier and LIME models are then deployed using CMLS real-time model deployment capability. Finally, a basic flask application is deployed that will let you interact with the real-time model to see which factors in the data have the most influence on the churn probability.

View on Github

Cancel Launch as Session

Structural Time Series Deep Learning for Anomaly Detection Fraud Detection

EXPLAINABILITY XGBOOST ANOMALY DETECTION DEEP NEURAL NETS FRAUD DETECTION ANOMALY DETECTION

Neural Question Answering

NLP

Image Analysis Model Application

Model Results Output

The final response will be sent to the selected mode and the response will be returned to the user.

The model is 95.8% confident that is Negative.

Sentiment Analysis with Sparklyr and Tensorflow

SENTIMENT ANALYSIS SPARK

< 1 2 >

CLOUDERA Machine Learning

All Projects Overview Sessions Experiments Models Jobs Applications Files Team

Get Started Help

### css0\_shubho / Churn Modeling with XGBoost - 1... / Prototype Status

Executing Prototype Setup Steps for new Project

Prototype Name: ML Churn Demo (v1)

Prototype to demonstrate building a churn model on CML

Completed all steps

Step 3 Job to train models. View details completed 10/19/2020 11:32 AM

Step 4 Run model training job. View details completed 10/19/2020 11:32 AM

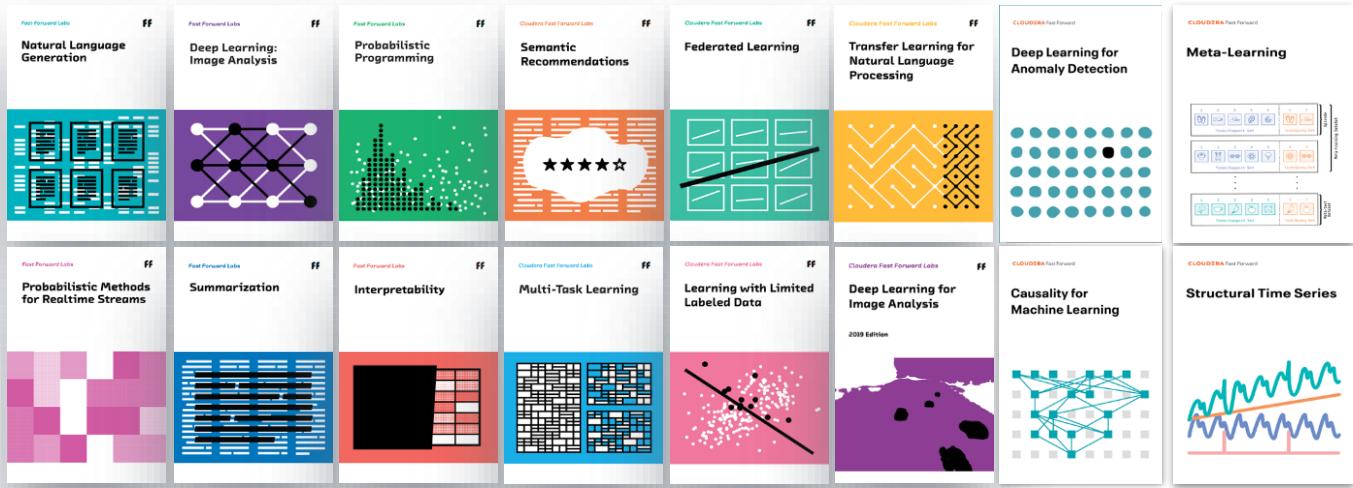
Running the job to train models.

Step 5 Create the churn model prediction api endpoint View details completed 10/19/2020 11:32 AM

Step 6 Build model View details completed 10/19/2020 11:37 AM

Step 7 Deploy model View details completed 10/19/2020 11:37 AM

- AMPs Seeded with Applied ML Research from Cloudera Fast Forward



- Example: Deep Learning for Question Answering

**NeuralQA**

Select any sample question/passage pair below

what is the goal of the fourth amendment? Who was the first woman to serve on the supreme court in America  
Where did Sandra Day grow up?

Enter question

what is the goal of the fourth amendment? Get Answer

Enter passage

The Fourth Amendment of the U.S. Constitution provides that the right of the people to be secure in their persons, houses, papers, and effects, against unreasonable searches and seizures, shall not be violated, and no Warrants shall issue, but upon probable cause, supported by Oath or affirmation, and particularly describing the place to be searched, and the persons or things to be seized.'The ultimate goal of this provision is to protect people's right to privacy and freedom from unreasonable intrusions by the government. However, the Fourth Amendment does not guarantee protection from all searches and seizures, but only those done by the government and deemed unreasonable under the law.

**1 Answer returned 0.211 seconds.**

**A0** Time: 0.209s | 0.4220  
“to protect people’s right to privacy and freedom from unreasonable intrusions by the government”  
what is the goal of the fourth amendment? the fourth amendment of the u . s . constitution provides that the right of the people to be secure in their persons , houses , papers , and effects , against unreasonable searches and seizures , shall not be violated , and no warrants shall issue , but upon probable cause , supported by oath or affirmation , and particularly describing the place to be searched , and the persons or things to be seized . ' the ultimate goal of this provision is to protect people ‘ s right to privacy and freedom from unreasonable intrusions by the government . however , the fourth amendment does not guarantee protection from all searches and seizures , but only those done by the government and deemed unreasonable under the law

© NeuralQA 2020. Learn more on [Github](#).

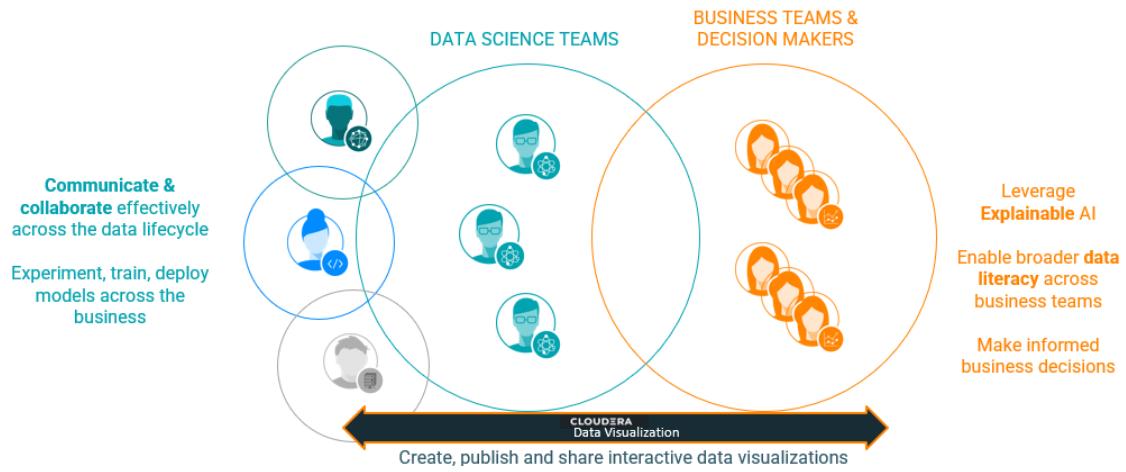
- You can play with this in your browser right after the session at: [neuralqa.fastforwardlabs.com](http://neuralqa.fastforwardlabs.com)
- For more prototypes exploring the state of the art in machine learning capabilities: [blog.fastforwardlabs.com/prototypes](http://blog.fastforwardlabs.com/prototypes)

- Integrated Arcadia Business Intelligence

The screenshot shows the Cloudera Machine Learning (CML) interface. On the left, there's a sidebar with options like Projects, Sessions, Experiments, Models, Jobs, Applications, Settings, and Admin. A red arrow points from this sidebar to the main dashboard area. The main dashboard is titled "priyankp / Projects" and shows a list of projects, including "Flight Data Analy". To the right, there's a grid of various data visualizations and dashboards, such as "Example analysis", "Churn Demo App", "Interactive & Dynamic dashboards", "Animated world population - GDP vs life expectancy", "Mapping customers", "SKO'21 ML Demo - IRA", "SKO - IRA bank analysis", and "Deficiency Details <<county Queens>>".

Native Data Visualizations in CML (*acquired Sept 2019 and fully integrated in 2020*)

- Visual Collaboration for Getting to Production

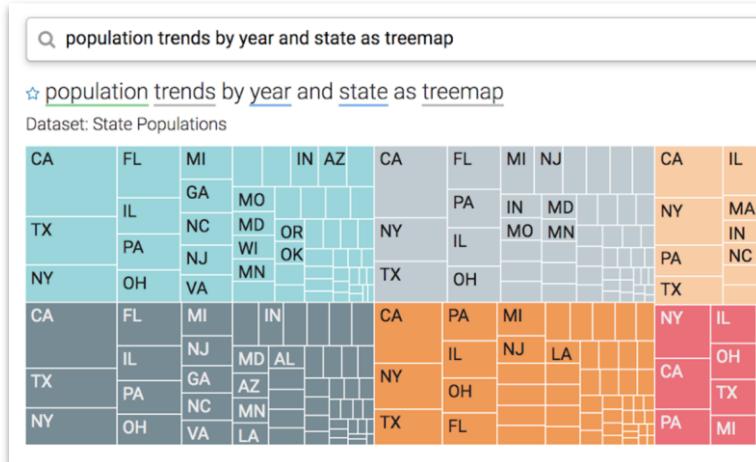


Accelerating Production ML Workflows from Raw Data to Business Impact

- Business User Friendly App Creation for Improved Data Literacy

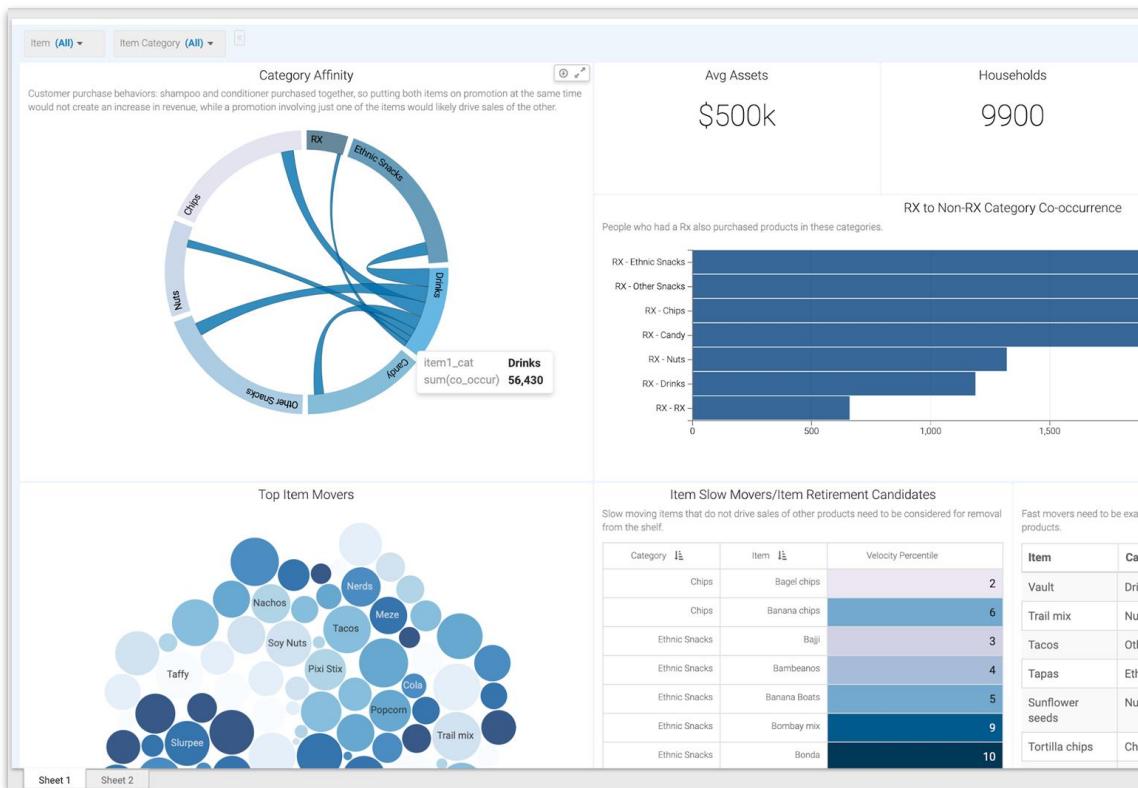


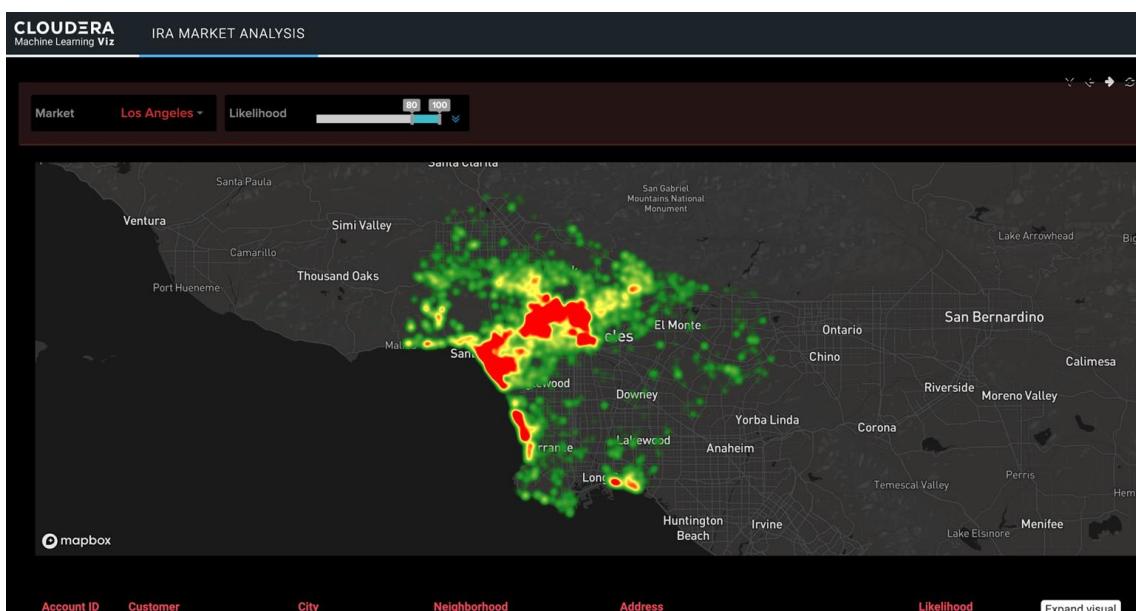
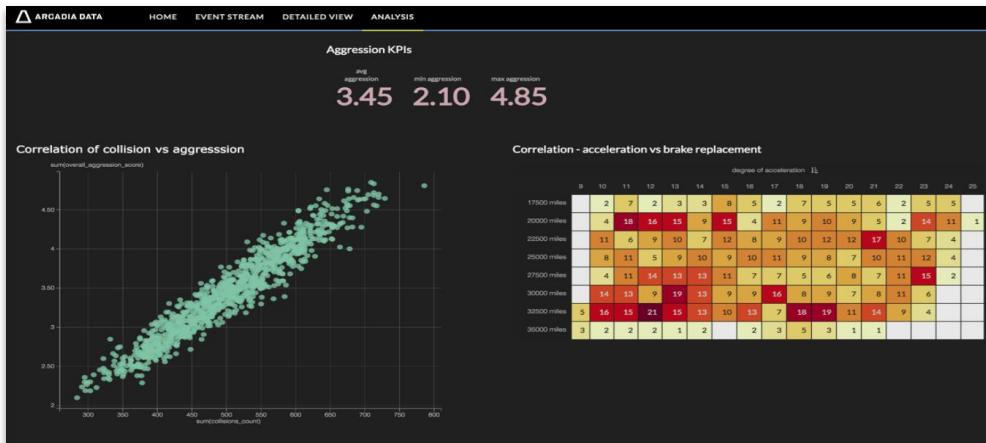
Visual recommendations



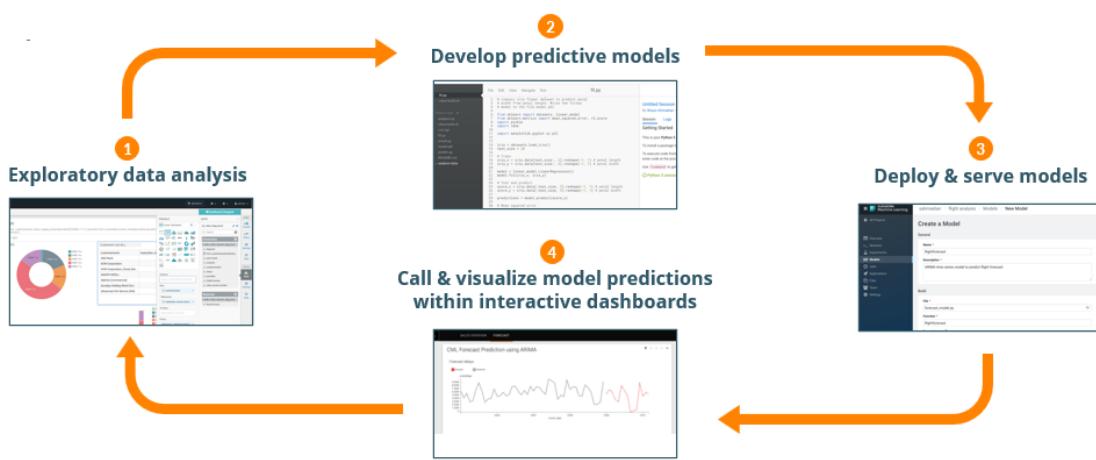
Natural language search

- Build Rich Visual Applications



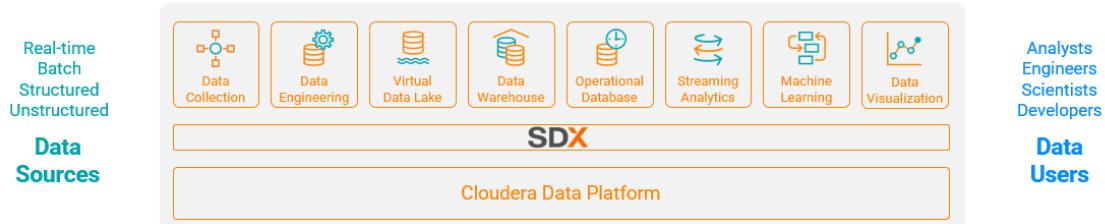


- Integrated Data Visualizations for ML Workflows



## LAMPIRAN 6: CLOUDERA DATA WAREHOUSE

### A HYBRID / MULTI-CLOUD DATA PLATFORM AND AN INTEGRATED SUITE OF SECURE ANALYTIC APPS

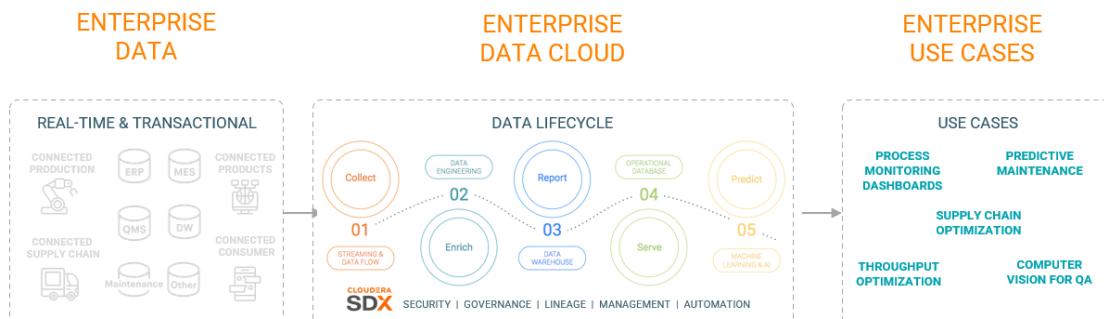


- HOW DO CUSTOMERS USE CLOUDERA?

	BANKING	TECHNOLOGY	TELCO	LIFE SCIENCES	MANUFACTURING
USE CASES	<ul style="list-style-type: none"> <li>• Fraud detection</li> <li>• Anti-money laundering</li> <li>• Spend analytics</li> </ul>	<ul style="list-style-type: none"> <li>• Customer analytics</li> <li>• Threat detection</li> <li>• Predictive support</li> </ul>	<ul style="list-style-type: none"> <li>• Churn analysis</li> <li>• Customer care</li> <li>• Network optimization</li> </ul>	<ul style="list-style-type: none"> <li>• Patient care (IoT)</li> <li>• Genomics research</li> <li>• Regulatory compliance</li> </ul>	<ul style="list-style-type: none"> <li>• Predictive maintenance (IoT)</li> <li>• Supply chain optimization</li> <li>• Remote monitoring</li> </ul>
KEY CUSTOMERS	<ul style="list-style-type: none"> <li>• Barclays</li> <li>• Citi</li> <li>• Santander UK</li> </ul>	<ul style="list-style-type: none"> <li>• Cisco</li> <li>• Intel</li> <li>• Reef Technology</li> </ul>	<ul style="list-style-type: none"> <li>• Globe Telecom</li> <li>• Deutsche Telekom</li> <li>• Robi Axiata</li> </ul>	<ul style="list-style-type: none"> <li>• GlaxoSmithKline</li> <li>• Clearsense</li> <li>• Cerner</li> </ul>	<ul style="list-style-type: none"> <li>• Navistar</li> <li>• Micron</li> <li>• Sikorsky</li> </ul>

Every business use case is a data lifecycle use case

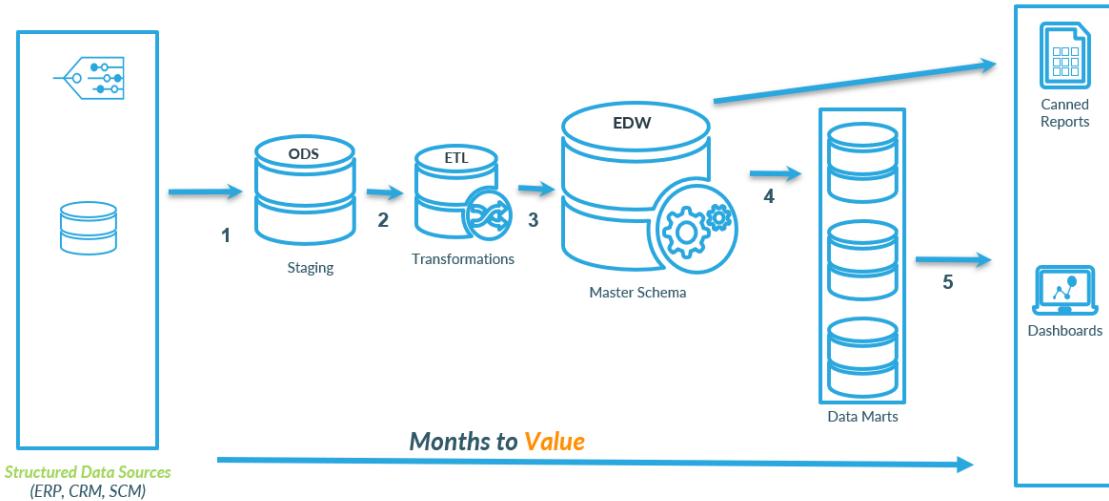
- BUSINESS USE CASES REQUIRE THE DATA LIFECYCLE



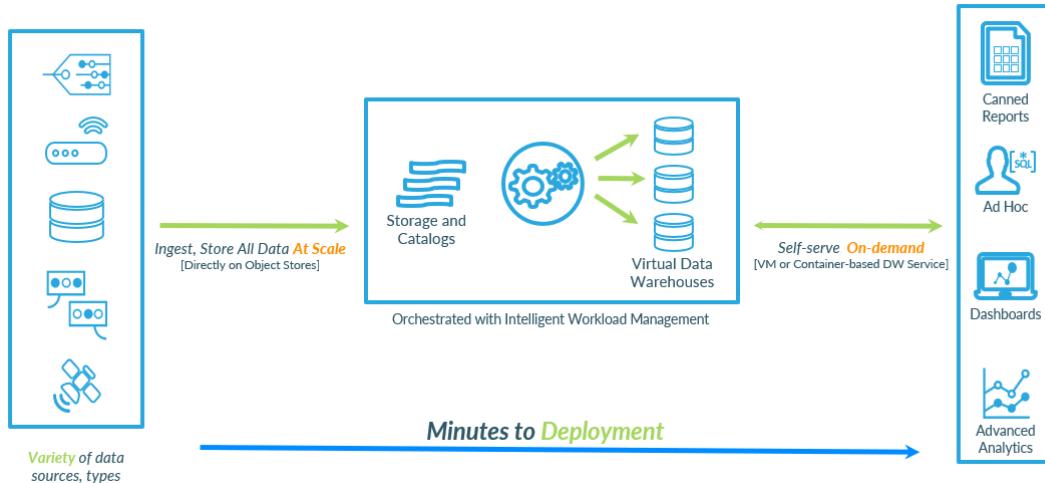
An integrated lifecycle is easier to use, manage and secure

## CDP Data Service - Cloudera Data Warehouse

- CUMBERSOME: TRADITIONAL DATA WAREHOUSE



- ELEGANT: MODERN DATA WAREHOUSE



- WHY CLOUDERA DATA WAREHOUSE?

Self-service analytics



Businesses demand data democracy

Faster Insights for faster decisions



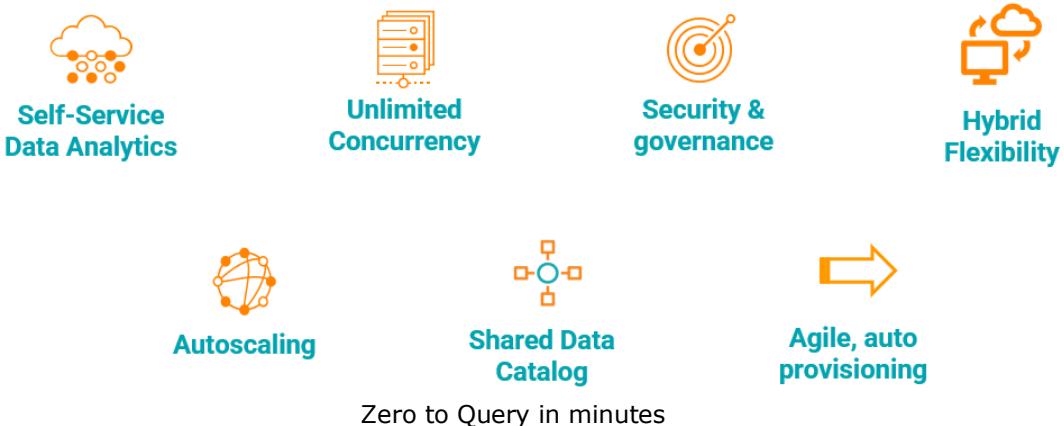
LOBs cannot wait

Access to modern data use cases

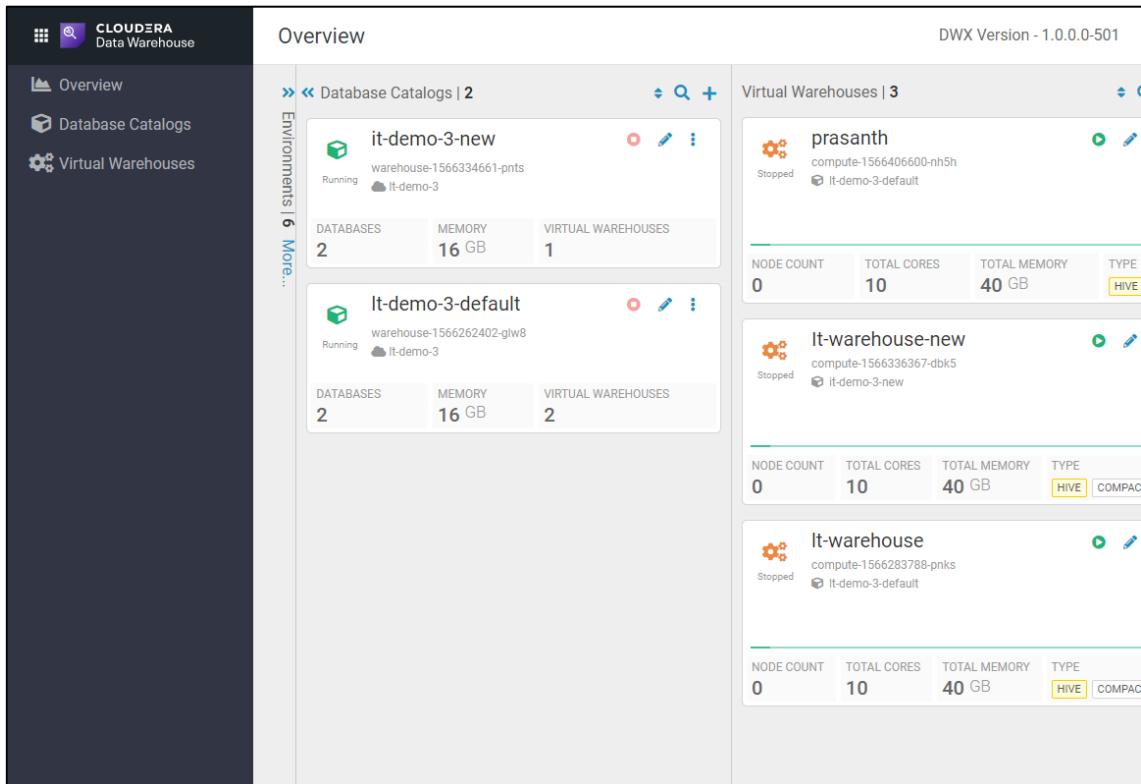


Need data to build competitive advantage

- BENEFITS OF CLOUDERA DATA WAREHOUSE



- Cloudera Data Warehouse



The screenshot shows the Cloudera Data Warehouse (DWX) interface with the following details:

- Left Sidebar:** Includes links for Overview, Database Catalogs, and Virtual Warehouses. It also shows 6 environments and a More... button.
- Top Bar:** Displays the DWX Version - 1.0.0.0-501.
- Overview Section:**
  - Database Catalogs | 2:** Shows two entries: "it-demo-3-new" (Running) and "lt-demo-3-default" (Running).
  - Virtual Warehouses | 3:** Shows three entries: "prasanth" (Stopped), "lt-warehouse-new" (Stopped), and "lt-warehouse" (Stopped).
- Table Headers:** NODE COUNT, TOTAL CORES, TOTAL MEMORY, TYPE.
- Data Rows:**

	NODE COUNT	TOTAL CORES	TOTAL MEMORY	TYPE
it-demo-3-new	0	10	40 GB	HIVE
lt-demo-3-default	0	10	40 GB	HIVE
prasanth	0	10	40 GB	HIVE
lt-warehouse-new	0	10	40 GB	HIVE
lt-warehouse	0	10	40 GB	HIVE

A data warehousing service optimized for concurrency, caching, and isolation

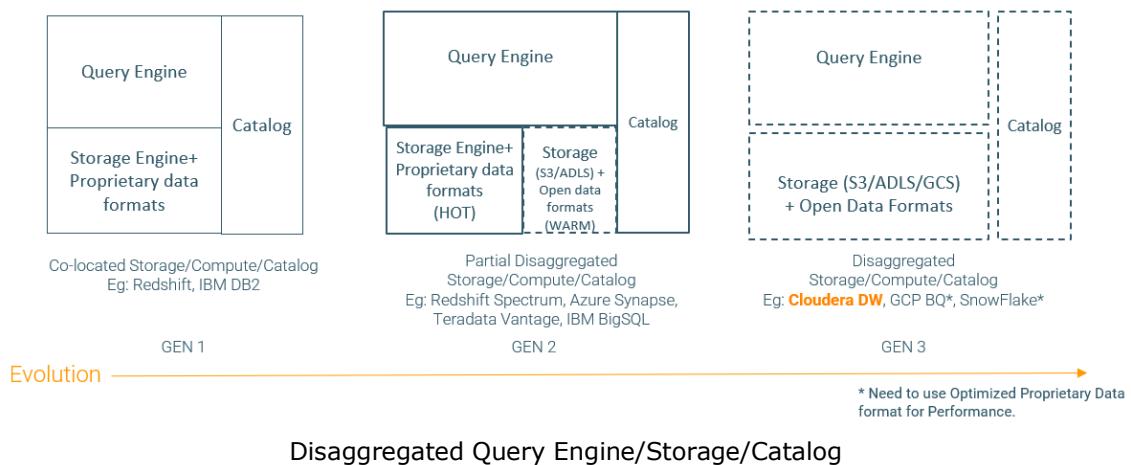
### 1. Quick Time to Value

- Ease of Use
- Integrated Visual Applications
- Shared Data Catalog
- Traditional & New Data Types

### 2. Lower Cost of Ownership

- Multi-Engine & lowest Price/Performance
- Elasticity, Automation & Intelligence
- Workload Management

- EVOLUTION OF THE ARCHITECTURE

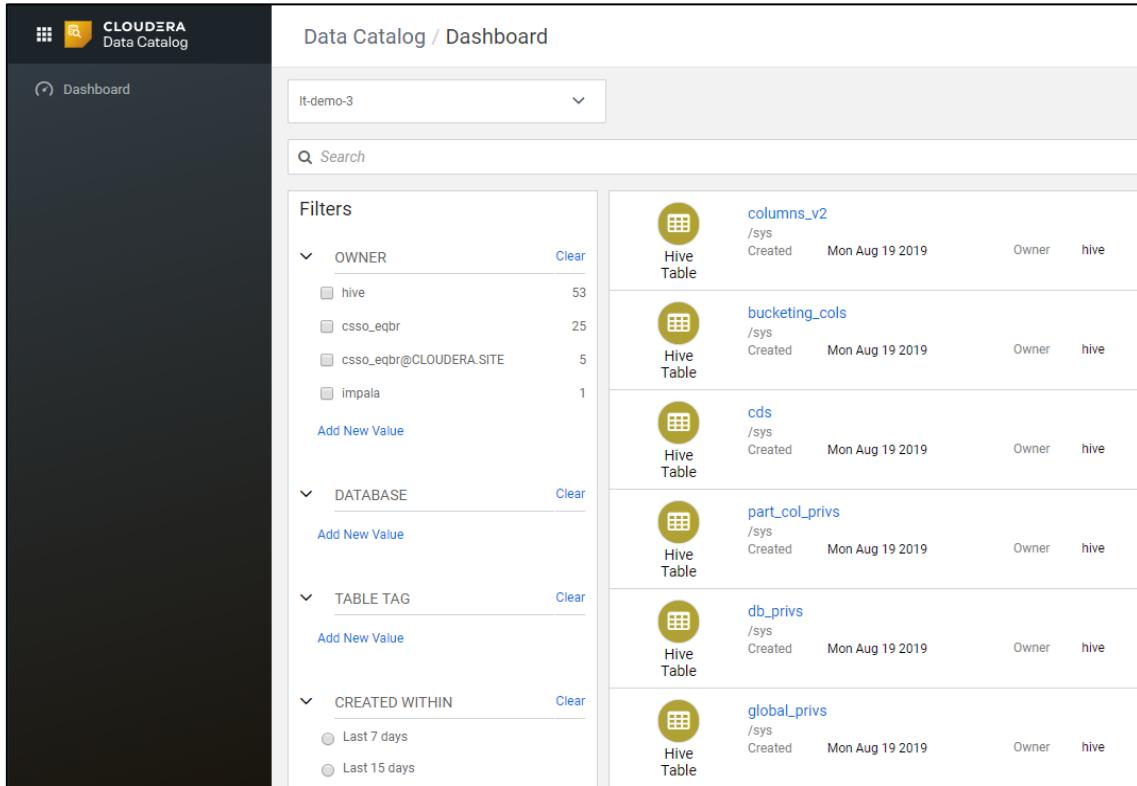


- EASE OF USE

Zero to Query in 10 minutes or less

- Simplified provisioning**
  - Relieve Central IT from LoB projects with the simplicity to enable self-service
- Simplified data lifecycle**
  - Simplified ingest: Integrated Data Flow and Engineering enables any data ingest at any speed
  - Simplified data exploration: Integrated data catalog enables easy exploration of all data
  - Simplified data visualization: Integrated visualization tool for building faster insights
- Elastic architecture**
  - Simplified capacity planning with automated scale up, down, & auto-suspend to ZERO nodes
  - Guardrails prevent runaway costs
- Contention elimination**
  - Simplify resource allocation and workload optimization with isolated compute resources
  - Simplify SLA adherence, as users access the same data with independent compute clusters
  - Simplify data availability with sharing between experimental and production workloads

- SHARED DATA CATALOG



The screenshot shows the Cloudera Data Catalog interface. On the left, there's a sidebar with a 'Dashboard' button. The main area is titled 'Data Catalog / Dashboard' and shows a dropdown menu set to 'lt-demo-3'. Below it is a search bar. To the right is a table of data. The table has columns for icon, name, description, created date, owner, and type. There are six entries:

	name	description	Created	Owner	Type
	columns_v2	/sys	Created Mon Aug 19 2019	Owner	hive
	bucketing_cols	/sys	Created Mon Aug 19 2019	Owner	hive
	cds	/sys	Created Mon Aug 19 2019	Owner	hive
	part_col_privs	/sys	Created Mon Aug 19 2019	Owner	hive
	db_privs	/sys	Created Mon Aug 19 2019	Owner	hive
	global_privs	/sys	Created Mon Aug 19 2019	Owner	hive

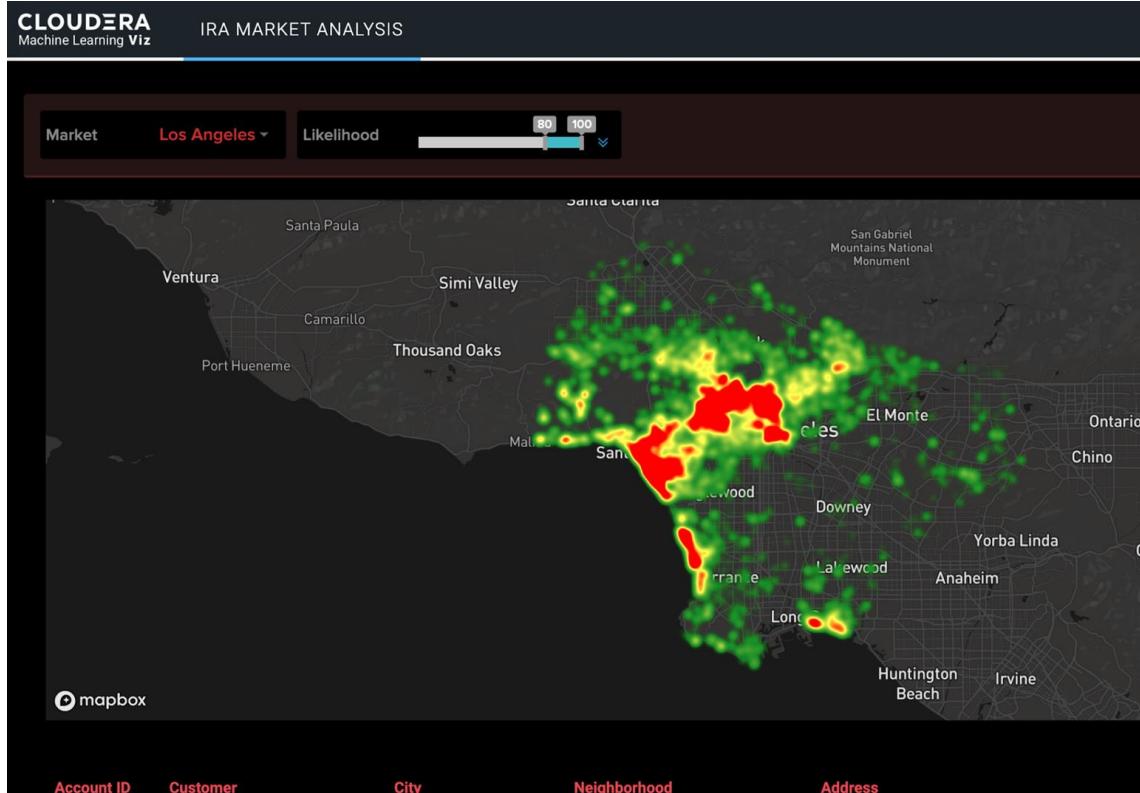
**Filters** (on the left):

- OWNER: Options include 'hive' (53), 'csso\_eqbr' (25), 'csso\_eqbr@CLOUDERA.SITE' (5), and 'impala' (1). Includes 'Add New Value'.
- DATABASE: Includes 'Add New Value'.
- TABLE TAG: Includes 'Add New Value'.
- CREATED WITHIN: Options include 'Last 7 days' and 'Last 15 days'. Includes 'Clear'.

### Organize & curate data

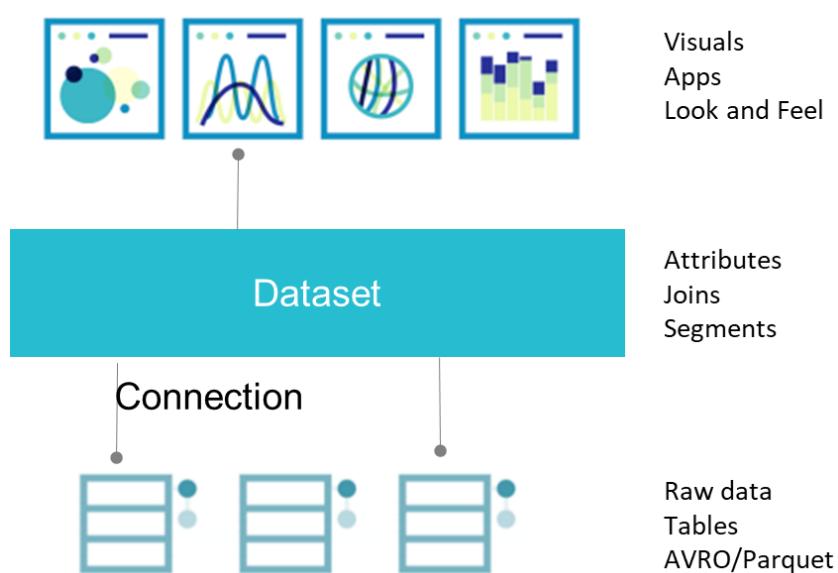
- Find
- Interpret
- Trace
- Audit
- Profile

- VISUALIZATION



Integrated visualization and dashboarding with CDW

- Collection of dashboards styled for the organization that built it
  - Consolidate related visualizations on same screen
  - Easy collaboration by linking and sharing visuals
  - Customized using styles and out of the box settings
- TRADITIONAL & NEW DATA TYPES - ONE SQL STATEMENT



## Data Sets

- Logical collections of data, NOT a copy of the underlying source data
- May be one table or several joined tables
- Used for semantic modeling
- Queries are evaluated at runtime using the most up-to-date data

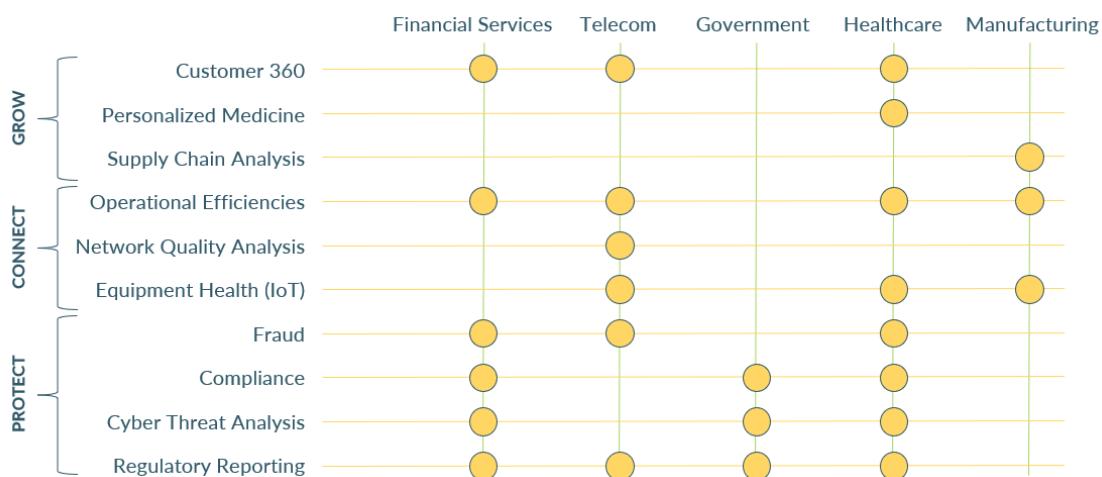
## Data Types

- Same Table - Multiple Open-Source Formats
- Real-Time Data Warehouse (Druid/Kudu): Time-series & Streaming
- Discovery Data Warehouse (Solr): Unstructured & Semi-Structured
- TRANSFORMATIONAL DATA WAREHOUSE FOCUS AREAS



**Use Case Explosion** across all three

- TOP CLOUDERA DATA WAREHOUSE USE CASES BY INDUSTRY



## LAMPIRAN 7: APACHE OZONE

### CUSTOMER USE CASE PERSPECTIVES

- CUSTOMER USE CASE #1: Scale

#### Requirements

- 1000+ nodes, Billions of Files
- Mix of Data Science, Analytics ad hoc querying, ETL ingestion via spark, Hive, HBase, streaming, mapreduce workloads

#### Challenges

- HDFS limit of 400M files results in extreme compaction needs
- Restarts take a very long time

- CUSTOMER USE CASE #2: Density



#### Requirements

- Support dense storage nodes.

#### Challenges

- HDFS block space scaling
  - Maximum blocks per datanode recommended < 10 million
  - Limit of 100 TB per DN (Average file size 10-100 MB)
- Namenode block reports limitations
  - NN IOPS suffer when internal load is too high (Too many FBRs)
  - DN overloaded and slow starts
  - NN performance impacted as density increases

- CUSTOMER USE CASE #3: TCO



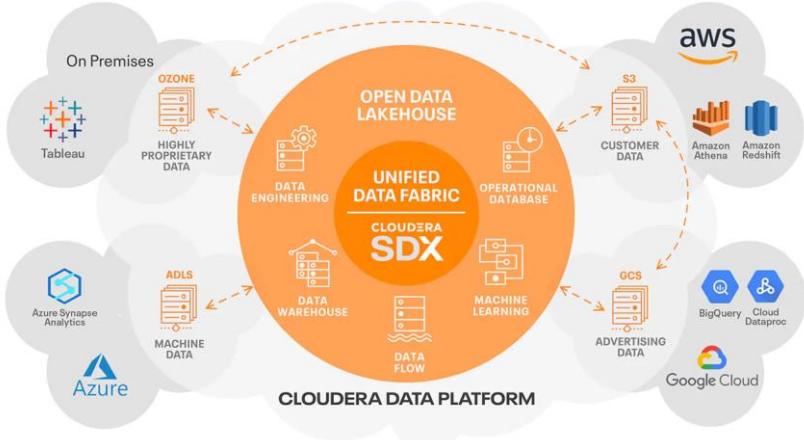
#### Requirements

- Reduction in the storage footprint without sacrificing reliability
- Predictable performance at scale without 3X replication cost

#### Challenges

- Market drivers
  - Chip Shortage after 2020

- Prohibitive cost of DC footprint
- Increased cloud spends with reduced budgets (2022-2023)
- CUSTOMER USE CASE #4: Hybrid



#### Requirements

- Ingestion built for HDFS using Spark, Nifi and Kafka
- DE pipelines written for HDFS - pySpark, Stream sets, Informatica, Talend etc ..
- Data Scientists want to use newer tools (Tensorflow, pySpark, H2O, pyTorch) on the same data

#### Challenges

- Newer Data Science tools support S3 API but not HDFS API
- Either deny access to the data, or replicate the data out of the data lake into a separate S3 data store

- CUSTOMER USE CASE #5: Ease of use

#### Requirements

- Faster restarts and cold starts
- Lesser canonical knowledge to operate
- OOTB integration with compute, warehousing, DE, streaming, analytical engines
- OOTB security and governance

#### Challenges

- HDFS operational limitations
  - JVM fine tuning required, GC overhead
  - Slow restarts - hours for XXL clusters
- External Object Stores limitations
  - Limited OOTB support for authoring DW, DE, ML applications

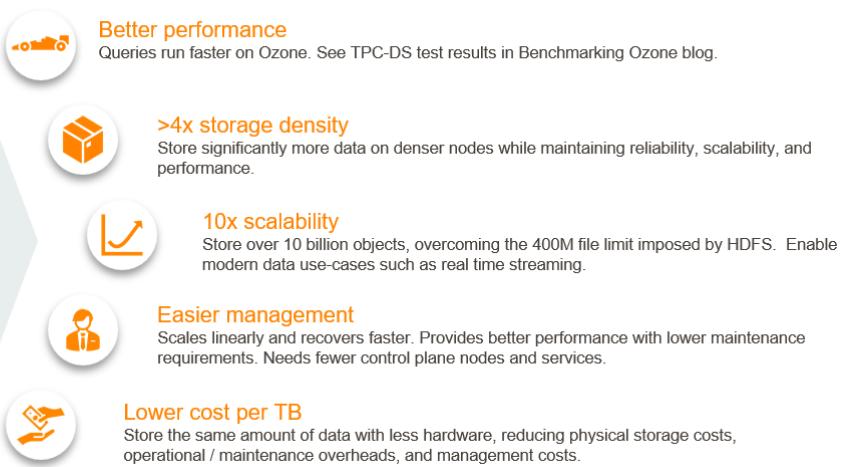
#### OZONE KEY OBJECTIVES

- OZONE DESIGN GOALS
  - Object storage optimized for Analytics
  - Scale to billions of objects of varying sizes
  - Strong Read-After-Write consistency
  - Best next gen storage platform for traditional analytics apps (Spark, Impala, Hive)
  - Modular architecture using industry standard building blocks
  - Preserve and reuse good parts of HDFS
  - Enable new use cases - S3 protocol access, auto-tiering

- Save customers money, denser nodes, large clusters.
- Administrative ease of use (easier HA, better cluster insights via Recon)

#### • KEY BENEFITS

Ozone delivers a vastly superior object storage solution for modern data workloads



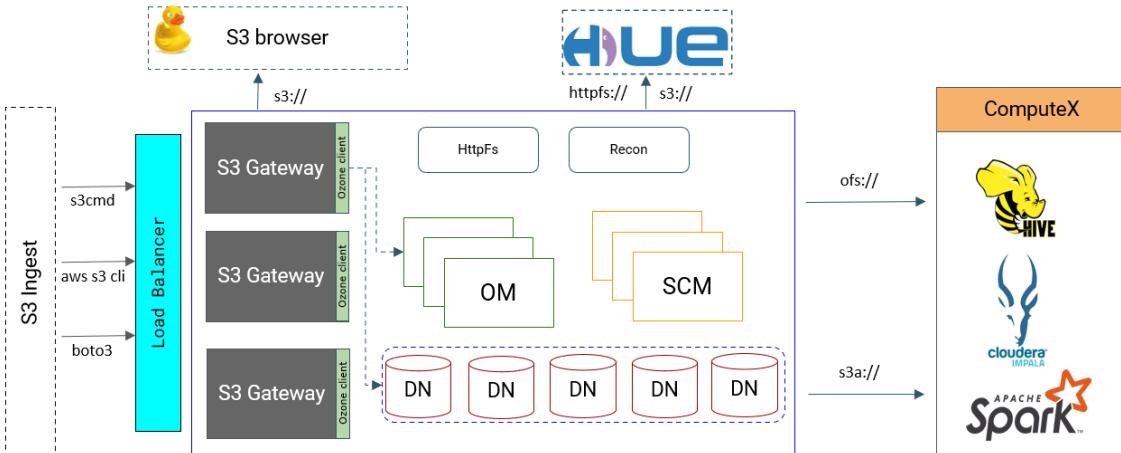
#### • OZONE vs HDFS KEY BENEFITS

Key benefits of Ozone compared to HDFS

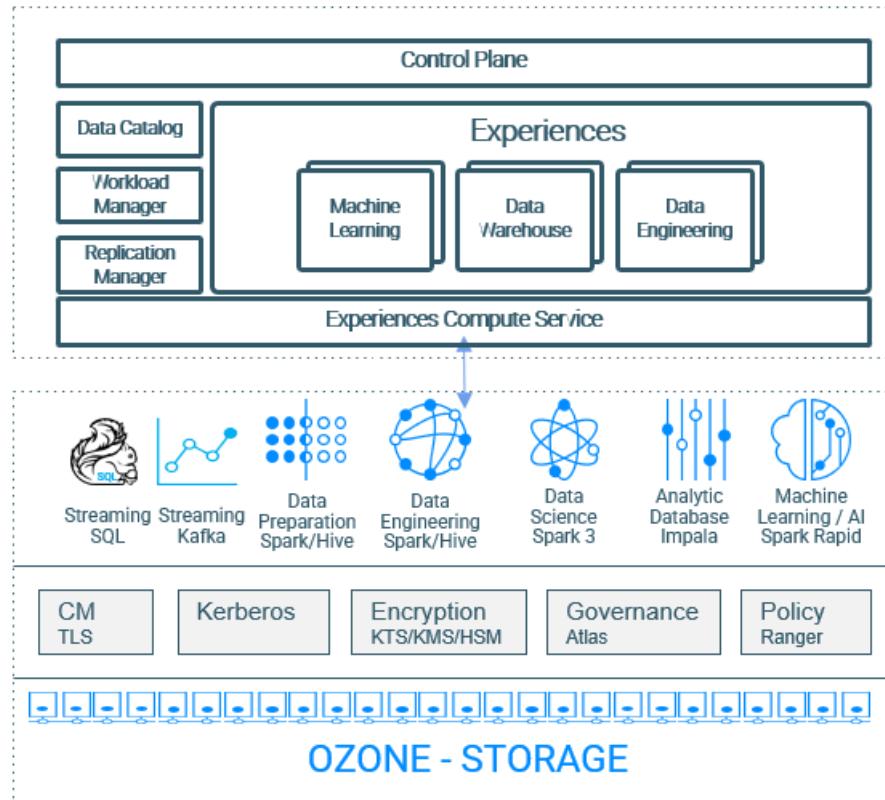
Capability	Ozone	HDFS
Storage Density	1000s of nodes at up to 600TB per node	1000s of nodes at 100TB per node
Scalability	10B Objects	400M Objects
Recovery	Fast recovery (< 5 min restart)	Slow startup based on size
High Availability	Painless deployment, RAFT quorum	Dependency on JN, ZK, ZKFC
Protocol Support	Hadoop and S3 APIs	Hadoop API

#### ARCHITECTURE DECISIONS

##### • OZONE - MULTI-PROTOCOL SUPPORT

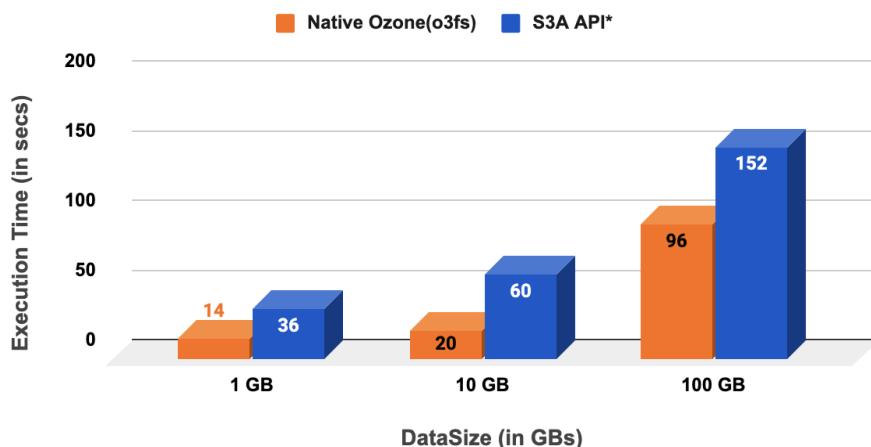


- OOTB - VAST ECOSYSTEM - FULLY INTEGRATED



- Ozone Provides Complete stack integration E2E for the entire data lifecycle
- 2 in 1 - Hybrid Object Store that provides the scale of an Object Store and the benefits of an Hierarchical file store
- Strong security with TLS, TDE and a single policy engine
- Native support for **Hadoop** and **S3 API** that enables cloud native architectures
- OZONE USP: DUALITY AT SCALE

Teragen Performance Comparison with Apache Ozone  
(Lower is Better)



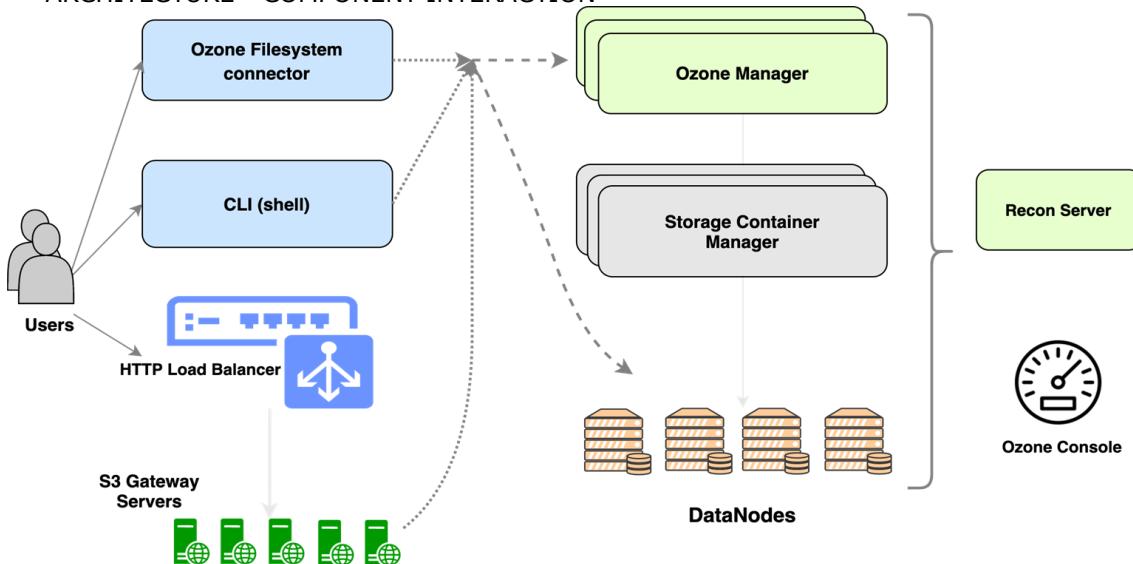
Only object store with native FS support

- Hive, Impala, Spark, MR require delete and rename ops to be strictly **atomic**
- Job committers performance is directly impacted by the **speed** of a delete or a rename op (map/shuffle outputs)

- Other object stores including S3 do not provide such guarantees
- **FileSystem Optimized** (FSO) buckets in Ozone provide  $O(1)$  rename and deletes, atomically like native FS semantics

## HIGH LEVEL ARCHITECTURE

- ARCHITECTURE - COMPONENT INTERACTION

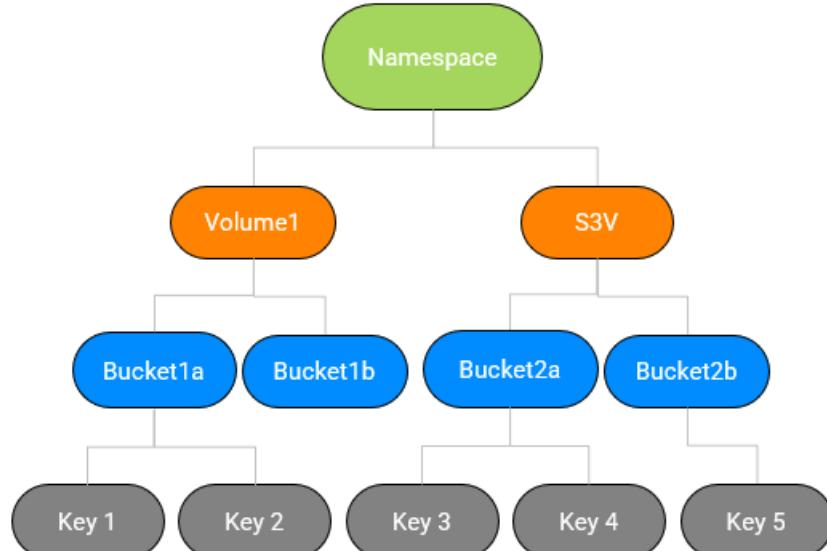


- OZONE BUILDING BLOCKS

Use proven technologies

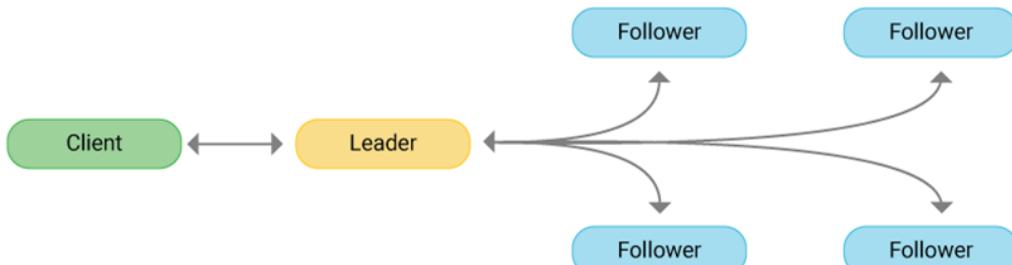
- **RAFT** replication – <http://raft.github.io>
  - Open-source Java implementation of RAFT - Apache Ratis Library
- **Storage Containers** – use off-the-shelf open-source key-value store (**RocksDB**)
  - Supported by, and battle-tested at Facebook.
- **OM** – a namespace manager (also uses RocksDB to store the namespace)
- **HDDS** – a distributed container management layer
- Hadoop security model and Hadoop RPC

- OZONE DATA LAYOUT



- Ozone consists of **volumes, buckets, and keys**
- **Volumes** are similar to user accounts. Only administrators can create or delete volumes
- **Buckets** are similar to Amazon S3 buckets. A bucket can contain any number of keys, but buckets cannot contain other buckets
- **Keys** are similar to files
- The hierarchical file system builds on top of the flat key-value store

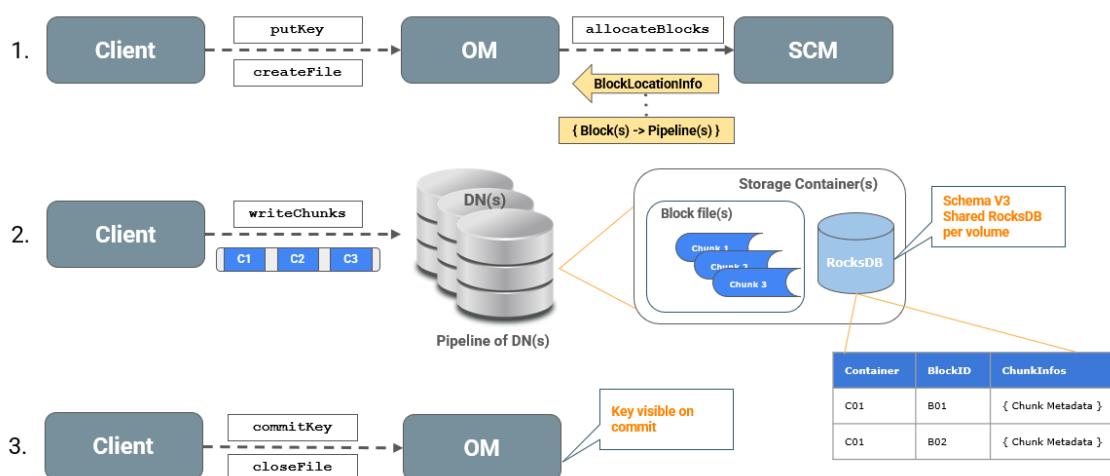
- APACHE RATIS - DISTRIBUTED CONSENSUS



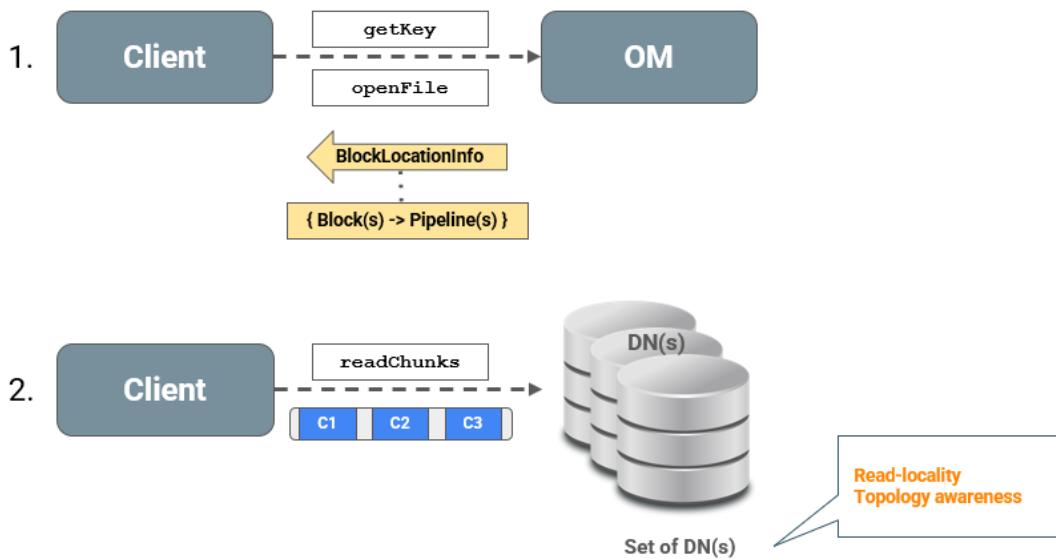
Homegrown performant implementation of the well-known RAFT algorithm

- High-performance **RAFT** implementation - formally proven and understandable
- **Finite State Machines** - distributed across a cluster of computing systems
- Ideal for **serializable updates** aka small transactions
- Extensively used in Ozone for **all metadata transactions** on OM, SCM as well as Datanodes
- Master Nodes use Ratis for HA
- Client uses Ratis library to write data to the containers on the Datanode (**Pipelines ~ RAFT rings**)

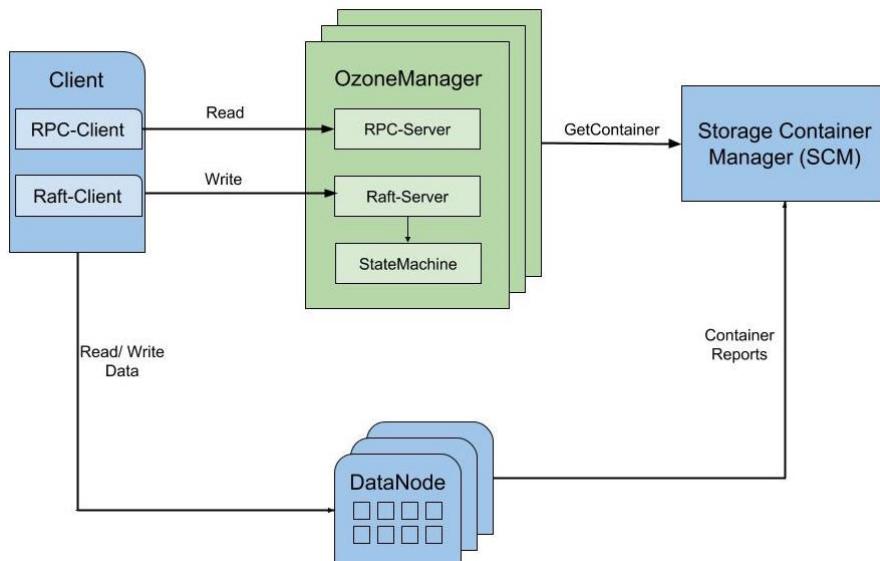
- ANATOMY OF A WRITE



- ANATOMY OF A READ



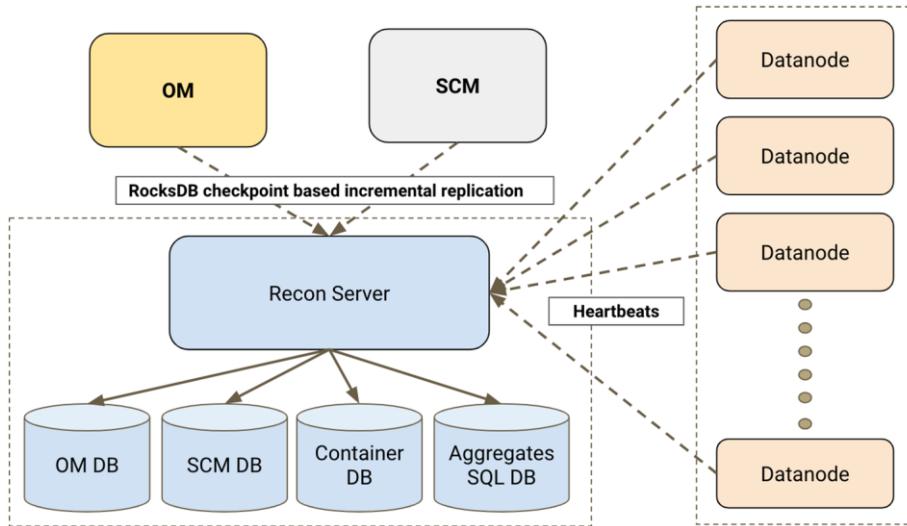
- SIMPLIFIED HIGH AVAILABILITY MODEL



#### OOTB full HA with zero dependencies

- First class HA solution
  - No Journal nodes
  - No Zookeeper
  - No Standby (**Active-Active**)
- Apache Ratis used as a library to keep a **quorum** of OM servers ( $2N+1$ ) in sync
- Fast** failover (seconds)
- Transparent** failover for the application
- All OM State changes go through Ratis
- All write requests go to the Leader OM
- All read requests currently served by Leader OM
- Client caches the last known leader

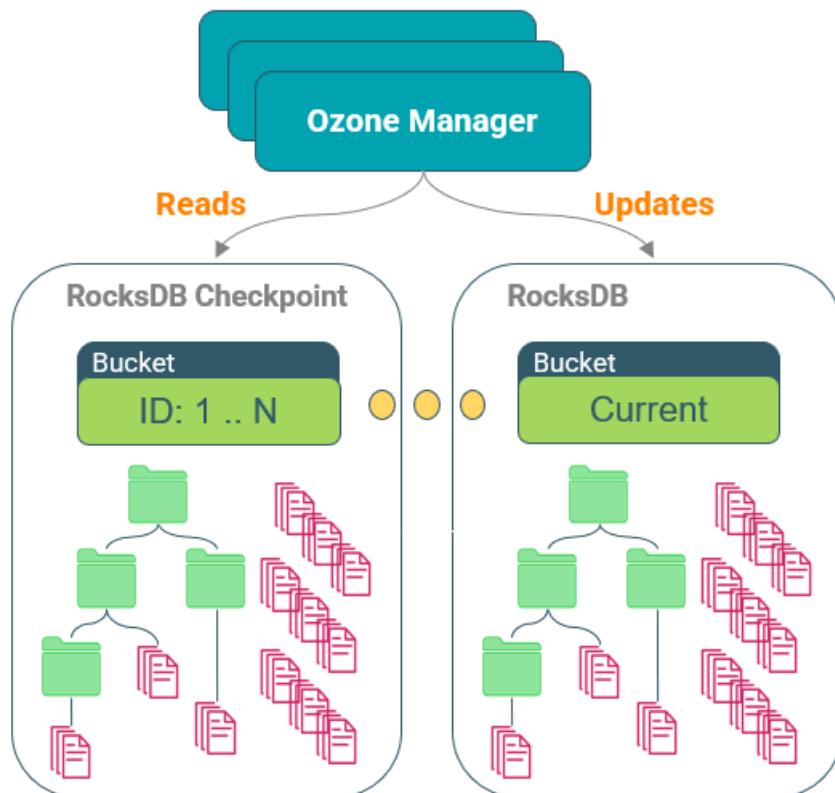
- OZONE RECON



Built-in observability - The monitoring console

- Collects and **aggregates metadata** from OM, SCM and DN heartbeats
  - **Disjoint** sets: {C} -> {K} solution
  - Offloads **heavyweight computations**, e.g.: du, health, heatmaps, etc.
- **Container Health** tracking
  - Missing containers, blocks and keys
  - Replication health: Over/Under/Mis-replicated
- Higher order insights
  - File size distribution
  - Pipeline health and distribution

- OZONE SNAPSHOTS



### Key Differentiator

- Instantaneous snapshots
- Immediately available for reads and restores
- Bucket level (and Volumes)
- Efficient snapshot diffs
- Out of order deletion
- Insights into space utilization
- Simple yet powerful
  - OM metadata path independent
  - Overall code simplicity