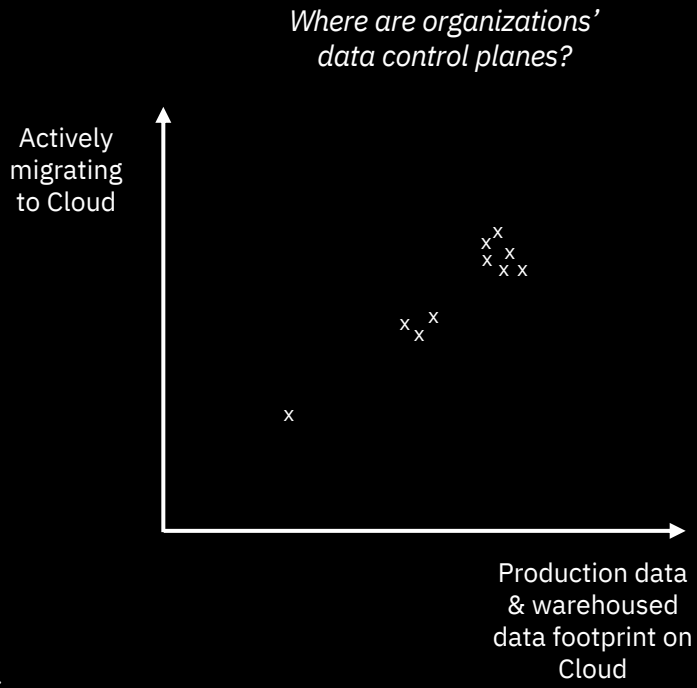
Next Generation DataStage

Value proposition for
Existing DataStage Customers

Market Trends

Today's distributed data processing workloads require a Cloud-native data integration product

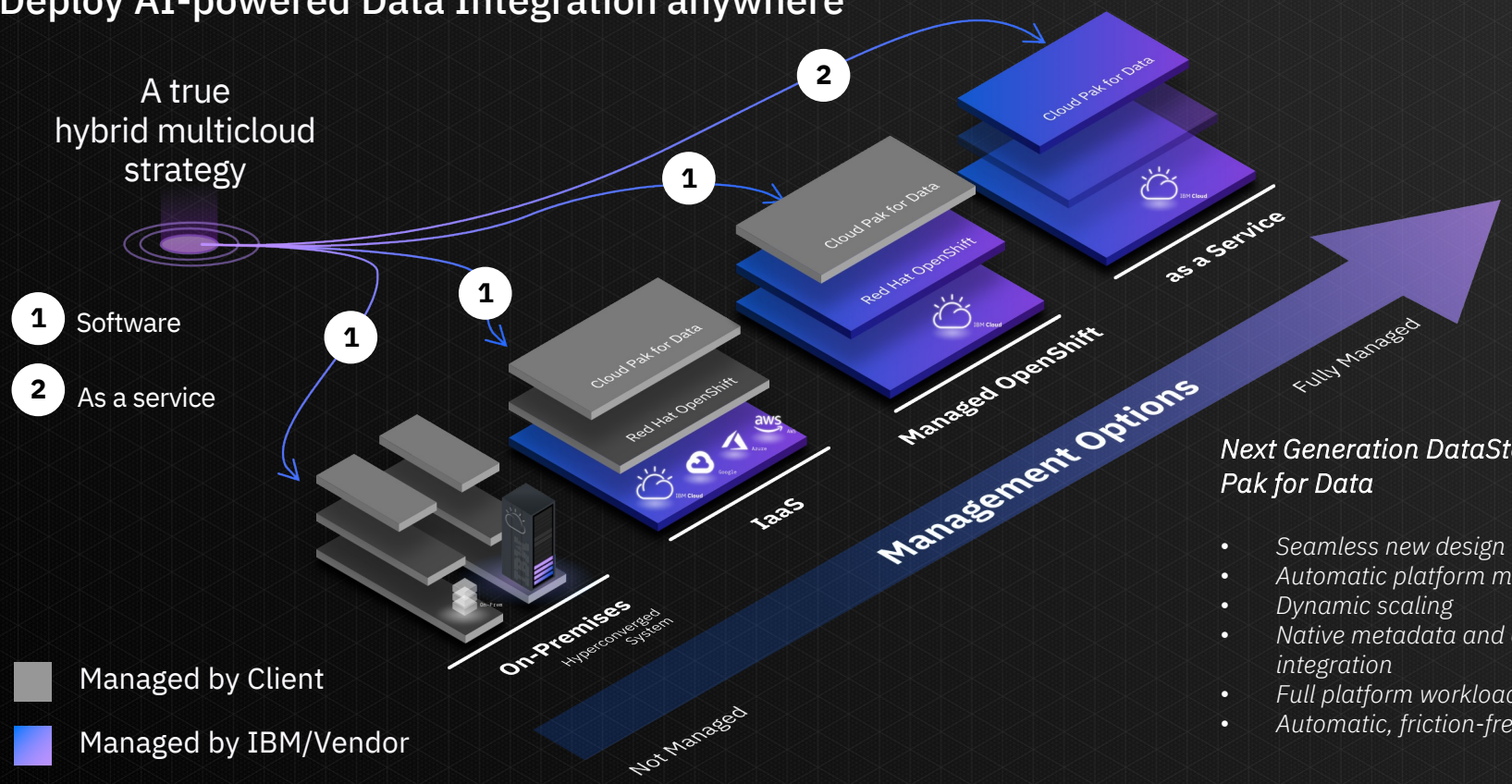


Source:
IDC Cloud Database Survey

- Movement of on-premises data to cloud data warehouses
- Cloud data warehouse modernization & hydration
- Containerized, cloud-native applications and integration capabilities
- Multicloud data integration patterns
- Business-ready data for AI, Analytics, and BI
- CI/CD for faster development

DataStage on IBM Cloud Pak for Data

Deploy AI-powered Data Integration anywhere

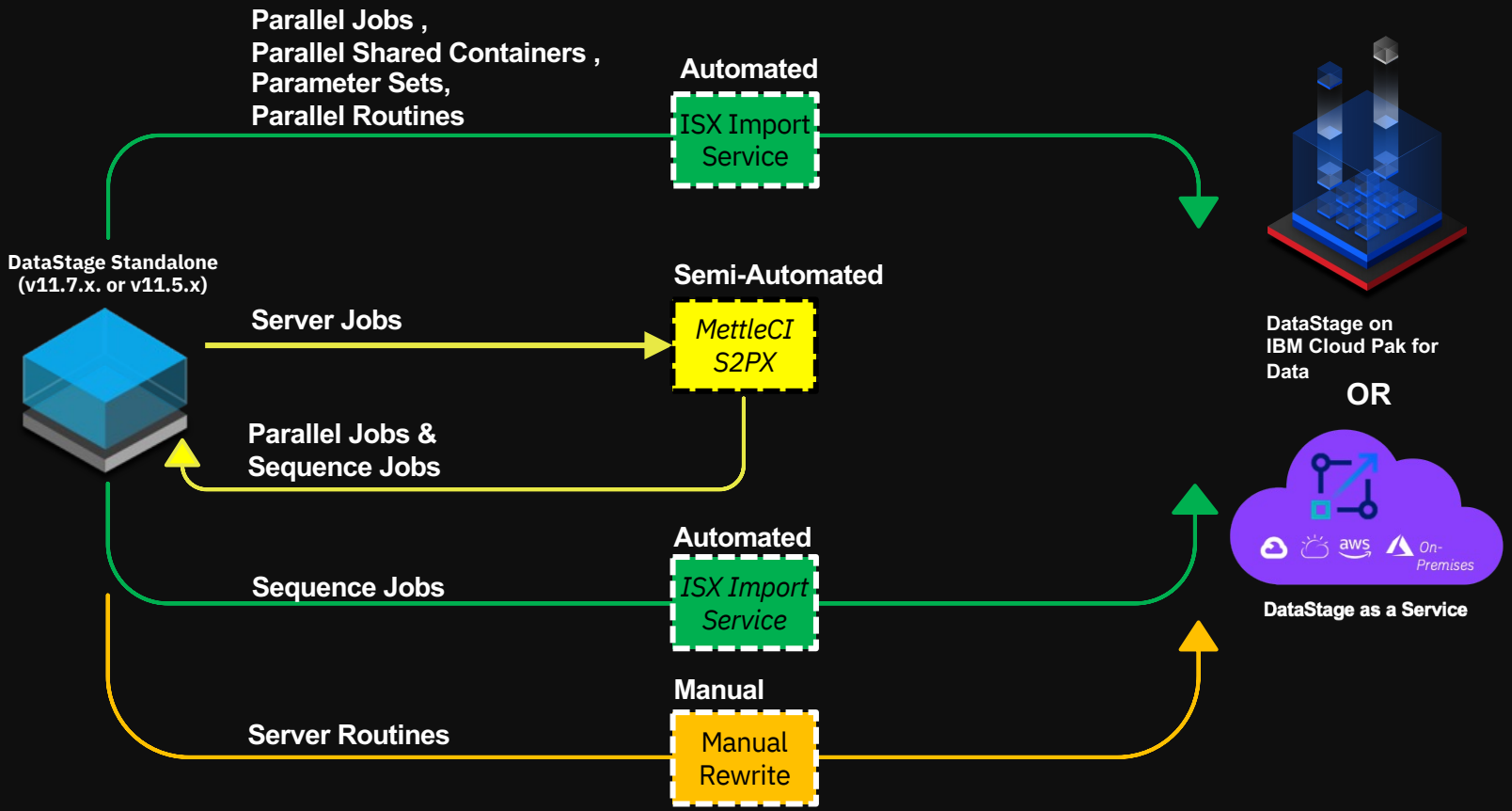


Next Generation DataStage on Cloud Pak for Data

- *Seamless new design experience*
- *Automatic platform management*
- *Dynamic scaling*
- *Native metadata and catalog integration*
- *Full platform workload orchestration*
- *Automatic, friction-free conversion*

On-Premises, any public Cloud

Migration to Next Generation DataStage



Next Generation DataStage:
Modernizing the industry's most performant data integration solution

Leverage existing DataStage investment and skills

- Bring forward existing DataStage parallel jobs with an automatic, friction-free migration.
- No re-training for your developers. Developers can design DataStage flows on a modernized, web-based canvas.

Different data integration styles with polyglot execution engines

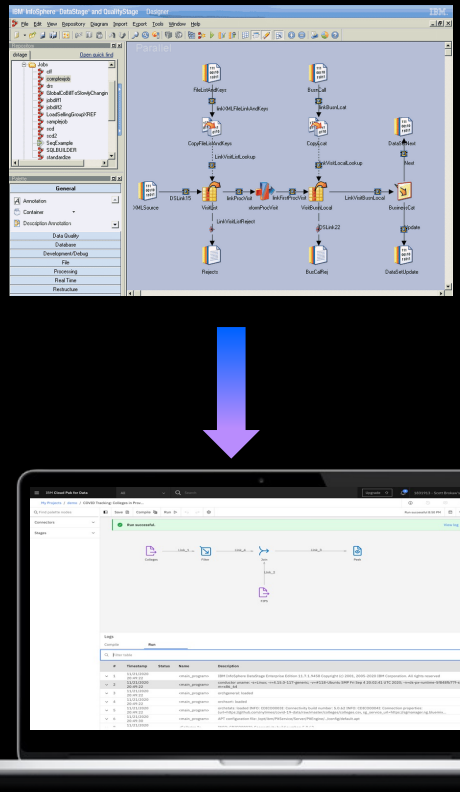
- Combine batch integration with other capabilities in Cloud Pak for Data to actualize data integration styles like parallel processing, virtualization, replication, streaming, and preparation.

Multicloud scalability, elasticity, and runtime execution

- Design once, dynamically run anywhere with built-in automatic workload balancing, parallelism and dynamic scalability.
- Increase data gravity with distributed processing for hybrid-cloud or multicloud requirements.

Integrated with IBM's data and AI ecosystem

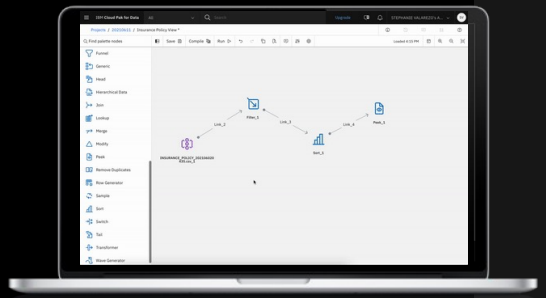
- Common canvas on Cloud Pak for Data.
- Data integration, machine learning, data science.

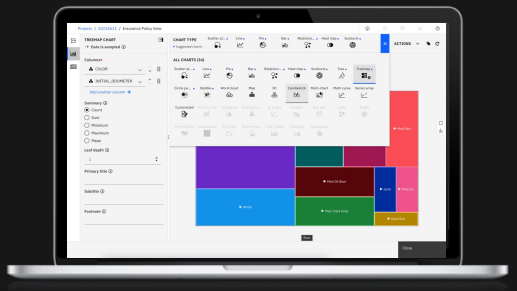


Why DataStage on Cloud Pak for Data?

Benefits and outcomes of containerized DataStage

- Design at speed through smart automation and high level of reusability
- No additional development cost when scaling out to new environments
- Runtime independence through Enterprise container platform foundation (also enables automatic workload balancing)
- Cost, speed and scale optimized data integration
- Increased compliance and deploy @ scale operate through full CI/CD integration
- Significantly reduce effort in management and operation





Key customer benefits

5x

Faster execution than on Spark parallel engine

9x

Faster design than hand-coding

85%

Reduction in infrastructure management time & effort

Next Generation DataStage

Benefits you can measure

85%
Reduction in infrastructure management time and effort

87%
Savings in development cost when using visual design compared to hand coding

30%
Throughput execution improvement over standalone DataStage

60%
Increased efficiency (minimal upgrade efforts)

50%+
Reduction in development cost when running on multiple (cloud) environments utilizing design once - run anywhere feature

The value proposition with Next Generation DataStage

There are 2 main components involved for calculating the value proposition of Next Generation DataStage versus on-prem standalone DataStage

- Current running cost of standalone DataStage versus the 2 implementation options of DataStage on Cloud Pak for Data software and DataStage on Cloud Pak for Data as a service

- Cost of one-time migration of DataStage jobs to Next Generation DataStage. There are multiple tools provided to understand complexity and parity

Once the above can be estimated, it is possible to do a Business Value Assessment. The following slides provide an overview on the above 2 aspects

Calculating the value proposition – potential parameters considered

Software costs charges for S&S, software charges or SaaS charges

Upgrades
hardware upgrade charges

Citrix
charges for using DataStage Designer on a fat client

Designer setup
installing DataStage designer for Windows

Ad hoc requests
changes and additional operations costs

MettleCI
Continuous Integration tool for build and deployment

Hardware
hardware cost for on prem DataStage

Administration
cost to administer DataStage

Software Patch
patches to upgrade the software

Cloud infra costs
Storage costs, EC2, ROSA, OCP, ELBs etc

- Value proposition will vary based on deployment option
- One time migration cost calculated separately
- There can be additional parameters considered as well like use of Data Refinery etc

Business Value Assessment - comparing deployment options

Output cost option estimates - visual example:

Current DataStage on-prem		DataStage CP4D software		DataStage CP4D as a service	
Annual IBM S & S	$1,42,880	Annual S/W charges	$2,05,000	Annual SaaS charges	$1,65,000
People, Processes, Hardware Upgrades	$36,000	People, Processes, Hardware			
Citrix	$90,000	AWS CP4D/RHOS infra EC2,S3,ELB,Support	$59,000		
Designer setup	$20,000				
Ad hoc requests	$80,000	Ad hoc requests	$80,000		
Administration	$20,000	Administration	$20,000		
Patch installation	$3,000	Patch installation	$3,000		
Hardware	$120,000	One time service migration	TBD	One time service migration	TBD
MettleCI	$50,000				
Subtotal	**$561,880**	**Subtotal**	**$367,000 + services**	**Subtotal**	**$165000 + services**
Three year price	**$1,685,640**	**Three year price**	**$1,101,000 + one time migration services (TBD)**	**Three year price**	**$495,000 + one time migration services (TBD)**

Business Value Assessment - Assessment for a customer

The below is an illustration of typical savings arrived post taking inputs from the customer

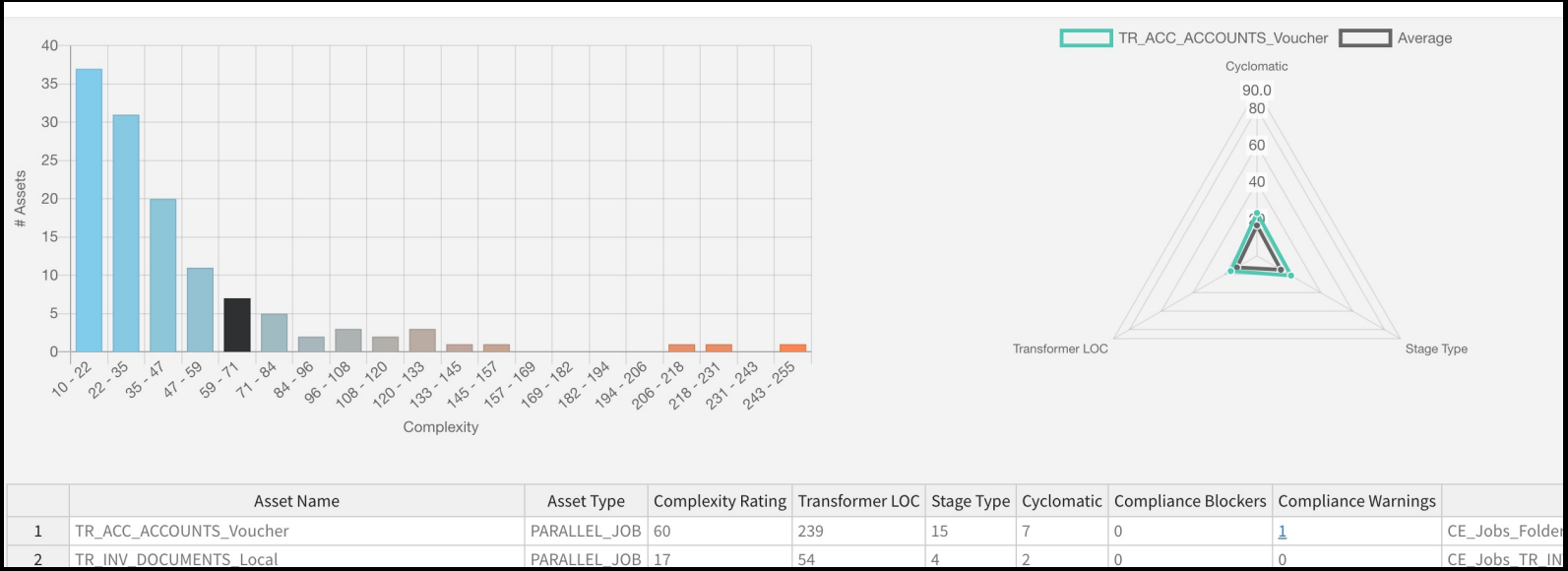
$1.2 Million USD Estimated Annual savings with DataStage NextGen

- **$700,000 USD** in Citrix savings (upgrade/prep/testing and deployment)
- **$200,000 USD** in migration savings with MettleCI
- **$100,000 USD** in production environment downtime and recovery savings
- **$160,000 USD** in business requests turnaround time savings with Data Refinery
- **$40,000 USD** in logistics savings (shuttling jobs, repeat testing)

Modernization assessment – understanding the migration complexity

- **DataStage Report card** - provides on idea on complexity of the on-prem DataStage assets to be migrated.

The link to upload the file is here if client is sharing .isx file.
Technical details provided in Appendix section - DataStage project report card



	Asset Name	Asset Type	Complexity Rating	Transformer LOC	Stage Type	Cyclomatic	Compliance Blockers	Compliance Warnings	
1	TR_ACC_ACCOUNTS_Voucher	PARALLEL_JOB	60	239	15	7	0	1	CE_Jobs_Folder
2	TR_INV_DOCUMENTS_Local	PARALLEL_JOB	17	54	4	2	0	0	CE_Jobs_TR_IN

- Graph Analysis
- Computational Complexity
- Compliance Rules

DataStage Report card

Modernization assessment – understanding the migration complexity continued

XMETA Query for stage analysis - Perform an asset query on the metadata repository XMETA if .isx export file is not provided by client. The details on XMETA stage type and the count is got at project level. This is less detailed than the earlier DataStage project report card.
Technical details and sample output provided in Appendix section - XMETA Query for stage analysis

IBM stage review of project in CP4DaaS - IBM can do an analysis of all stages in the project and addresses cases where there is **missing parity** by providing milestone dates.
Details and sample output provided in Appendix section - IBM stage review of project in CP4DaaS

Capturing Workload to convert to SaaS capacity unit hours Model - estimate the amount of compute a customer would need to run their jobs in DataStage as a Service.
Technical details and sample output provided in Appendix section - SaaS consumption historical job run information

IBM Expert Labs engagement – A 6 week pilot engagement. Modernize and validate representative subset of DataStage assets. Also provides plan to complete modernization

APPENDIX

DataStage project report card

Typical sample with details of existing on prem DataStage assets , complexity rating etc by project
Can be used for estimating the migration complexity

Asset Name	Asset Type	Complexity Rating	Transformer LOC	Stage Type	Cyclomatic	Compliance Blockers	Compliance	Archive	Engine	Project
TR_ACC_ACCOUNTS_Voucher	PARALLEL_JOB	60	239	15	7	0	1	CE_Jobs_Folder1_TR_ACC_ACCOUNTS_Voucher.isx	x	mci_testing_master
TR_INV_DOCUMENTS_Local	PARALLEL_JOB	17	54	4	2	0	0	CE_Jobs_TR_INV_DOCUMENTS_Local.isx	TEST1-ENGN.DATAMIGRATORS.IO	mci_testing_master
TR_INV_CUSTOMER_INVOICES_Debtor	PARALLEL_JOB	18	100	3	2	0	0	CE_Jobs_TR_INV_CUSTOMER_INVOICES_Debtor.isx	TEST1-ENGN.DATAMIGRATORS.IO	mci_testing_master
TR_ACC_ACCOUNTS_TOLLPRODUCTS_Lo	PARALLEL_JOB	18	60	5	2	0	0	CE_Jobs_Folder1_TR_ACC_ACCOUNTS_TOLLPRODUC	TEST1-ENGN.DATAMIGRATORS.IO	mci_testing_master
TR_INV_POSTPAID_EVENTS_MUF	PARALLEL_JOB	49	182	11	6	0	3	CE_Jobs_TR_INV_POSTPAID_EVENTS_MUF.isx	TEST1-ENGN.DATAMIGRATORS.IO	mci_testing_master
TR_ACC_ACCOUNTS_Interop	PARALLEL_JOB	40	198	9	4	0	0	CE_Jobs_Folder1_TR_ACC_ACCOUNTS_Interop.isx	TEST1-ENGN.DATAMIGRATORS.IO	mci_testing_master
TR_INV_POSTPAID_EVENTS_Debtor_Bill	PARALLEL_JOB	33	130	8	4	0	1	CE_Jobs_TR_INV_POSTPAID_EVENTS_Debtor_Billed_	TEST1-ENGN.DATAMIGRATORS.IO	mci_testing_master
TR_ACC_ACC_CUST_PAYMETHODS_Loca	PARALLEL_JOB	16	44	4	2	0	0	CE_Jobs_Folder1_TR_ACC_ACC_CUST_PAYMETHODS_	TEST1-ENGN.DATAMIGRATORS.IO	mci_testing_master
TR_INV_PREPAID_EVENTS_Debtor_Bille	PARALLEL_JOB	32	170	7	3	0	0	CE_Jobs_TR_INV_PREPAID_EVENTS_Debtor_Billed_F	TEST1-ENGN.DATAMIGRATORS.IO	mci_testing_master
TR_ACC_VEHICLES_Interop	PARALLEL_JOB	18	97	3	2	0	0	CE_Jobs_Folder2_TR_ACC_VEHICLES_Interop.isx	TEST1-ENGN.DATAMIGRATORS.IO	mci_testing_master
TR_LOG_TAG_ORDERS	PARALLEL_JOB	41	164	12	4	0	1	CE_Jobs_TR_LOG_TAG_ORDERS.isx	TEST1-ENGN.DATAMIGRATORS.IO	mci_testing_master
TR_DEB_NOMINATIONS	PARALLEL_JOB	38	155	5	6	0	2	CE_Jobs_TR_DEB_NOMINATIONS.isx	TEST1-ENGN.DATAMIGRATORS.IO	mci_testing_master
TR_BIL_DEDUCTION_FEE_EVENTS_Loca	PARALLEL_JOB	72	263	19	8	0	4	CE_Jobs_TR_BIL_DEDUCTION_FEE_EVENTS_Local_M	TEST1-ENGN.DATAMIGRATORS.IO	mci_testing_master
TR_BIL_FUNDS_EVENTS_RefundsTopup	PARALLEL_JOB	20	135	3	2	0	0	CE_Jobs_TR_BIL_FUNDS_EVENTS_RefundsTopups.isx	TEST1-ENGN.DATAMIGRATORS.IO	mci_testing_master
TR_BIL_DEDUCTION_FEE_EVENTS_Debt	PARALLEL_JOB	106	427	24	13	0	2	CE_Jobs_Folder2_TR_BIL_DEDUCTION_FEE_EVENTS_	TEST1-ENGN.DATAMIGRATORS.IO	mci_testing_master
TR_ACC_ACCOUNT_STATUS_HISTORY_L	PARALLEL_JOB	10	n/a	2	2	0	0	CE_Jobs_Folder1_TR_ACC_ACCOUNT_STATUS_HISTC	TEST1-ENGN.DATAMIGRATORS.IO	mci_testing_master
TR_ACC_ADDITIONAL_CONTACTS	PARALLEL_JOB	79	167	28	8	0	2	CE_Jobs_Folder1_TR_ACC_ADDITIONAL_CONTACTS.is	TEST1-ENGN.DATAMIGRATORS.IO	mci_testing_master
TR_ACC_VEHICLES_Debtors	PARALLEL_JOB	25	80	6	3	0	0	CE_Jobs_Folder2_TR_ACC_VEHICLES_Debtors.isx	TEST1-ENGN.DATAMIGRATORS.IO	mci_testing_master
TR_ACC_TAG_LOCATIONS	PARALLEL_JOB	28	72	9	3	0	1	CE_Jobs_Folder2_TR_ACC_TAG_LOCATIONS.isx	TEST1-ENGN.DATAMIGRATORS.IO	mci_testing_master
TR_ACC_DUNNING_LOCKS_Postpaid	PARALLEL_JOB	19	69	3	3	0	0	CE_Jobs_Folder2_TR_ACC_DUNNING_LOCKS_Postpai	TEST1-ENGN.DATAMIGRATORS.IO	mci_testing_master
TR_ACC_TAGS_Interop	PARALLEL_JOB	38	121	9	5	0	1	CE_Jobs_Folder2_TR_ACC_TAGS_Interop.isx	TEST1-ENGN.DATAMIGRATORS.IO	mci_testing_master
TR_BIL_DEDUCTION_FEE_EVENTS_Loca	PARALLEL_JOB	36	186	9	3	0	0	CE_Jobs_TR_BIL_DEDUCTION_FEE_EVENTS_Local_In	TEST1-ENGN.DATAMIGRATORS.IO	mci_testing_master
TR_ACC_VEHICLES_Local	PARALLEL_JOB	16	77	3	2	0	0	CE_Jobs_Folder2_TR_ACC_VEHICLES_Local.isx	TEST1-ENGN.DATAMIGRATORS.IO	mci_testing_master
TR_ACC_DUNNING_LOCK_HISTORY_Adc	PARALLEL_JOB	42	214	10	4	0	0	CE_Jobs_Folder1_TR_ACC_DUNNING_LOCK_HISTORY	TEST1-ENGN.DATAMIGRATORS.IO	mci_testing_master
TR_DEB_ENFORCEMENT_RECORDS	PARALLEL_JOB	41	219	6	5	0	1	CE_Jobs_TR_DEB_ENFORCEMENT_RECORDS.isx	TEST1-ENGN.DATAMIGRATORS.IO	mci_testing_master
TR_INV_CHARGE_ORDERS_Debtors	PARALLEL_JOB	18	101	3	2	0	0	CE_Jobs_TR_INV_CHARGE_ORDERS_Debtors.isx	TEST1-ENGN.DATAMIGRATORS.IO	mci_testing_master

XMETA Query for stage analysis

Provide this SQL query to the customer, to run against XMETA. Customer can perform this query by DataStage project

```
SELECT J.DSNAMESPACE_XMETA PROJECT, J.NAME_XMETA JOBNAME, S.STAGETYPE_XMETA STAGE, COUNT(STAGETYPE_XMETA) AS COUNT FROM
XMETA.DATASTAGEX_DSSTAGE S INNER JOIN XMETA.DATASTAGEX_DSJOBDEF J ON J.XMETA_REPOS_OBJECT_ID_XMETA = S.CONTAINER_RID WHERE
J.DSNAMESPACE_XMETA LIKE '%${project}' GROUP BY J.NAME_XMETA, S.STAGETYPE_XMETA, J.DSNAMESPACE_XMETA ORDER BY COUNT DESC;
```

```
PROJECT,STAGETYPE_XMETA,COUNT
LINUX-HOST:ADMIN,CEndLoopActivity,3
LINUX-HOST:ADMIN,CExceptionHandler,3
LINUX-HOST:ADMIN,CExecCommandActivity,6
LINUX-HOST:ADMIN,CNotificationActivity,3
LINUX-HOST:ADMIN,CStartLoopActivity,3
LINUX-HOST:ADMIN,CTerminatorActivity,3
LINUX-HOST:ADMIN,CTransformerStage,19
LINUX-HOST:ADMIN,DB2ConnectorPX,3
LINUX-HOST:ADMIN,NetezzaConnectorPX,6
LINUX-HOST:ADMIN,OracleConnectorPX,1
LINUX-HOST:ADMIN,PxAggregator,6
LINUX-HOST:ADMIN,PxCopy,8
LINUX-HOST:ADMIN,PxExternalSource,1
LINUX-HOST:ADMIN,PxFunnel,2
LINUX-HOST:ADMIN,PxLookup,8
LINUX-HOST:ADMIN,PxPeek,7
LINUX-HOST:ADMIN,PxRemDup,7
LINUX-HOST:ADMIN,PxRowGenerator,5
LINUX-HOST:ADMIN,PxSequentialFile,7
```

IBM stage review of project in CP4DaaS

Besides stage counts, Milestone Name and Milestone Date for each stage is also provided in the report

Archive,Engine,Project,StageType,Count,Milestonename,Milestonedate
EXPORT.isx,LINUX:31538,DW_DB,PIVOT ENTERPRISE (PARALLEL),11,Oct-21,10/31/21
EXPORT.isx,LINUX:31539,DW_DB,SEQUENTIAL FILE (PARALLEL),737,Mar-21,3/31/21
EXPORT.isx,LINUX:31540,DW_DB,COPY (PARALLEL),28,Launch DataStage aaS Beta,11/30/20
EXPORT.isx,LINUX:31541,DW_DB,ROW GENERATOR (PARALLEL),49,Launch DataStage aaS Beta,11/30/20
EXPORT.isx,LINUX:31542,DW_DB,AGGREGATOR (PARALLEL),26,Feb-21,2/26/21
EXPORT.isx,LINUX:31543,DW_DB,TRANSFORMER (PARALLEL),623,DataStage SaaS GA,6/18/21
EXPORT.isx,LINUX:31544,DW_DB,DATASET (PARALLEL),109,DataStage SaaS GA,6/18/21
EXPORT.isx,LINUX:31545,DW_DB,FUNNEL (PARALLEL),31,DataStage SaaS GA,6/18/21
EXPORT.isx,LINUX:31546,DW_DB,FILTER (PARALLEL),27,DataStage SaaS GA,6/18/21
EXPORT.isx,LINUX:31547,DW_DB,REMOVE DUPLICATES (PARALLEL),79,DataStage SaaS GA,6/18/21
EXPORT.isx,LINUX:31548,DW_DB,SORT (PARALLEL),3,DataStage SaaS GA,6/18/21
EXPORT.isx,LINUX:31549,DW_DB,JOIN (PARALLEL),83,DataStage SaaS GA,6/18/21
EXPORT.isx,LINUX:31550,DW_DB,MERGE (PARALLEL),1,DataStage SaaS GA,6/18/21
EXPORT.isx,LINUX:31551,DW_DB,SHARED CONTAINER (PARALLEL),18,Aug-21,8/31/21
EXPORT.isx,LINUX:31552,DW_DB,STORED PROCEDURE (PARALLEL),3,Q4 2021,12/31/21
EXPORT.isx,LINUX:31553,DW_DB,UNSTRUCTURED DATA (PARALLEL),1,Q3 2021,9/30/21
EXPORT.isx,LINUX:31554,DW_DB,CHANGE CAPTURE (PARALLEL),24,Q2 2021,6/30/21
EXPORT.isx,LINUX:31555,DW_DB,TERMINATOR ACTIVITY (SEQUENCE),488,Nov-21,11/30/21

SaaS consumption historical job run information

To be run against the DataStage Operations database (DSODB) in all environments

Script Details:

--Display DataStage job run details by Year and Month since 2021-12-01

--Excludes ISD jobs or incomplete jobs

Note: Please include headers

SQL for Db2 DSODB:

```
SELECT YEAR(T2.RUNENDTIMESTAMP) AS RUNYEAR, MONTH(T2.RUNENDTIMESTAMP) AS RUNMONTH, T1.JOBID,  T1.HOSTID,  T1.PROJECTNAME,  T1.JOBNAME,  T1.JOBTYPE,
T2.RUNSTARTTIMESTAMP,  T2.RUNENDTIMESTAMP,  T2.ELAPSEDRUNSECS,  T2.RUNMINORSTATUS,  T2.TOTALROWSCONSUMED,  T2.TOTALROWSPRODUCED,  T2.TOTALCPU,
T2.CONFIGFILENAME,  T2.TOTALPHYSICALNODES,  T2.TOTALLOGICALNODES FROM DSODB.JOBEXEC T1,      DSODB.JOBRUN T2 WHERE T1.JOBID = T2.JOBID AND
T1.WEBSERVICESENABLED = 0 AND T2.RUNENDTIMESTAMP IS NOT NULL  AND T2.RUNENDTIMESTAMP >= CAST('2021-12-01' as timestamp) ORDER BY RUNYEAR DESC,
RUNMONTH DESC, PROJECTNAME, JOBTYPE, TOTALLOGICALNODES, JOBNAME
```

SQL for Oracle DSODB:

```
SELECT EXTRACT(YEAR FROM T2.RUNENDTIMESTAMP) AS RUNYEAR, EXTRACT(MONTH FROM T2.RUNENDTIMESTAMP) AS RUNMONTH, T1.JOBID, T1.HOSTID,
T1.PROJECTNAME, T1.JOBNAME, T1.JOBTYPE, T2.RUNSTARTTIMESTAMP, T2.RUNENDTIMESTAMP, T2.ELAPSEDRUNSECS, T2.RUNMINORSTATUS, T2.TOTALROWSCONSUMED,
T2.TOTALROWSPRODUCED, T2.TOTALCPU, T2.CONFIGFILENAME, T2.TOTALPHYSICALNODES, T2.TOTALLOGICALNODES
FROM DSODB.JOBEXEC T1, DSODB.JOBRUN T2
WHERE T1.JOBID = T2.JOBID
AND T1.WEBSERVICESENABLED = 0
 AND T2.RUNENDTIMESTAMP IS NOT NULL
 AND T2.RUNENDTIMESTAMP >= CAST('01-DEC-2021' AS TIMESTAMP WITH LOCAL TIME ZONE)
ORDER BY RUNYEAR DESC, RUNMONTH DESC, PROJECTNAME, JOBTYPE, TOTALLOGICALNODES, JOBNAME
```

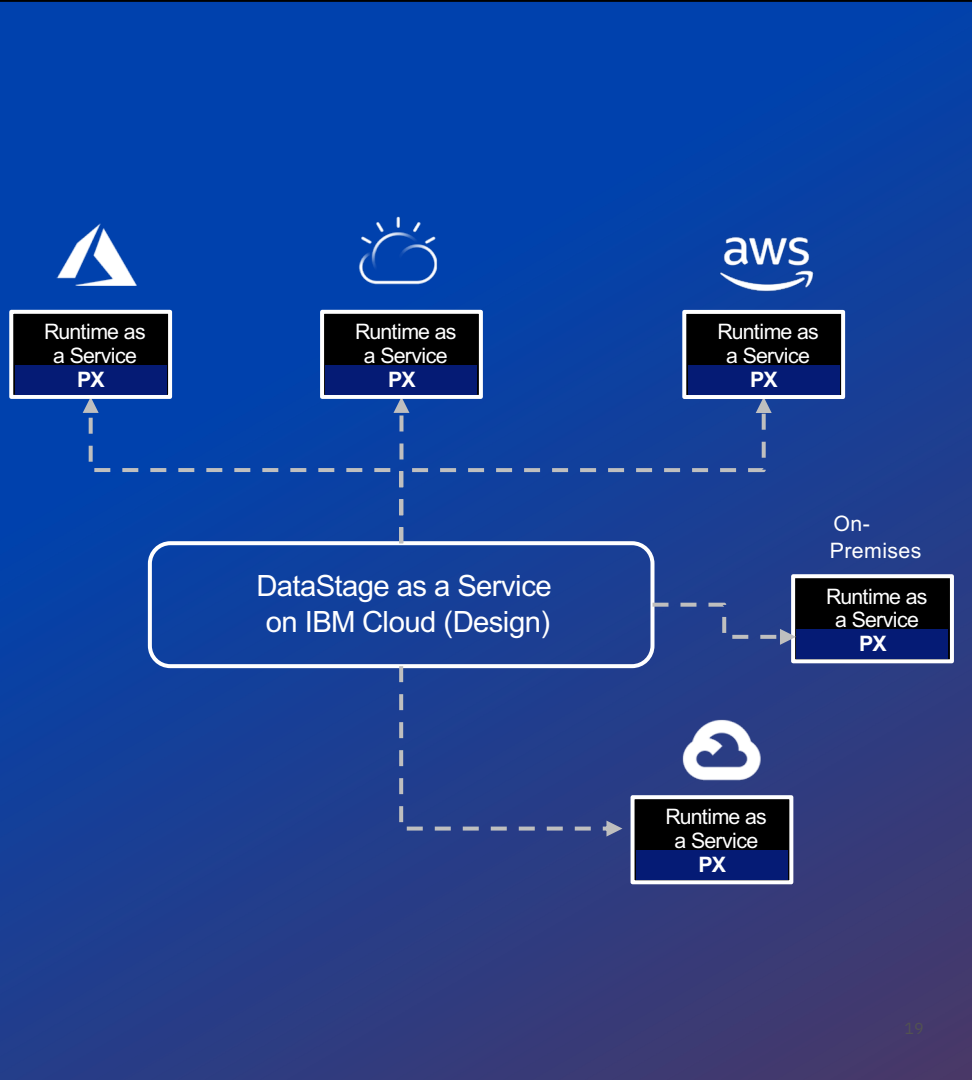
DataStage as a Service

Designed and built on cloud-native principles

- Provision the SaaS service to get started – **no install or configuration needed**
- **No upgrades needed** – fully managed
- DataStage data pipelines can be built and deployed in **minutes**
- **New user experience** with productivity enhancements for developers
- **Connect** to data stores you need to build DataStage flows
- **Elastic scaling** so you scale quickly and meet your workload requirements
- **Consumption based** pricing

Accelerate time to value and reduce TCO

Flexible, elastic scaling based on workload requirements



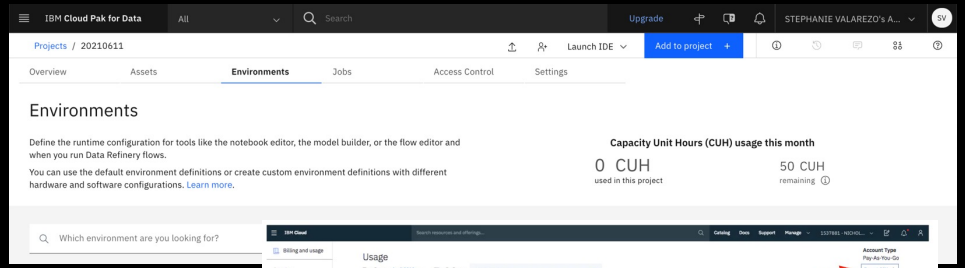
Consumption-based model for DataStage as a Service

Capacity Unit Hours-based

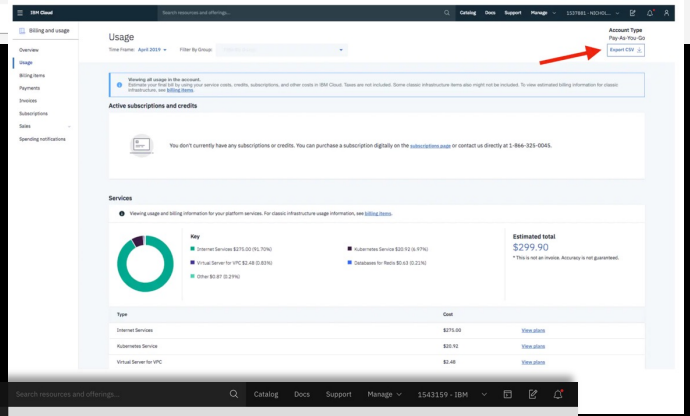
Single pane of glass from IBM Cloud console

- Dashboard for operations across environments
- Identity, key & certificate management
- Observability: Central logging & monitoring for apps and platform

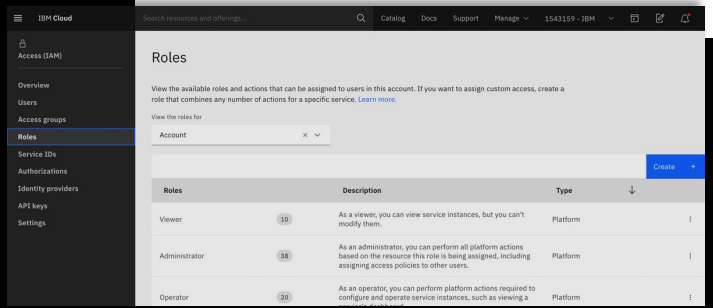
Auditable inventory of Cloud environment



Track usage by environment



Metered cloud services spend



Centralized administration dashboard for user roles

Consumption-based model benefits

Reduced up-front and variable costs

- Pay only for what you use with consumption-based pricing
  - Flexible, elastic scaling to meet workload requirements
  - No excess or unused capacity
- Increase productivity and automation
  - Focusing on building and deploying data pipelines
  - Reduce IT management activities
- Eliminate costs for infrastructure management
  - Reduce costs for monitoring, maintenance, license spend
  - Eliminate investments in hardware