IBM Db2 – The AI database
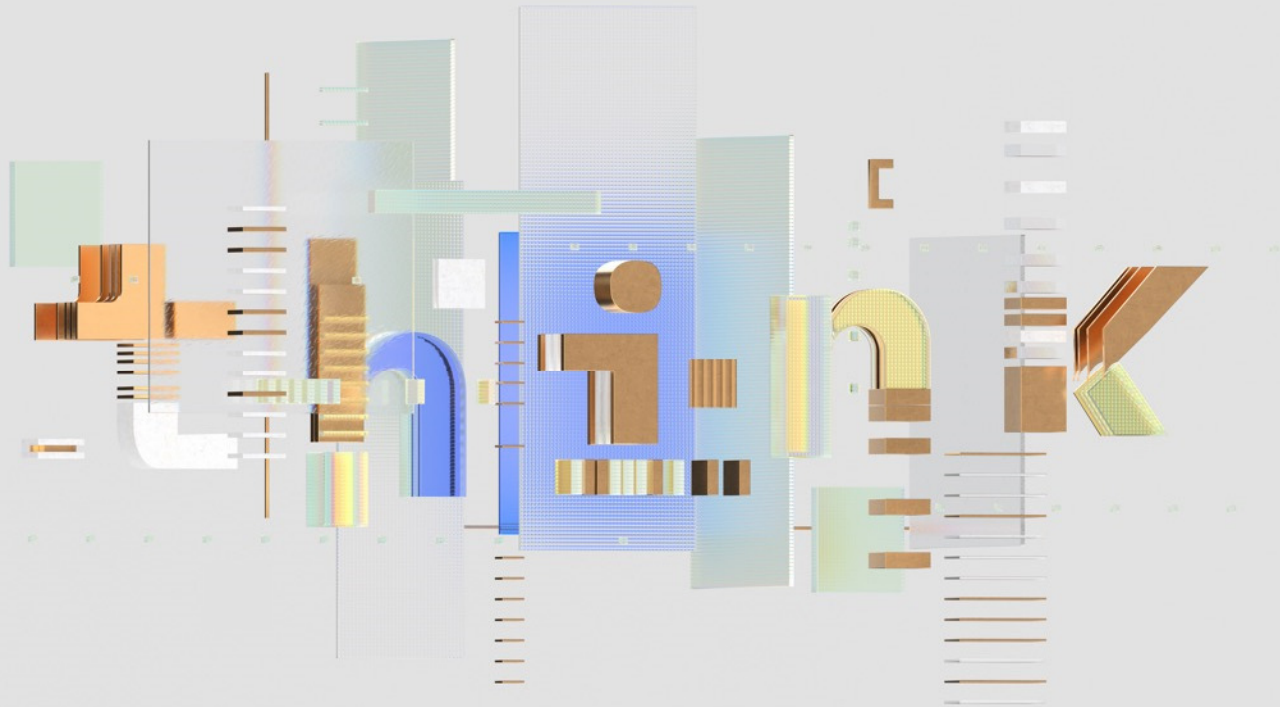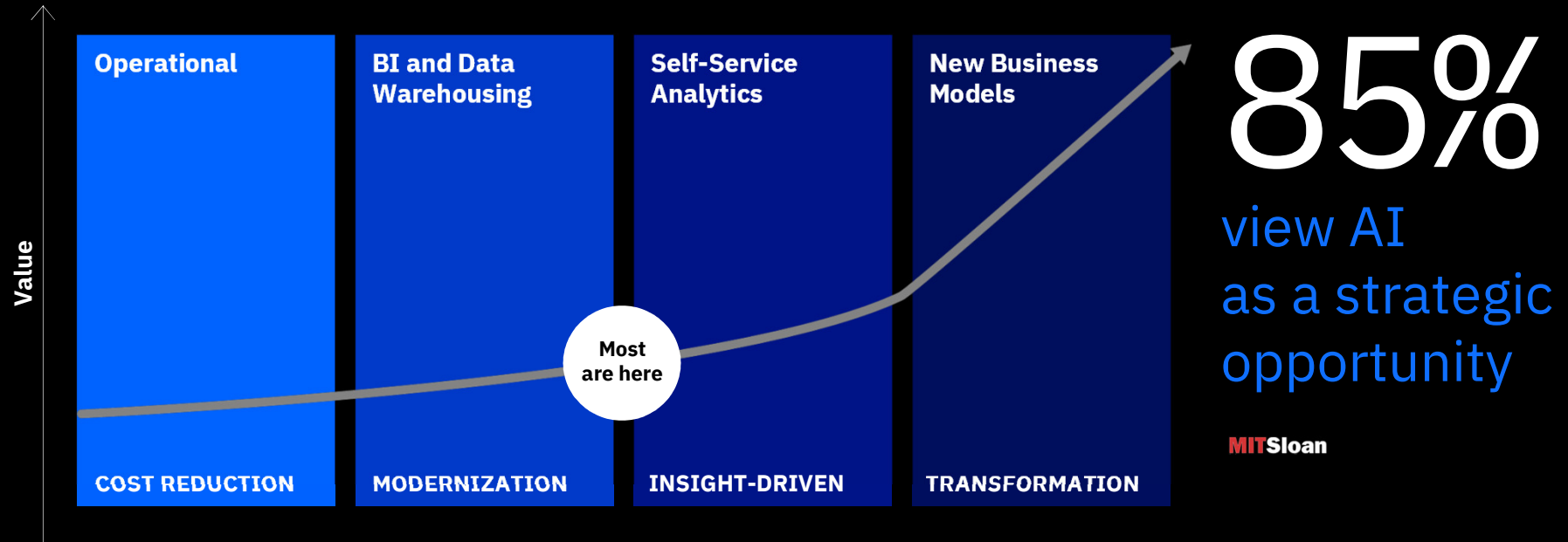
Pandit Prasad – Offering Manager, Db2 & Big Data

All business is becoming more data-driven

**"Data will be _your_
basis of competitive advantage"**
- Ginni Rometty

Our clients have declared the journey to AI a strategic priority

Value

Operational	BI and Data Warehousing	Self-Service Analytics	New Business Models
COST REDUCTION	MODERNIZATION	INSIGHT-DRIVEN	TRANSFORMATION

Most are here

85%

view AI as a strategic opportunity

MITSloan

33%

Accessing and preparing data is a barrier to using machine learning

88%

AI and ML important for their data platform and analytics initiatives

*451 Research Voice of the Enterprise: AI & Machine Learning, 2019

IBM delivers the capabilities you need to build a ladder to AI

Multi-Cloud

Infuse: Automate & scale across your processes

Trust: Achieve trust & transparency in outcomes

Analyze: Scale insights with ML everywhere

Organize: Create a trusted analytics foundation

Collect: Make data simple & accessible

IBM proudly presents:

IBM Db2

The AI database

– **Transform** your business with AI infused data management

– **Deliver** data to the business more quickly and with higher confidence

– **Create** AI apps faster with tools that make data and analytics accessible for AI development

The foundation of AI is data architecture. Be ready for everything that comes next, with

the new generation of IBM Db2

IBM Db2 - the AI database

Powered By AI

Confidence-based query results
leveraging ML-SQL

Up to 10x better query performance
powered by an ML-Optimizer

single view on all data
delivered by Data Virtualization

Auto resource optimization
delivered by Adaptive Workload
Management

Built for AI

Faster data exploration
by using NLQ in Augmented Data Explorer

Build AI based applications
with Python, GO , JSON and Jupyter notebooks
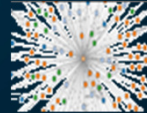
Model Complex Relationships
by using Db2 Graph and SQL

Blockchain Ready
using Db2 Blockchain Connector

Db2 – The database powered by AI
Enables a spectrum of automated tasks for the administrator

Figure 5: The spectrum of AI use cases as they relate data management
Source: 451 Research

Automating repetitive tasks:

Automated data identification and tagging.

Augmenting humans to do everything in between:

Automation to make databases easier and more intuitive to work with.
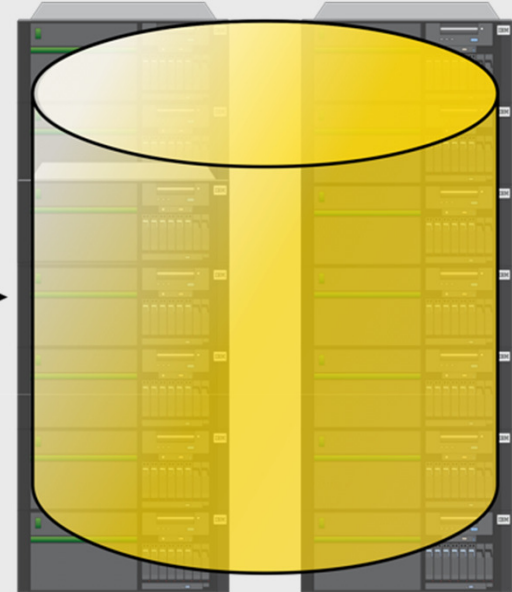
Automating complex tasks:

Integrated support for the development of data science and AI applications

Db2 scales … easily and is optimized for maximum performance

Db2 is the only database that offers exclusive features for maximum performance

Go from gigabytes to petabytes … without compromising performance

- Industry leading cost optimizer
- ML Optimizer (New)
- Adaptive Workload Management (New)

Db2 - the only database that offers ML Optimizer in addition to Cost Optimizer

New

Machine Learning Optimizer that improves from experience to make query optimization simple, reliable and stable

- Number of rows flowing through the various operations impacts performance for most common issues and is calculated via cardinality estimates

- Basic tuning to improve cardinality estimation is high impact

- **Phase 1 – Cardinality Estimation**

    - Initial phase – Support equality & range local operators with no expressions

    - Future phases – Cardinality support for expressions, predicates & BLU, support for join enumerations,

Db2 delivers automatic workload management

Ensures system stability and responsiveness with zero tuning

- **Adaptive Workload Management** is the "next generation" of Db2 workload management (WLM)

- Delivers automatic workload management within Db2 that ensures system stability and responsiveness with zero tuning
  - Make sure that the system is well-utilized but don't let it become overcommitted
  - Schedule jobs appropriately to ensure fairness and appropriate responsiveness
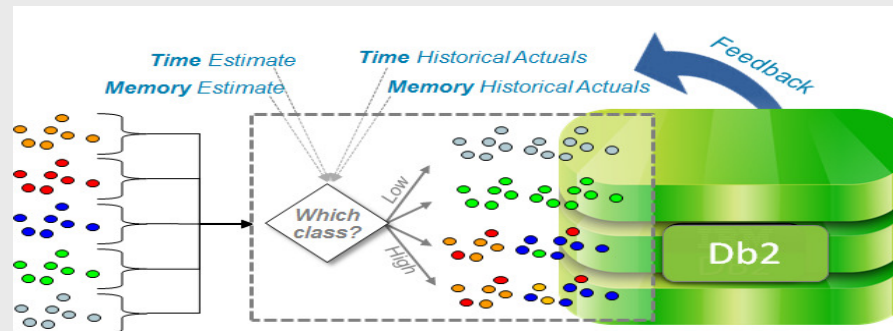
Some performance numbers for Adaptive WLM

Analytical workloads performed at par or better with Adaptive WLM compared to current WLM using default concurrency thresholds across a set of internal workloads

- Since work is admitted into the system based on matching of CPU and memory needs versus current usage, processing of work is generally more efficient due to smoother, steadier consumption patterns

- System stability is maintained throughout

Performance Improvement with Adaptive WLM

TPC-H 1 TB	TPC-H 1 TB (memory constrained)	BD Insights - 1 TB	BD Insights - 10 TB	BD Insights - 10 TB (Deep Analytics)
20%	30%	14%		
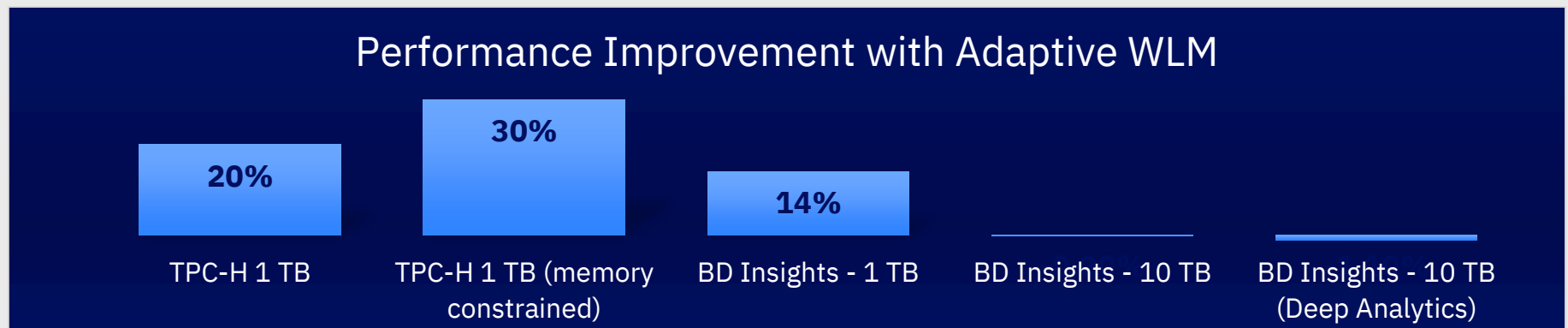
Db2 Data Virtualization is the next generation of federation

Data Virtualization will offer sophisticated capabilities to work with data distributed across multiple data stores
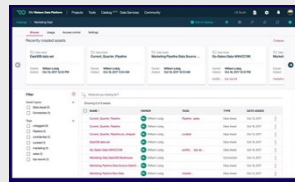
Unified data asset catalog, lineage and provenance
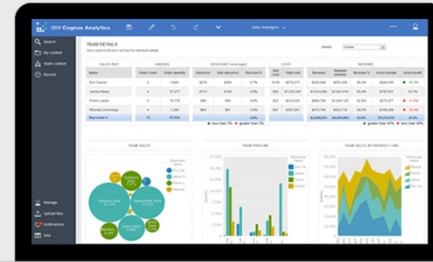
Unified access control and security policies

Data Virtualization
[+ caching layer]
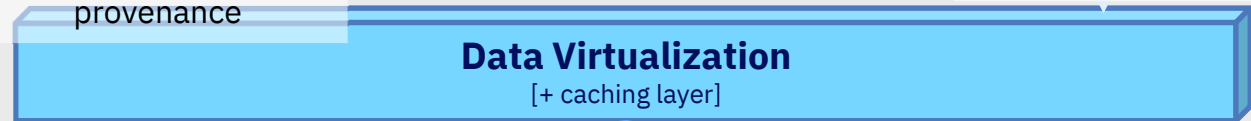
Data Warehouses & Marts

Relational Databases

No SQL

Spreadsheets & Text files

Big Data (Hadoop)

Locations: Private and public clouds, standalone systems, worldwide.

By the way we have auto-sharding too!

Sharding within Db2 (also known as DPF) lets you partition your database across multiple servers or within a large SMP server.

- *This allows for peta-byte scalability, since you can add new machines and spread your database across them.*

- *That means more CPUs, more memory, and more disks from each of the additional machines for your database!*

- Ideal to manage large databases, whether you are doing data warehousing, data mining, or online analytical processing (*OLAP*)
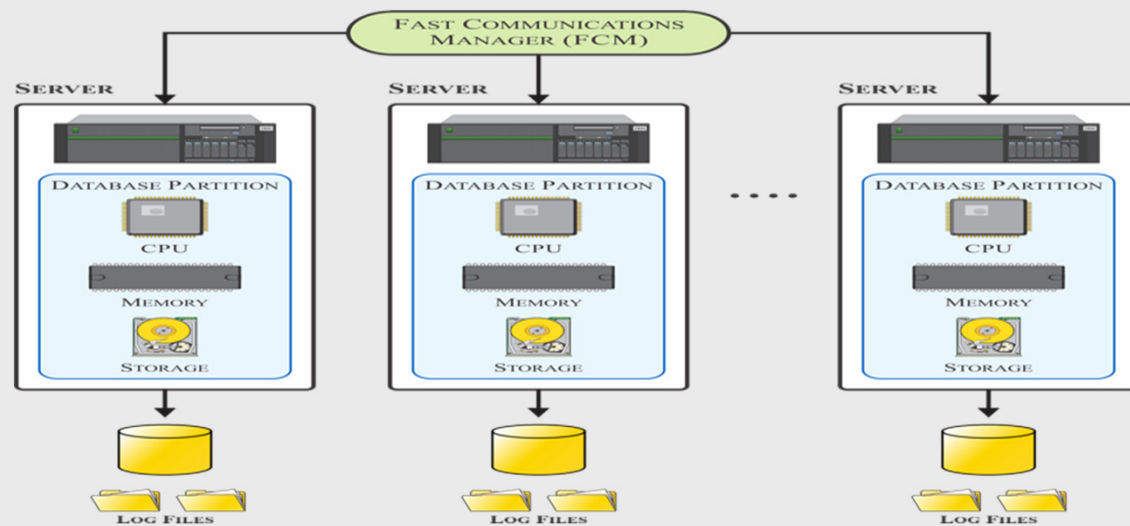
An AI Database automation unleashed

Machine learning has a key role to play in improving data base efficiency, performance and accuracy

Important but highly repetitive tasks can be accelerated through automation.

Improve query performance and accuracy

Increase system stability and reliability reducing manual tuning and monitoring of databases.

Decrease the time it takes to generate insights as well as make higher confidence-based business decisions.

Db2 – The database optimized for AI

Support modern AI applications & workloads ...
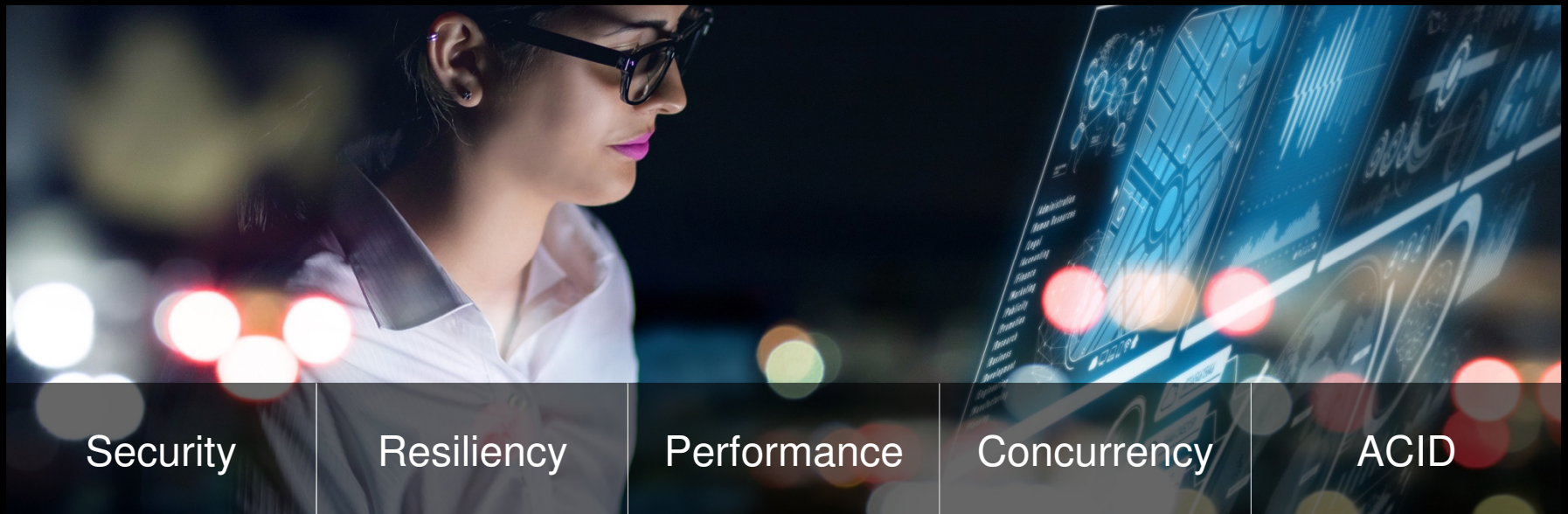
| Security | Resiliency | Performance | Concurrency | ACID |

... while offering a single comprehensive view on all data

Developers and data scientists can now use Db2 for their AI applications

Eight open source drivers

available from github -

https://github.com/ibmdb

- Go
- Python (Django, SQLAlchemy & Alembic)
- Jupyter Notebook
- Sequelize
- PHP
- Java (Spring)
- Node.js
- Ruby
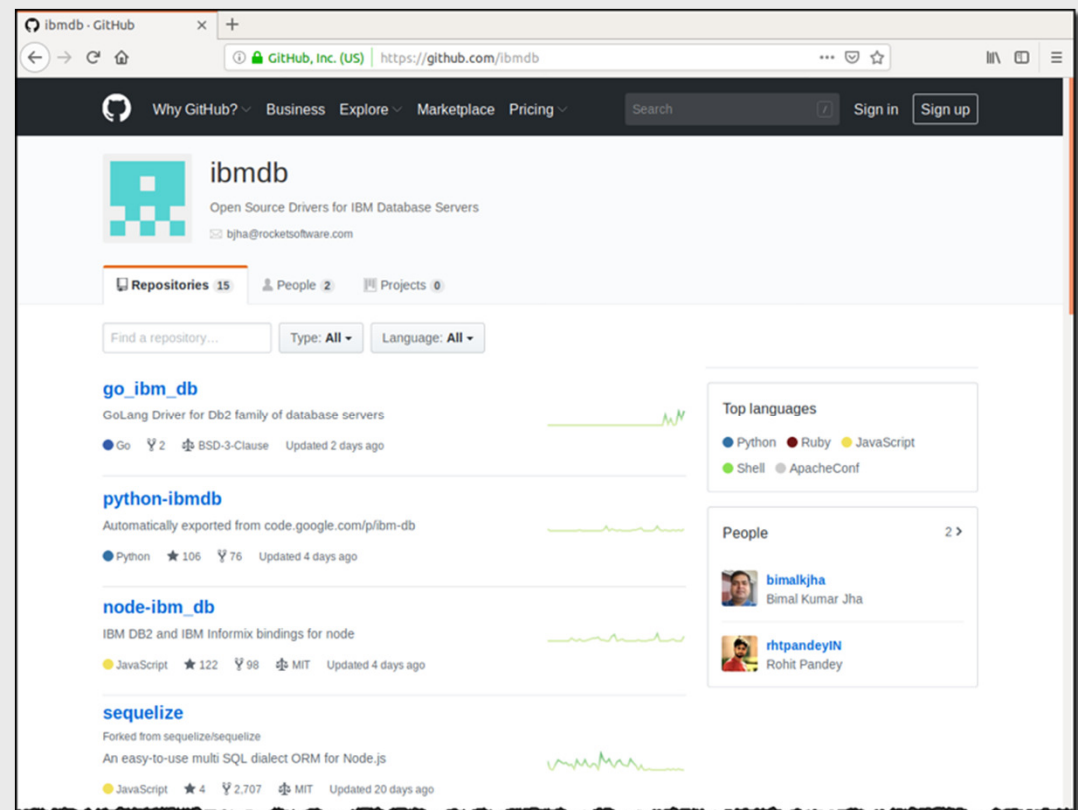
Follow "Code Patterns" to build cognitive applications today

Db2 Code Patterns – Learning by Doing
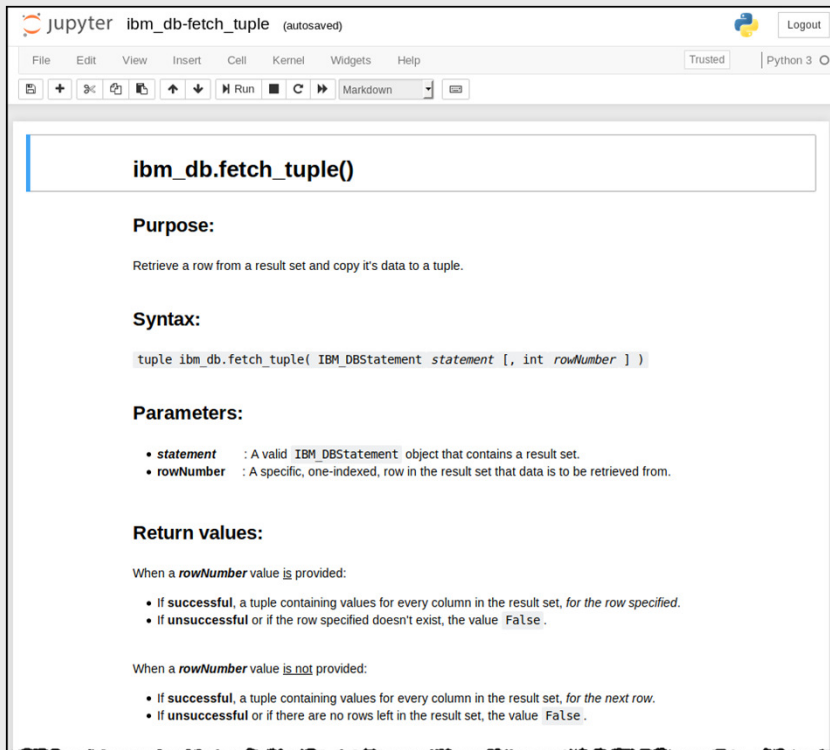
(https://developer.ibm.com/components/ibm-db2-database/)

NEW!

... and an extensive set of Jupyter Notebooks to kick start your Db2 development

https://github.com/IBM/db2-python

We just announced VSCode Extension for Db2!

NEW!

- Based on Stack Overflow Survey 2019, Visual Studio code was voted as the favorite IDE amongst the developer community

- This extension provides a graphical user interface designed to enable development with Db2 from within Visual Studio so that developers can accelerate and simplify application development.

- The extension is named "db2connect"

LedgerSQL – Query your Blockchain data hassle-free

What

Empowering end-users to query Blockchain data leverage their existing skills (SQL) seamlessly just like any other traditional database.

Why

Make it easy for adopting blockchain, perform analytics/reporting on it and integrate with other data stores to give 360 view of information.

How

Build a Db2 Federation Wrapper to offer SQL querying capabilities on Hyperledger Fabric shared ledgers using the Fabric SDK & security model

Use Db2 SQL to query Blockchain

1. A Db2 federation wrapper that provides:

   – Connectivity to Blockchain <u>Transactions</u> and state datbases via Hyperledger Rest API

   – Rewriting SQL queries into equivalent Hyperledger API calls to query data.

2. Using Zeppelin as end user tool to connect to Db2

3. Query Blockchain via Zeppelin using SQL.
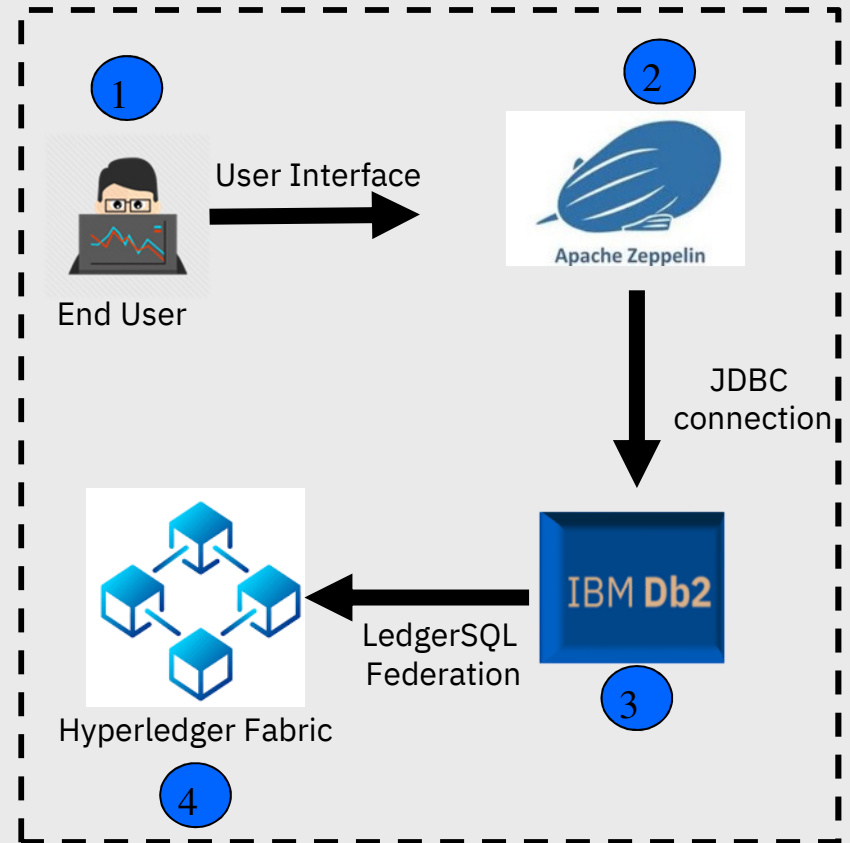
Combine Blockchain data with other data stores for Analytics

Easy and Secure: (Addressing the Skill Gap)

-In place querying, reporting & analytics of Blockchain data using the rich SQL language support provided without compromising the Blockchain security model.

Data Virtualization: (Fits into Analytics Infrastructure)

-Joining Blockchain data with other enterprise data set (Hadoop, Hive, Spark, external RDBMs) for complex use cases without additional work

-Immediate support for all DB2 based products through Federation

Connect to Blockchain just like any other database

Reporting Tools

Data Access Layer

Traditional Database

Blockchain

Data Repositories
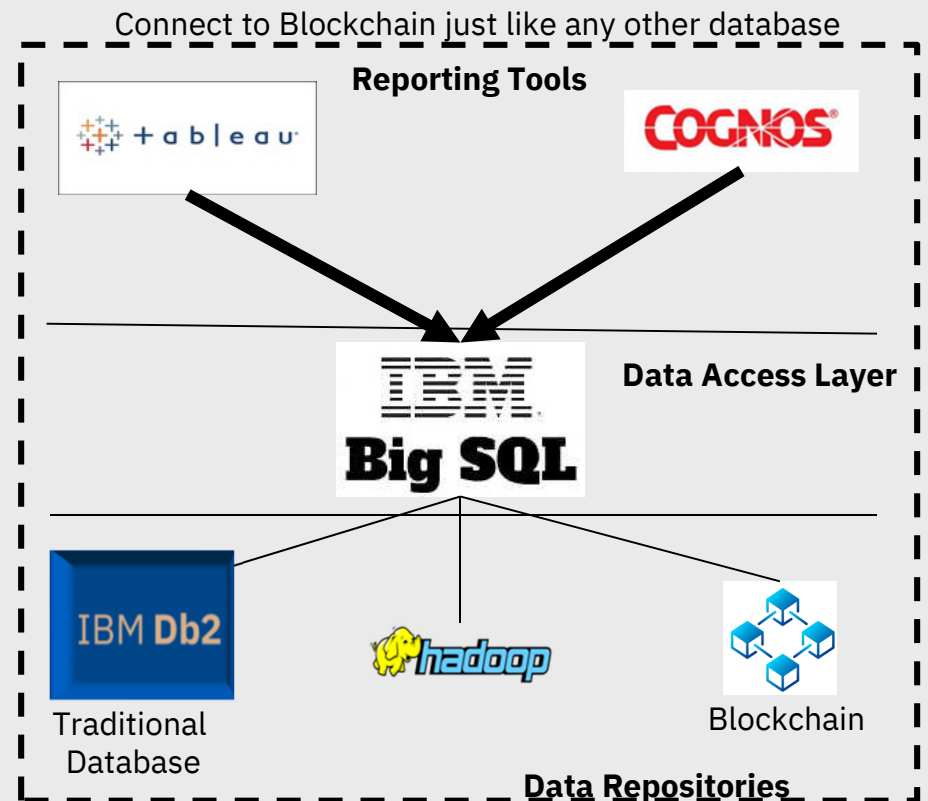
Data Exploration

"*accessing and preparing data is one of the most significant barriers to ML adoption...database and data management products that accelerate data exploration can lay the foundation for more efficient development of AI applications*" - 451 Research, Accelerating AI with Data Management, Accelerating Data Management with AI, 2019

AI DATA PIPELINE

Ingest	Data Prep.	Training	Testing	Deployment	Inference
Data collection	Data cleansing and transforming	Parallel model training	Validation on test dataset	Integration into applications	New, real-world inputs leading to actionable outputs
Pre-processing	Exploration and framework selection		Model fine-tuning	Scaling	
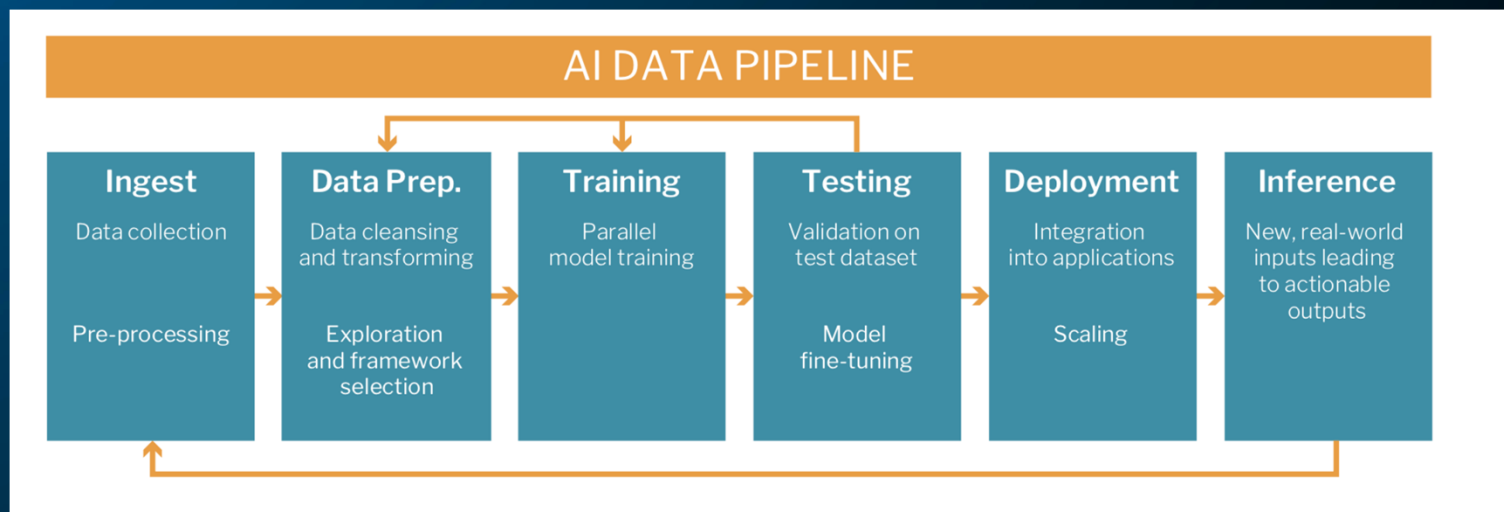
Use Augmented Data Explorer tool for rich and deeper insights

NEW!

Get to the insights faster – Use Natural Language Query (NLQ) for data analysis

Conversational analytics is the future...

- Data scientists often face the challenge of exploring data sets that are not familiar to them
  - Time consuming, Not sure what to look for
  - Need understanding of what is in the data set

- Need an intuitive data exploration tool
  - Alexa, Google Home, etc. prove trend
  - Natural extension of conversation to data
  - Provide multiple views of the data with minimal touch

- Free tool available to download - https://www.ibm.com/us-en/marketplace/db2-augmented-data-explorer
  - Currently in Beta
  - Seamlessly plugs and plays with Db2

Capabilities: Understanding NLQ to get accurate results
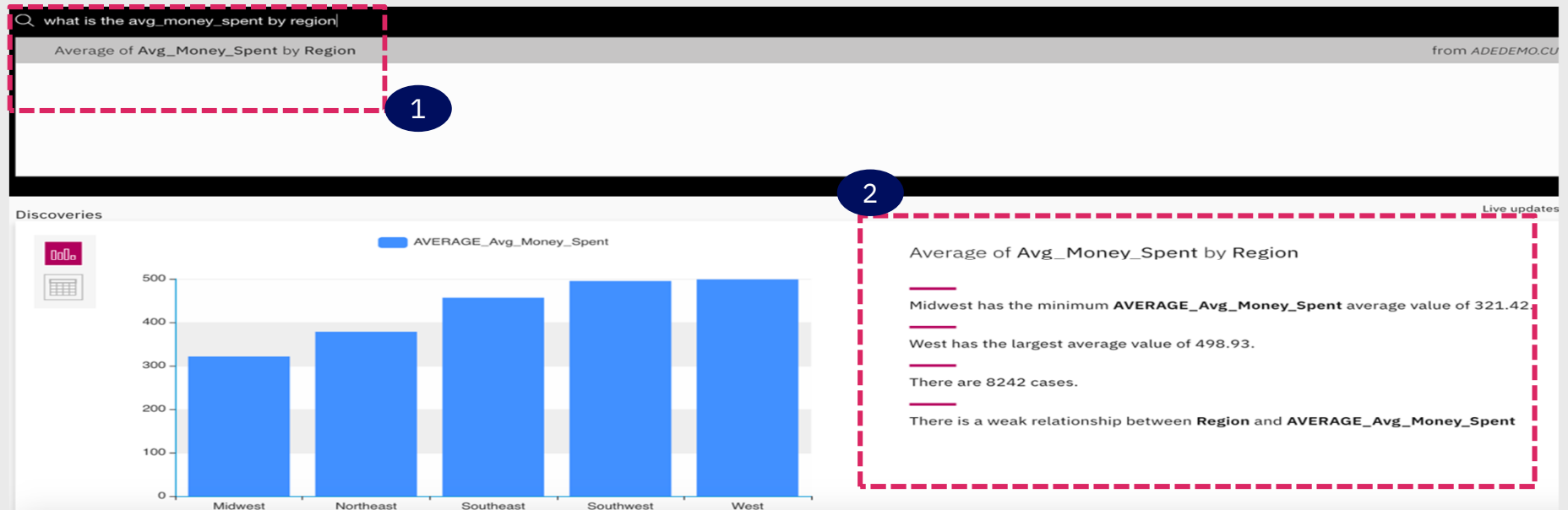
1. "avg sales by region" translates to: SELECT AVG(SALES), REGION FROM xyz GROUP BY REGION
2. Ability to automatically interpret the search result and create natural language based analysis that can be understood by user

Capabilities: Relevance

1. Understanding the user intent and proving type ahead queries

2. Providing suggestions for similar queries to the user

3. Providing source table of the data that the query will run against

4. Ability to use ML (Available) Model to augment insights

Try out the new Db2 Augmented Data Explorer

https://www.ibm.com/us-en/marketplace/db2-augmented-data-explorer

How to be the best database for ML applications?

Access to more Data
=
*Better Machine Learning Application

Let's try to bring Machine Learning closer to the data!

Testing and Deployment
Model Development
Data Quality and Transformation
Data Exploration
Data Access
Data Storage
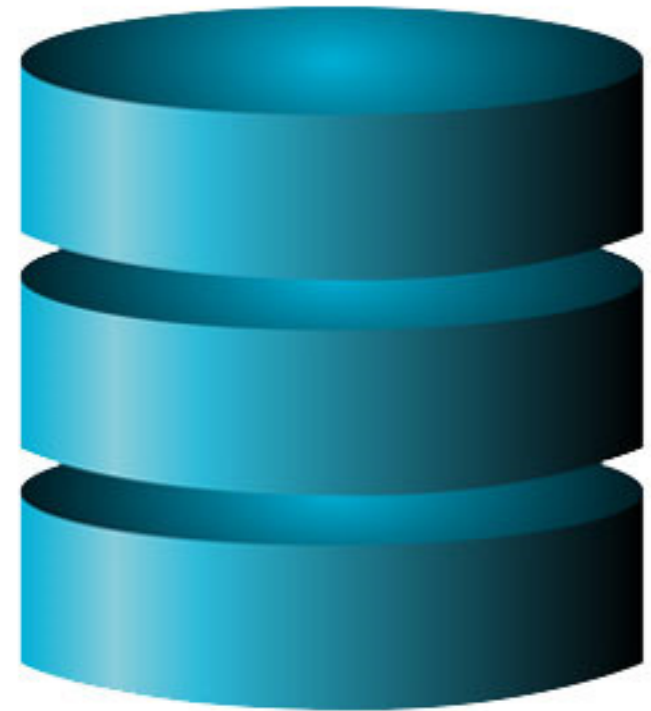
IBM Db2 - the AI database

"IBM is adding AI into Db2 through data science methods that will inject Machine Learning into it."
-Forbes

"Db2 11.5 will feature better hooks for the languages and development tools that data scientists work with"
-Datanami

"Db2 version 11.5 has added drivers for leading AI languages to help ease AI app dev with Db2."
-DBTA

One Db2 download, three Db2 licenses

	Community License	**Standard License**	**Advanced License**
Price	Free, unsupported	License + S&S (perpetual)	License + S&S (perpetual) subscription
License	Development / Production	Development / Production	Development / Production
Metric	Not applicable	Virtual Processor Core (VPC)	Virtual Processor Core (VPC)
Target	Developer community	Departmental	Enterprise
Caps	4 cores, 16 GB RAM	16 cores, 128 GB RAM	Unrestricted
Features	All features	All features	All features

Write once, run anywhere, with a Common SQL Engine

Anchored by a **Common SQL Engine** enabling true, highly scalable hybrid data warehousing solutions with portable analytics

Unified application and user experience across the Db2 family

Managed public Cloud DBaaS	Software defined warehouse On-Premises or in Cloud	Dedicated analytics appliance	Custom deployable database	Open source Hadoop with Hortonworks
Db2 on Cloud Db2 Warehouse on Cloud	Db2 Warehouse	Integrated Analytics System	Db2 Database	Db2 Big SQL

- **Application compatibility**
  Write once, run anywhere
- **Operational compatibility**
  Reuse operational and housekeeping procedures

- **Licensing**
  Single entitlement for flexible consumption enabling business agility and cost-optimization
- **Integration**
  Data virtualization capabilities for query federation and data movement

- **Standardized analytics**
  Common programming model for in-DB analytics
- **Ecosystem**
  One ISV product certification for all platforms

Deploy and scale data, when and where you want

IBM Hybrid Data Management Platform

One Db2 family

One HYBRID entitlement

| Consume anywhere | Transition as needed | One bill | Attractive price |
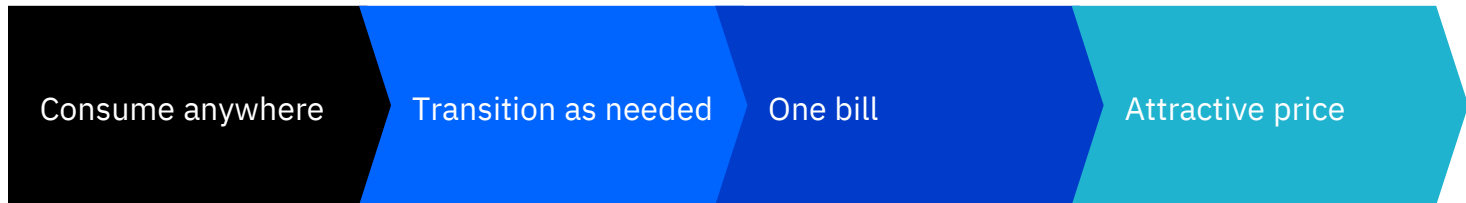
Available for purchase via two licensing models:

HDMP On-Premises	**HDMP Subscription**
(licensed *perpetually*)	(licensed *monthly*)

- Supports OLTP, warehousing, SQL on Hadoop, and Fast Data on private cloud and on-premises infrastructure
- Intended for customers who need to maintain ownership of their software licenses in perpetuity

- Supports OLTP, warehousing, SQL on Hadoop, and Fast Data on public cloud, private cloud, and on-premises infrastructure
- Intended for customers who want to procure software and cloud services as an operational expense

Hybrid Data Management Platform

Now supporting flexible, consumption-based licensing models

For both HDMP On-Premises and HDMP Subscription:

- **Step 1:** Customer purchases FlexPoints towards a license

- **Step 2:** Customer deploys those FlexPoints towards any of the bundled products and services included in the respective bundle

- Customer can mix/match their deployments and change them at any time

- No intervention is needed from IBM in order for customers to move between the bundled products and services!

FlexPoints

A   B   C

Estimate the costs - https://hdmp-calculator.mybluemix.net/
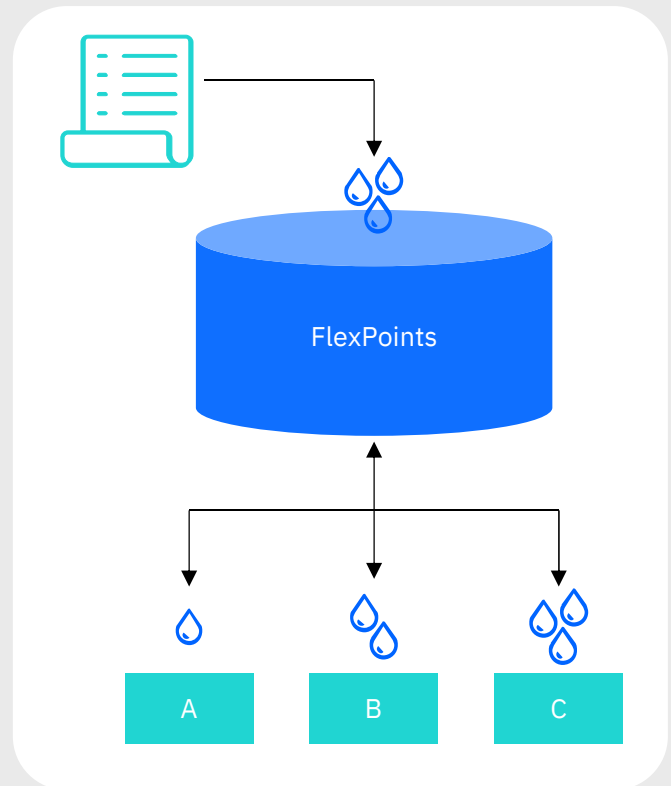
What's included in the platform?

Customers purchase a monthly subscription and gain entitlement to seven of IBM's best-in-class data management products, spanning across on-premises and the cloud.

Db2 Hosted

IBM's client-managed Db2 cloud service with fully scalable infrastructure

public cloud / DBaaS

Db2 on Cloud

IBM's fully-managed Db2 cloud database with fully scalable infrastructure

public cloud / DBaaS

Db2 Warehouse on Cloud

IBM's fully-managed cloud data warehouse for in-database analytics & MPP

public cloud / DBaaS

Db2 Advanced Enterprise

the most performant edition of IBM's flagship relational database management system

on-premises / private cloud

Db2 Warehouse

IBM's private cloud data warehouse for Docker container supported infrastructures

on-premises / private cloud

Db2 Event Store

IBM's premier in-memory database designed for massive structured data volumes

on-premises / private cloud

Db2 Big SQL

IBM's SQL on Hadoop solution product low latency support for ad-hoc or complex queries

on-premises / private cloud

The Hybrid Data Management Platform provides your business with **peace of mind**. No matter how your business needs evolve in the future, deployments can be changed at any time by simply reallocating your **FlexPoints**.

Hybrid Cloud Usage Hub

HDMP Subscription customers gain access to the **Hybrid Cloud Usage Hub**, a web-application used to manage and deploy your FlexPoints across all of the bundled products.

- Seamlessly integrate across on-premises and cloud using one dashboard

- Gain a 360 degree view of your resources and how they're being consumed, so that you can optimize your infrastructure costs

- Easily and quickly expand your hybrid footprint, thereby reducing cost of development to production and saving valuable time
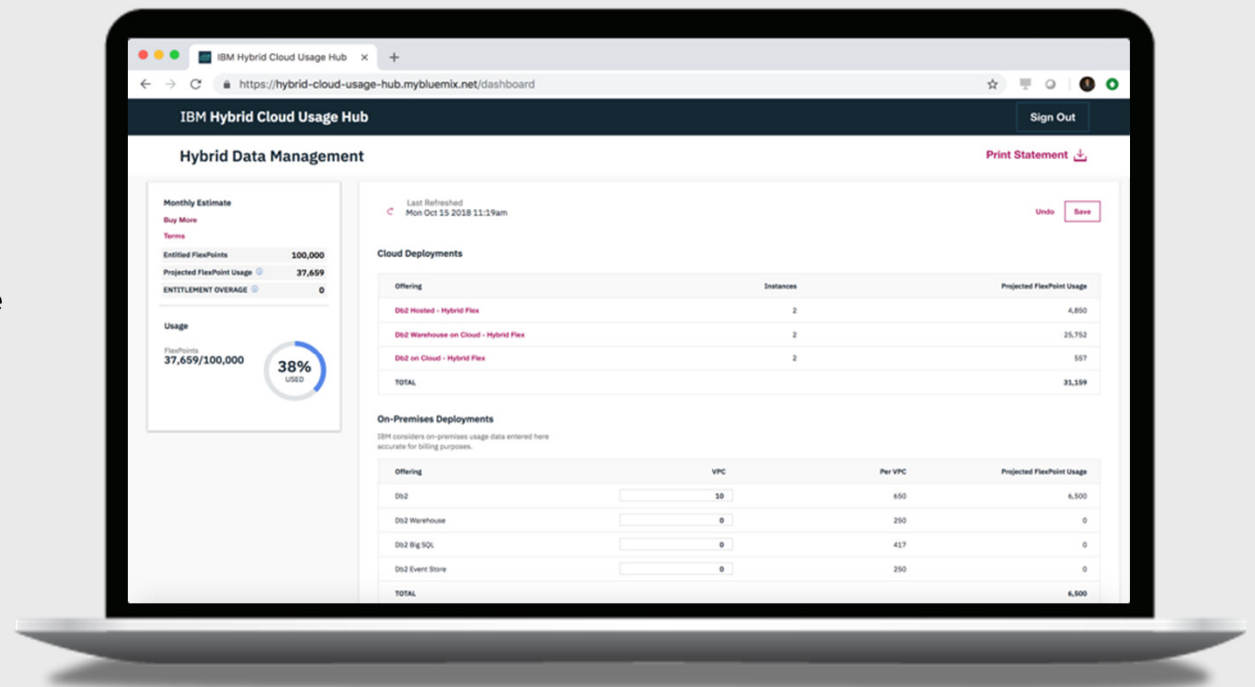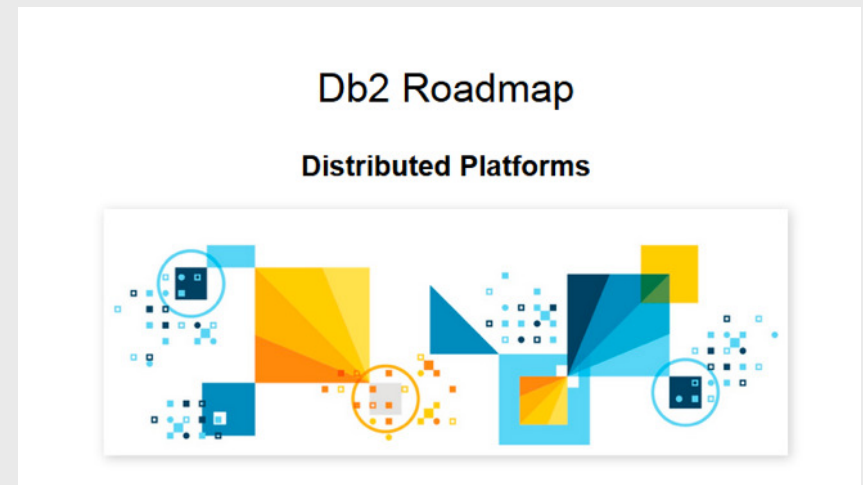
Follow our plans using Aha

We revisit development priorities frequently (e.g. every quarter) in response to customer and market demand/feedback

- As a result: some items move up, some down, some in, and some out.

We have committed to keeping our core roadmaps visible to the public eye using Aha

- http://ibm.biz/AnalyticsRoadmaps

Db2 Roadmap

Distributed Platforms

Pandit Prasad

Offering Manager – Db2, Big Data

pprasad@us.ibm.com

310-648-1934

Thank you!

IBM Developer
Code patterns,
Tutorials, Videos &
Articles
https://developer.ibm.com

Digital Technical Engagement for Db2
https://ibm-dte.mybluemix.net/

Read more about Db2 and application development: