Establish an Analytics Hub on Linux on IBM z Systems & LinuxONE

Wilhelm Mild

Executive IT Architect

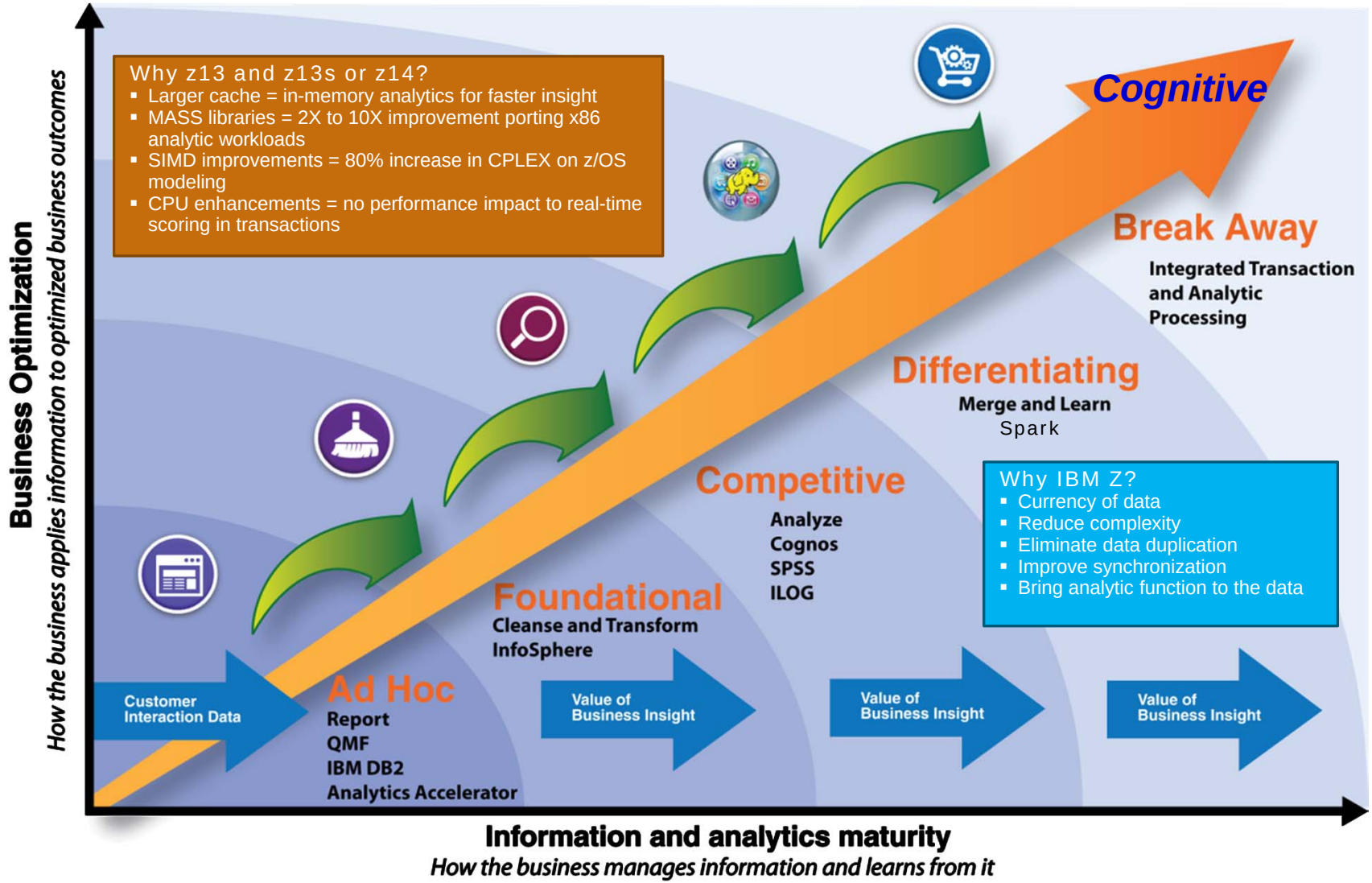
Integration Architectures

for Mobile, Linux & IBM Z

IBM Boeblingen Laboratory

IBM Z point-of-view

Building a foundation to grow with business needs



Evolution of Analytics

data is the resource for competitive advantage

	Descriptive *Analytics by spreadsheet*	Predictive *Alerts & suggest next actions*	Cognitive *Identify & immediately act*
	A business management exercise	*Studying business data to understand trends and predict the future*	*Expanding beyond business data for business opportunities*
who	Analysis by small teams	Owned by IT organizations	Lines of business
What	On historical structured data	Copy data to build data warehouses & data marts	Combine structured & unstructured data
How	Using utilities or roll your own tools	Tools build models to learn from historical data	Discover & act using open source tools
why	Internal executive decision support	Increase business opportunities	Embed analytics into business processes

IT infrastructure matters not because of what you are doing now But because of what happens **next**

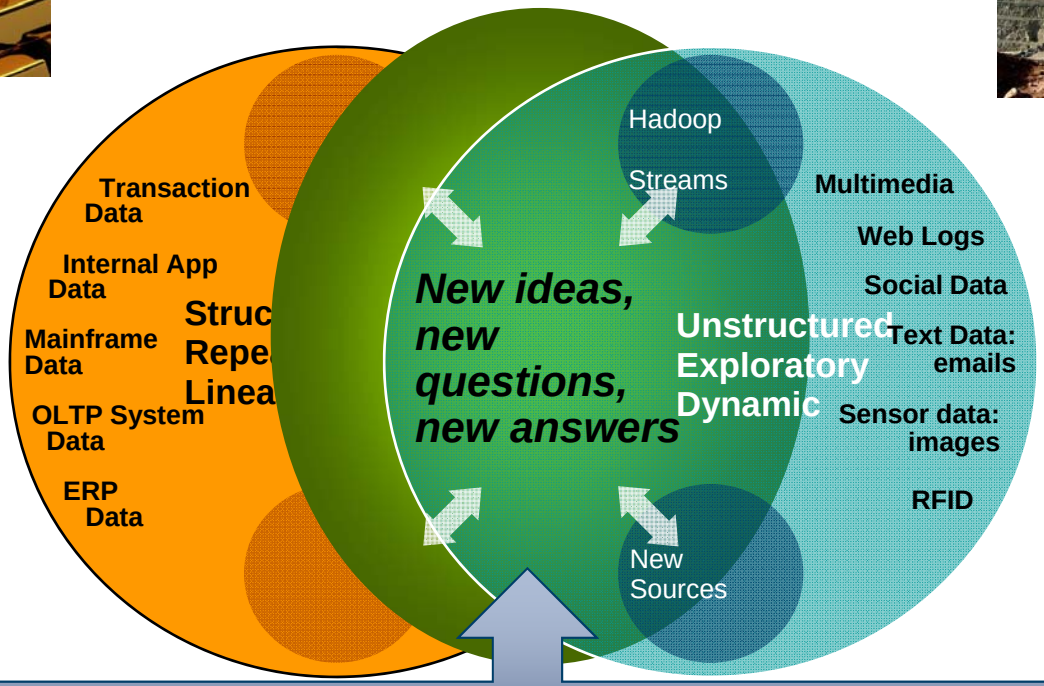
	Now	*What Happens Next*
	*Increasingly sophisticated analysis on static data*	• Currency of data becomes the driver of business value • Embedding real-time analytics into the business transaction • ***Privacy of client data and issues of ethical use*** • Integration of structured and unstructured data: Bringing business level QoS to a large scale with clustered and unstructured data • ***Analytics evolving into Cognitive: Power of Inference***
	*Virtualization*	• Orchestration, automation, and control of all resources: On-premise and off-premise • Splitting of application and data layers • ***Hybrid Cloud and Multi-architecture integration*** • ***Data privacy and security***
	*"Let's talk to the iPhone"*	• "Remember the lessons of the PC"- Today's toy can run tomorrow's world • ***Broader access → unprecedented and unpredictable transaction and data volumes → fluid scale*** • Unpredictable timing dictates 24x7 response and availability • ***Security: Data integrity and privacy with broader transaction access***

Imagine the possibility of leveraging all of your data assets

Traditional Technique
Structured
Analytical
Logical

Emerging Technique
Creative
Holistic thought
Intuition

"Here's a question, what's the answer?"

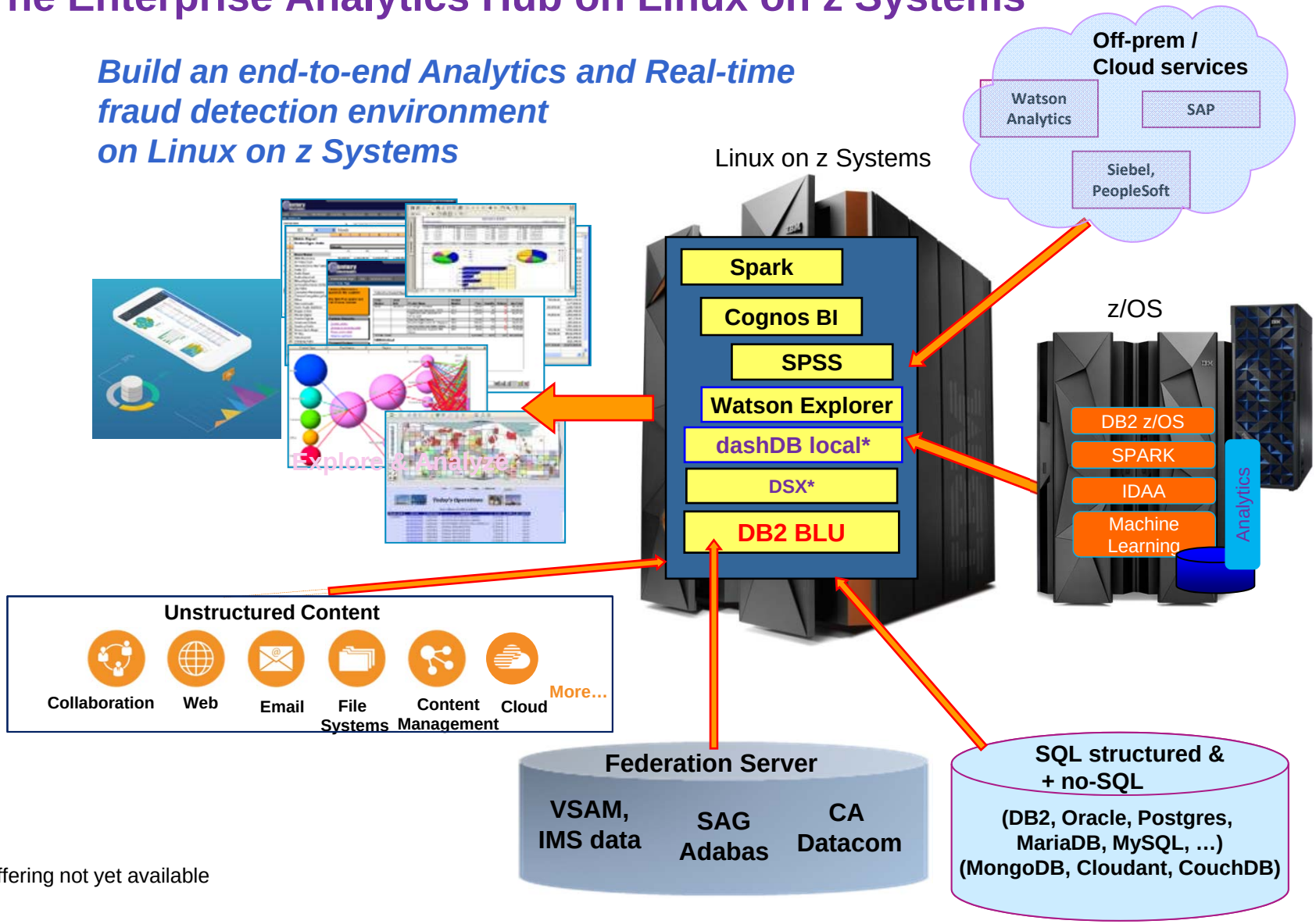


"Here's some data, are there correlations?"

Transformational **benefit comes from integration of new data sources with traditional corporate data**

(2) The Enterprise Analytics Hub on Linux on z Systems

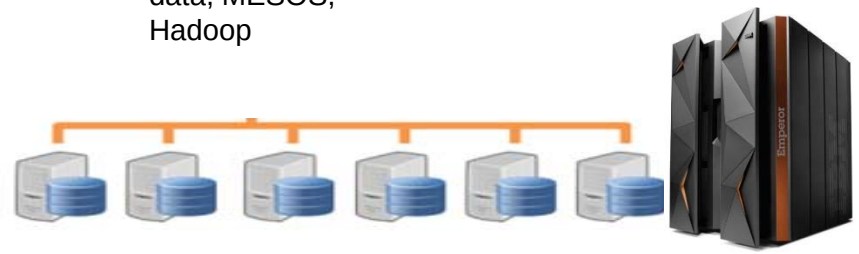
Build an end-to-end Analytics and Real-time fraud detection environment on Linux on z Systems



* - offering not yet available

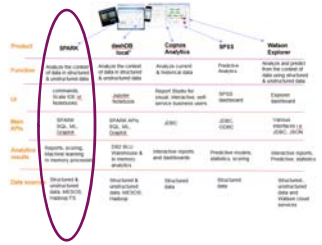
The Enterprise Analytics Hub on LinuxONE

Product	SPARK	dashDB local*	Cognos Analytics	SPSS	Watson Explorer
Function	Analyze the context of data in structured & unstructured data	Analyze the context of data in structured & unstructured data	Analyze current & historical data	Predictive Analytics	Analyze and predict from the context of data using structured & unstructured data
UI	commands, Scala IDE or Notebooks	Jupyter Notebook	Report Studio for visual, interactive, self-service business users	SPSS dashboard	Explorer dashboard
Main APIs	SPARK SQL, ML, GraphX	SPARK APIs SQL, ML, GraphX	JDBC	JDBC, ODBC	Various interfaces i.e JDBC, JSON
Analytics results	Reports, scoring, Machine learning In memory processing	DB2 BLU Warehouse & in memory analytics	Interactive reports and dashboards	Predictive models, statistics, scoring	Interactive reports, Predictive, statistics
Data sources	Structured & unstructured data, MESOS, Hadoop FS	Structured & unstructured data, MESOS, Hadoop	Structured data	Structured data	Structured-, unstructured data and Watson cloud services



* - offering not yet available

What is Apache Spark?



Java / Python / Scala / R — Languages

Spark SQL (Relational Operators) | Spark MLlib (Machine Learning) | Spark GraphX (Graph Processing) | Spark Streaming (Real-Time Streaming) — Spark Libraries

Spark Core (General Execution Engine) — Spark Core

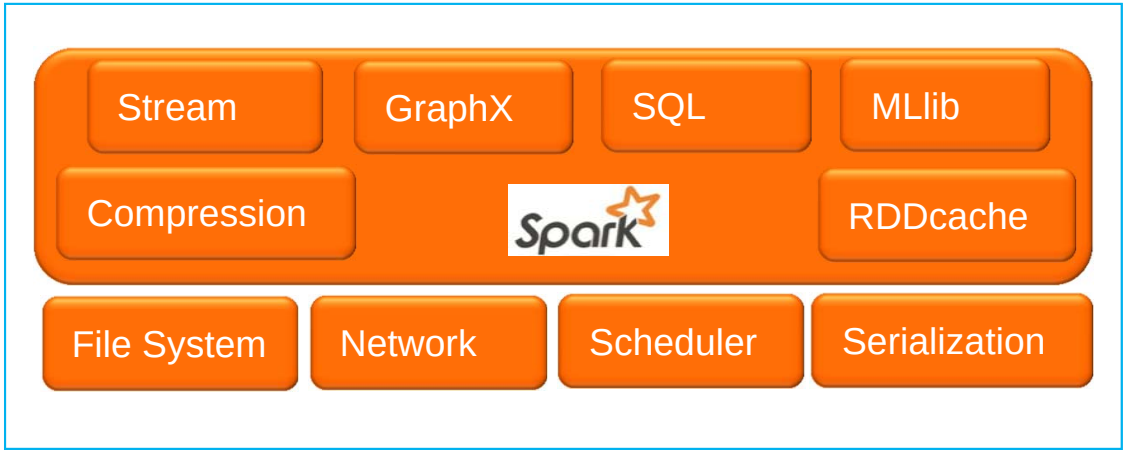
YARN | MESOS | Standalone — Cluster Manager

HDFS / Cassandra / HBase / Parquet / ... — Data Abstraction

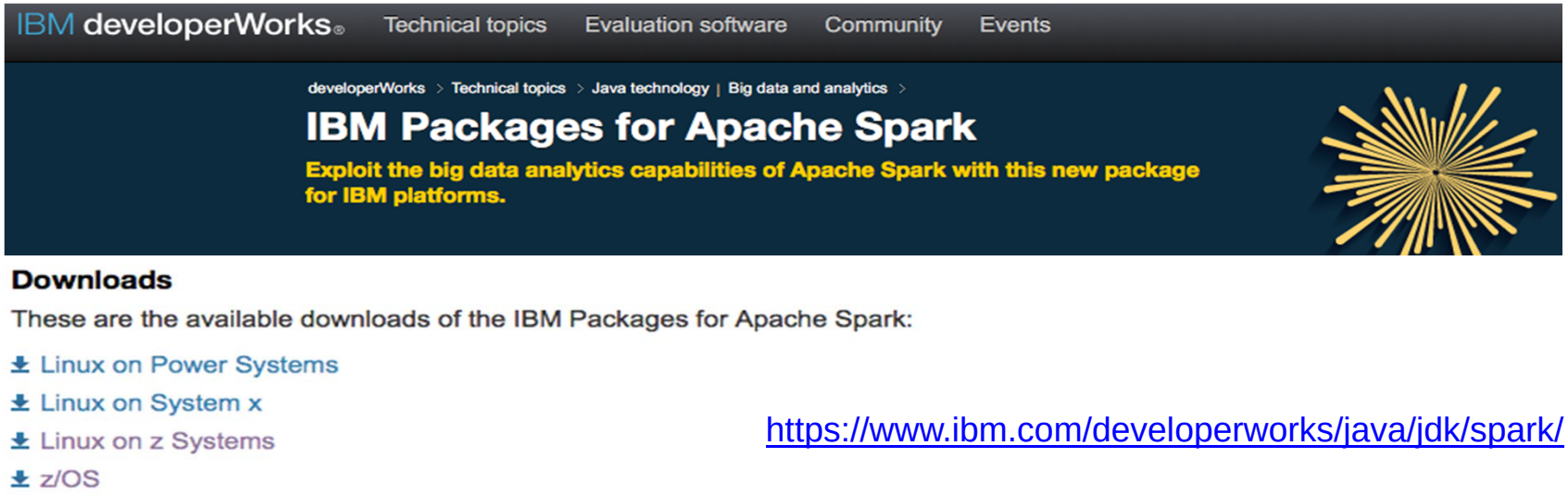
DB2 Connect Driver (JDBC)

Get Started – Exploit Your Data
download free software to start a project

- Federated, data-in-place analytics – reduce ETL
- Performance gains from co-location with data
- z Systems: SMT2, zEDC, SIMD, Large Pages, very high zIIP use



❑ Find insights from structured & unstructured data with Apache Spark

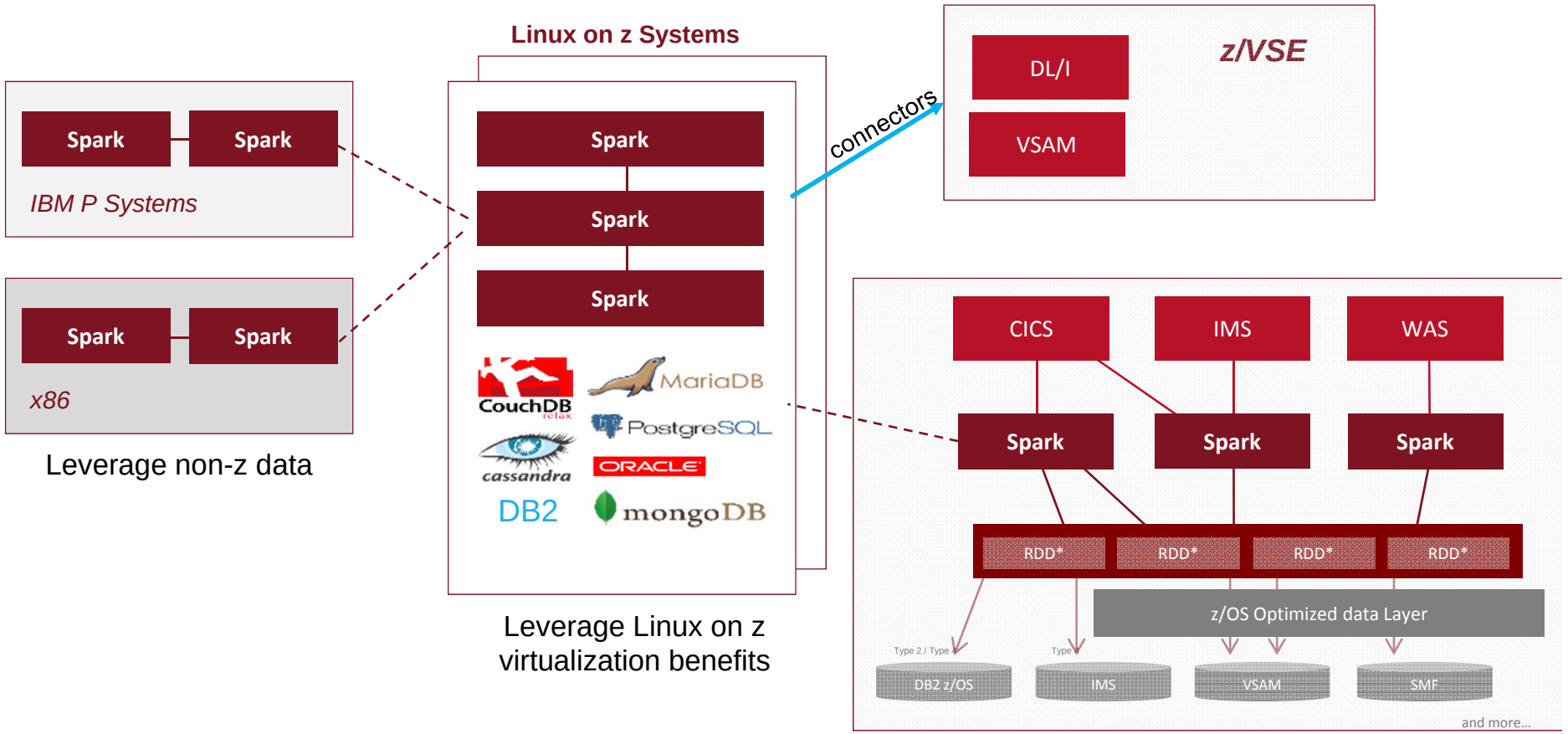


https://www.ibm.com/developerworks/java/jdk/spark/

Spark Analytics on Linux on z and z/OS

IBM z Systems provides an optimized platform to derive insights from all client data without moving it

Accurate – Secure – Federated analysis in a hybrid cloud model

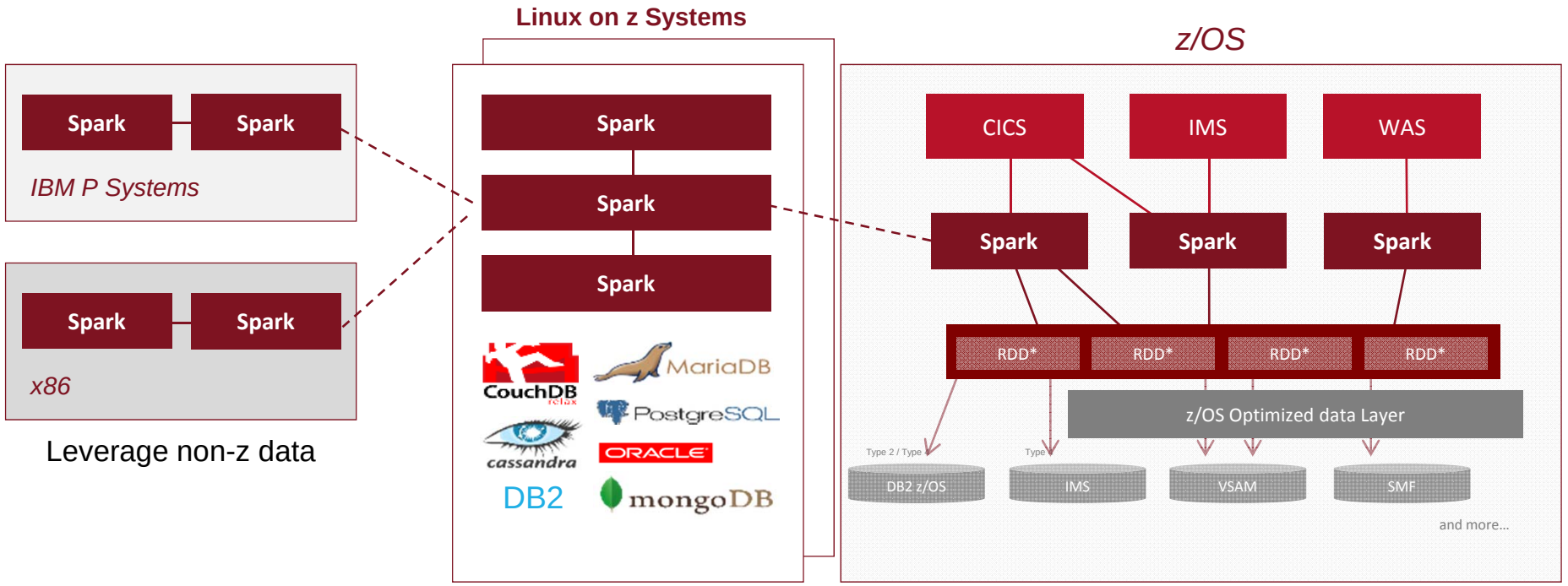


Spark is the Java of data analysis!

Spark Analytics on Linux on z and z/OS

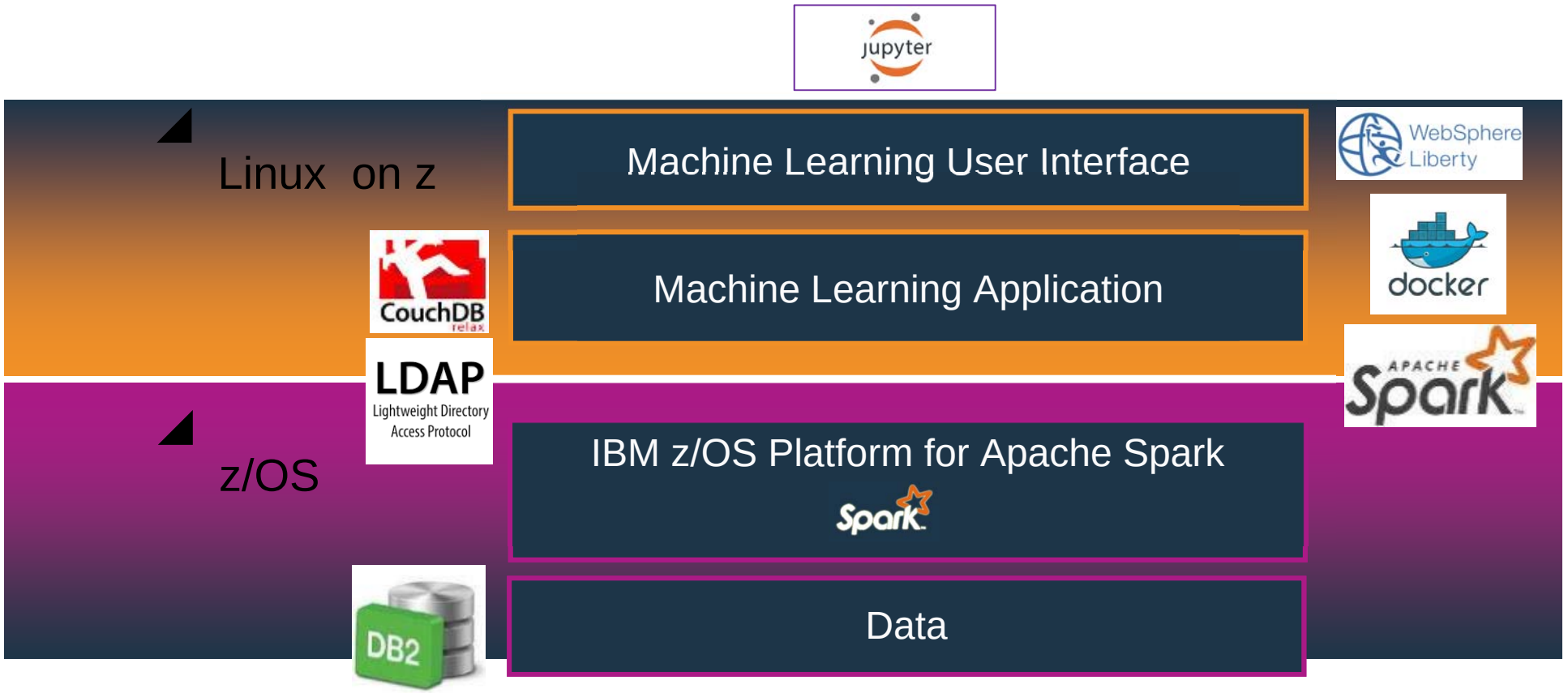
IBM z Systems provides an optimized platform to derive insights from all client data without moving it

Accurate – Secure – Federated analysis in a hybrid cloud model



Spark **is the Java of data analysis!**

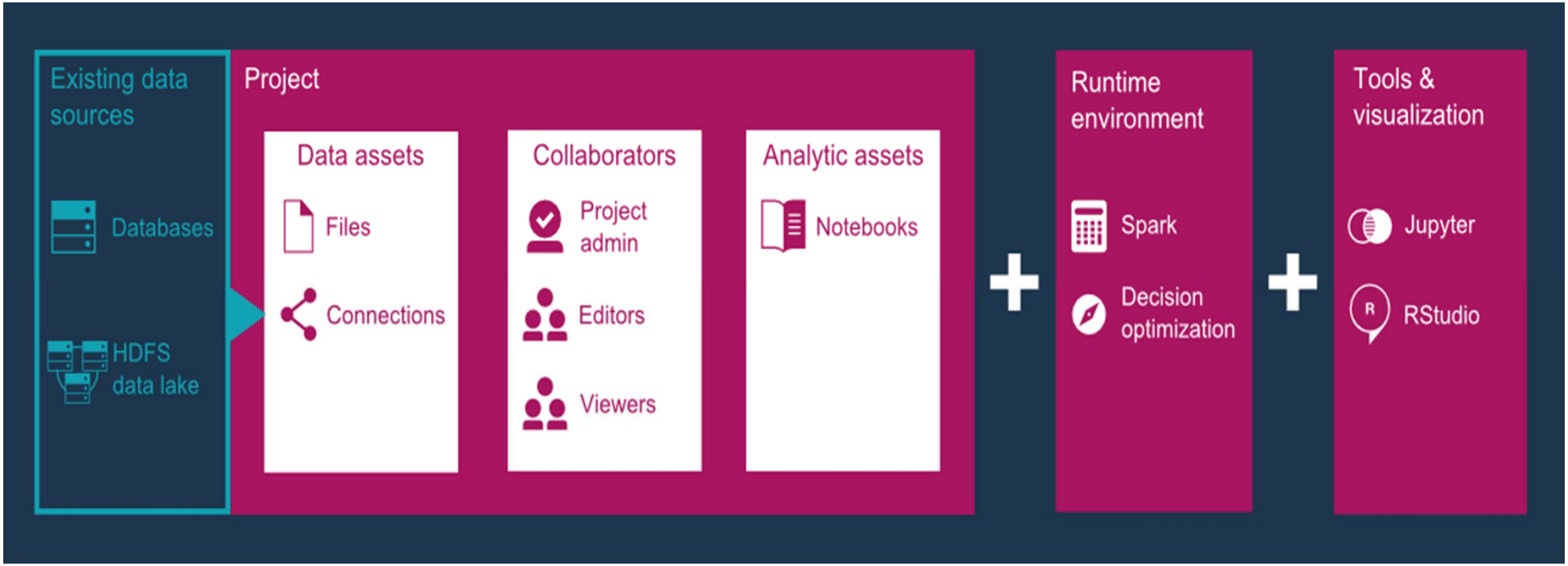
IBM Machine Learning for z/OS - Overview



Data Science Experience local*

Analytics and Machine Learning for Linux on IBM Z or LinuxONE

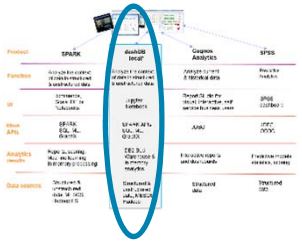
- out-of-the-box on premises enterprise solution for data scientists and data engineers.
- offers a suite of data science tools, such as Spark, Jupyter and Zeppelin notebooks, that are integrated with IBM technologies



*offering not yet available

https://datascience.ibm.com/docs/content/local/overview.html

IBM dashDB Local for Analytics*



dashDB Local is the premier private cloud data warehouse optimized for analytic workloads for Software Defined Environments (SDE) such as private clouds, virtual private clouds and other infrastructures that support Docker container.

Benefits of dashDB Technology with Fast Deployment into Private Cloud Environment

- Private or Virtualized Private Cloud
- Docker Container Technology
- dashDB Technology
- MPP with Automatic Scaling

- Highly flexible data warehouse
- Optimized for fast and flexible deployment into **private or virtual private clouds**
- Uses **Docker** container technology
- Built on top of **dashDB technology**, it shares the benefits of
  - BLU Acceleration in-memory columnar technology
- Massively Parallel Processing **(MPP)** with automated scaling capabilities to increase infrastructure efficiency

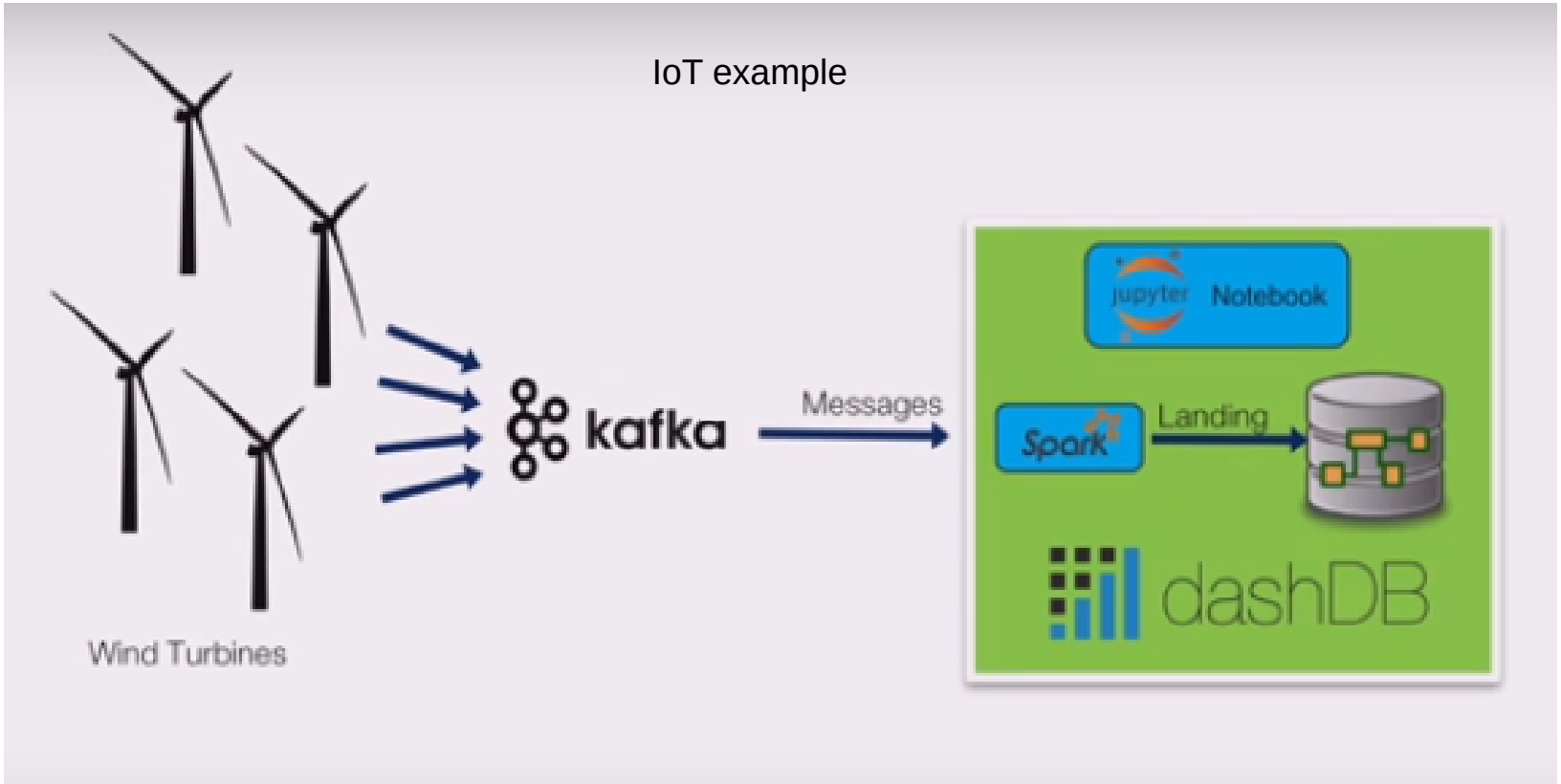
* - offering not yet available

dashDB local*

https://developer.ibm.com/clouddataservices/docs/dashdb/analyze/use-dashdb-local-spark-notebooks/
https://www.youtube.com/watch?v=mzOi45-KJN4

Streaming data using the built-in Apache Spark infrastructure in dashDB Local

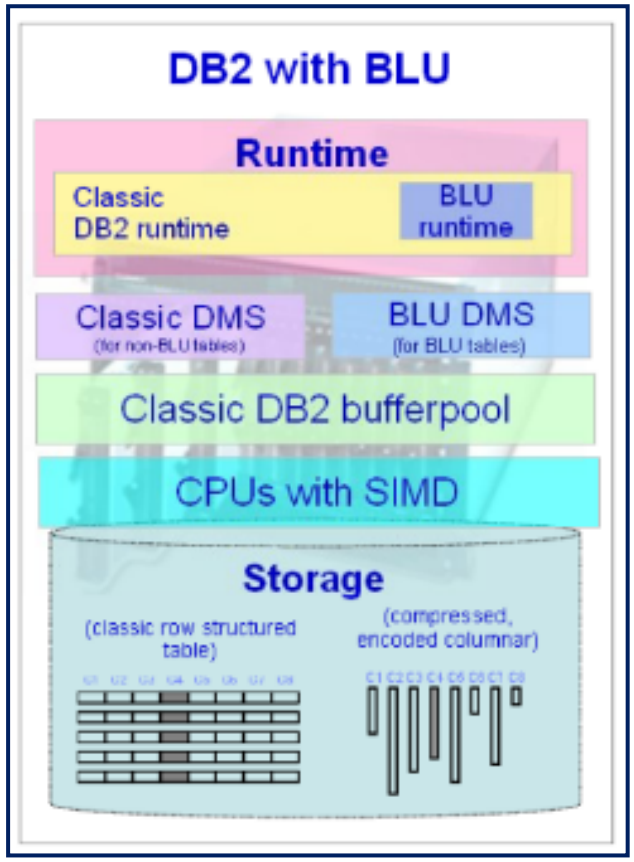
- **runs in Docker containers**
- **UI using Jupyter Notebook**



* - offering not yet available

DB2 w/ BLU Acceleration – inside dashDB

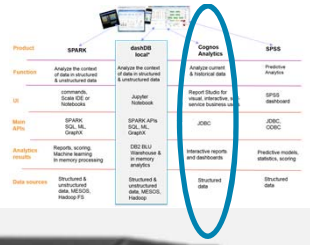
Super Simple. Super Fast.



Solution

- DB2 with BLU Acceleration is the preferred solution for customers who would like to run analytics with z Systems & Linux data
- Satisfy requirement for a columnar in-memory db
- Alternative of Linux on z Oracle installations
- Enhanced for distributed consolidations onto z Systems

Predictive Analytics: IBM SPSS on Linux on z
learn from historical data to make predictions



Techniques used to analyze data
- Data mining
- Statistics
- Modeling
- Machine learning
- Artificial intelligence

Example use cases
- Market Basket Analysis
- Fraud Detection
- Cross Sell – Up Sell Opportunities

Common Problems
- ***Copies of data created for specific needs***
- Created complexity in managing data
- Data Synchronization
- Excessive Costs

Predictive Analytics with Linux on z Systems technology
an open ecosystem for innovation

Co-locate for Right-Time insights

- Reduced latency allowing for better insights
- Minimize cost & complexity
- Improve data governance & security
- Open the aperture on innovation
- Cost efficiencies through consolidation
- Efficient HiperSocket LPAR connections

Freedom & Agility | **Standards Based** | **Developer Productivity**

Why IBM Z

SIMD delivers accelerated analytics processing for mathematical optimization

SMT delivers more throughput for Linux and zIIP-eligible workloads

zEDC reduces data transfer time and storage cost by up to **75%**

16 Gbps FICON links reduce latency for workloads such as Db2

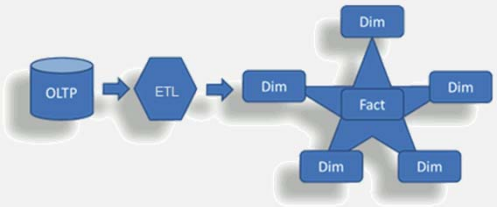
Cache increased to enable faster in-memory insights

10-32 TB of real memory ← even more critical with Linux

Descriptive Analytics : IBM Cognos Analytics
the analysis of historical data

1960s to early 1970s: Analytics expanded with the introduction of computers

Decision Support Systems provided business data for analysis.



Late 1970s: Relational databases created to *eliminate data redundancy / inconsistency* and improve structure of data
- data was organized around records
- relationships were enforced
- indexing for hi-speed access
- SQL was standardized

Reporting evolved from DB utilities to the creation of specific data marts for data manipulation
- creating cubes
- providing dashboards

Cognos Analytics

IBM Cognos Analytics is an enterprise BI platform for governed data discovery and managed reporting that automates the creation of reports and dashboards so users have the freedom to do it on their own. The user experience is designed for business professionals so they can easily prepare, create and visualize content using the built-in intelligence to guide them.

"We are excited about the new self service and visualisation capabilities of Cognos Analytics, it will enable our users to make more informed decisions."
Lizette Robles
BI Project Leader
Universidad de Guadalajara

Bring your data to life
Cognos Analytics is ready when you are with a unified experience that works the same on web or mobile devices, enabling you to quickly find, analyze, create and share insight.
•Intuitive interface lets all users quickly author content
•Dashboards created using drag and drop on mobile device or desktop
•Best visualizations automatically recommended
•Templates and styles to let you format reports instantly
•On demand menus for access to full capabilities over a clean workspace
•Single interface to create ad hoc or pixel perfect reports, frees up IT

Analytics when, where, and how you need them
- Simple intuitive interface
- Smart search works in context
- Personalized experience
- Scheduling and alerts
- Interactive content available on-line or off-line

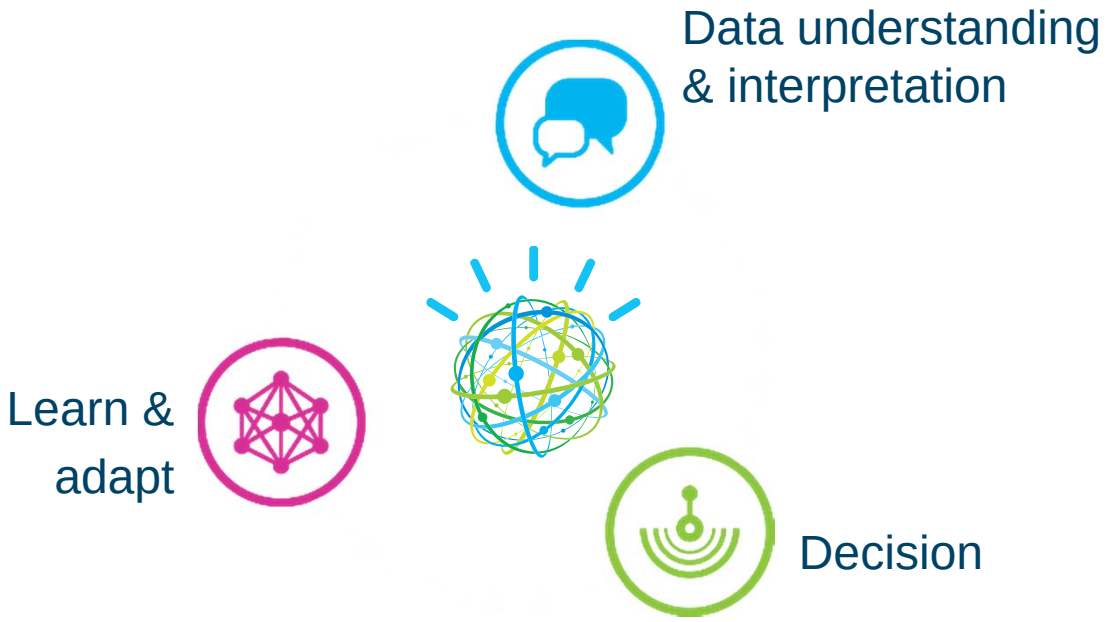
Analytics you can trust for confident action
Confident action comes from access to curated data that eliminate risk and debate over numbers.
- Data protected with layers of permissions, authentication, and history
- Report integrity maintained regardless of range of inputs across business
- Controls to protect data whether you're creating one report for many or many are creating one report
- Scheduling and alerts

http://www-03.ibm.com/software/products/en/cognos-analytics#othertab1

What is Cognitive Analytics ?

***Cognitive**: Psychological processes involved in acquisition and understanding of knowledge, formation of beliefs and attitudes, and decision making and problem solving.*

What is Cognitive computing?



Cognitive Analytics: IBM Watson Explorer

Systems that learn, understand, reason & interact

Cognitive Business requires **access to the right data, a trusted system** to hold that data and the ability to gain meaningful insights in time to affect outcomes.

Cognitive requires keeping up with the customer via mobile devices and **connecting to other sources of information** through the cloud to see the full picture of what's happening.

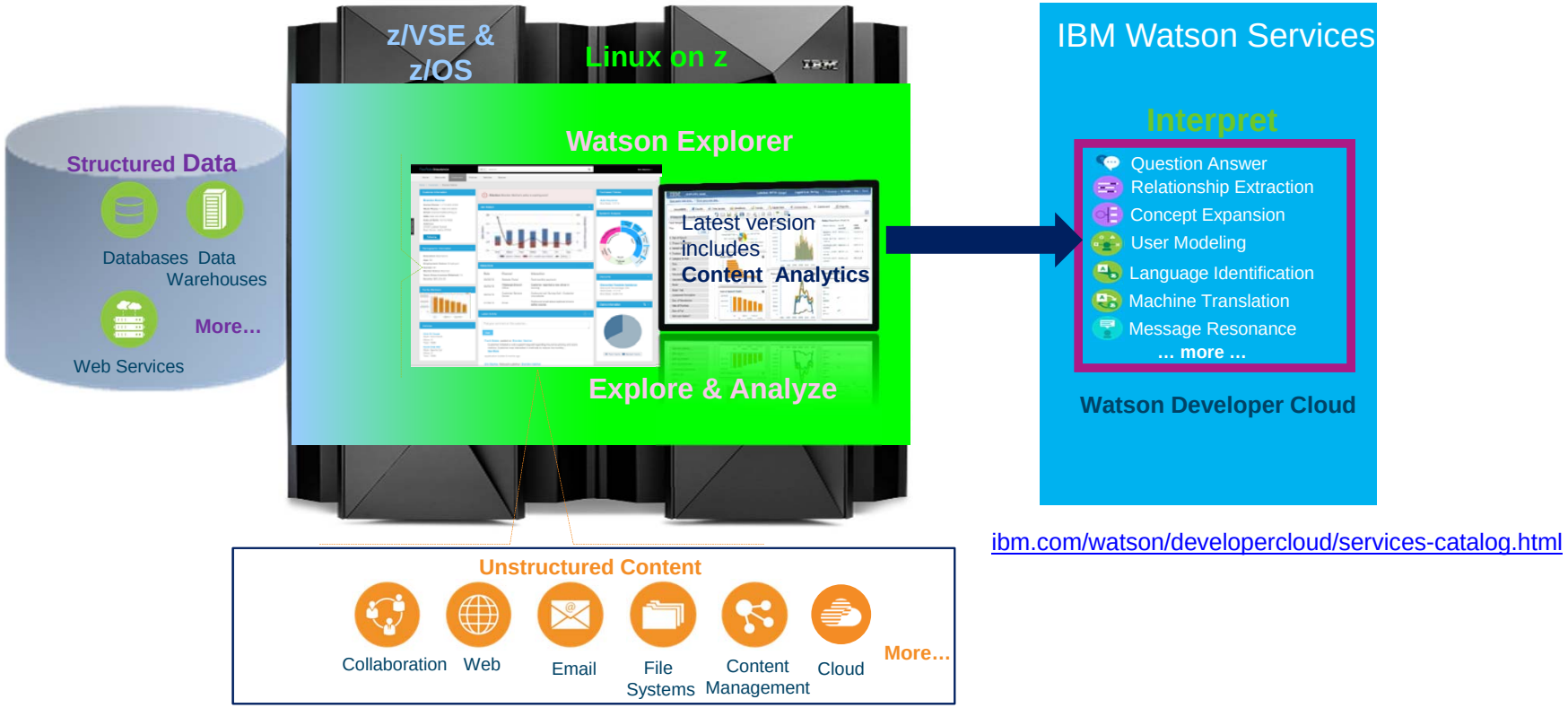
Cognitive requires systems with **analytics integrated into the business process** so that intelligence can be gained and actions taken while they still matter.

Cognitive brings all these qualities together

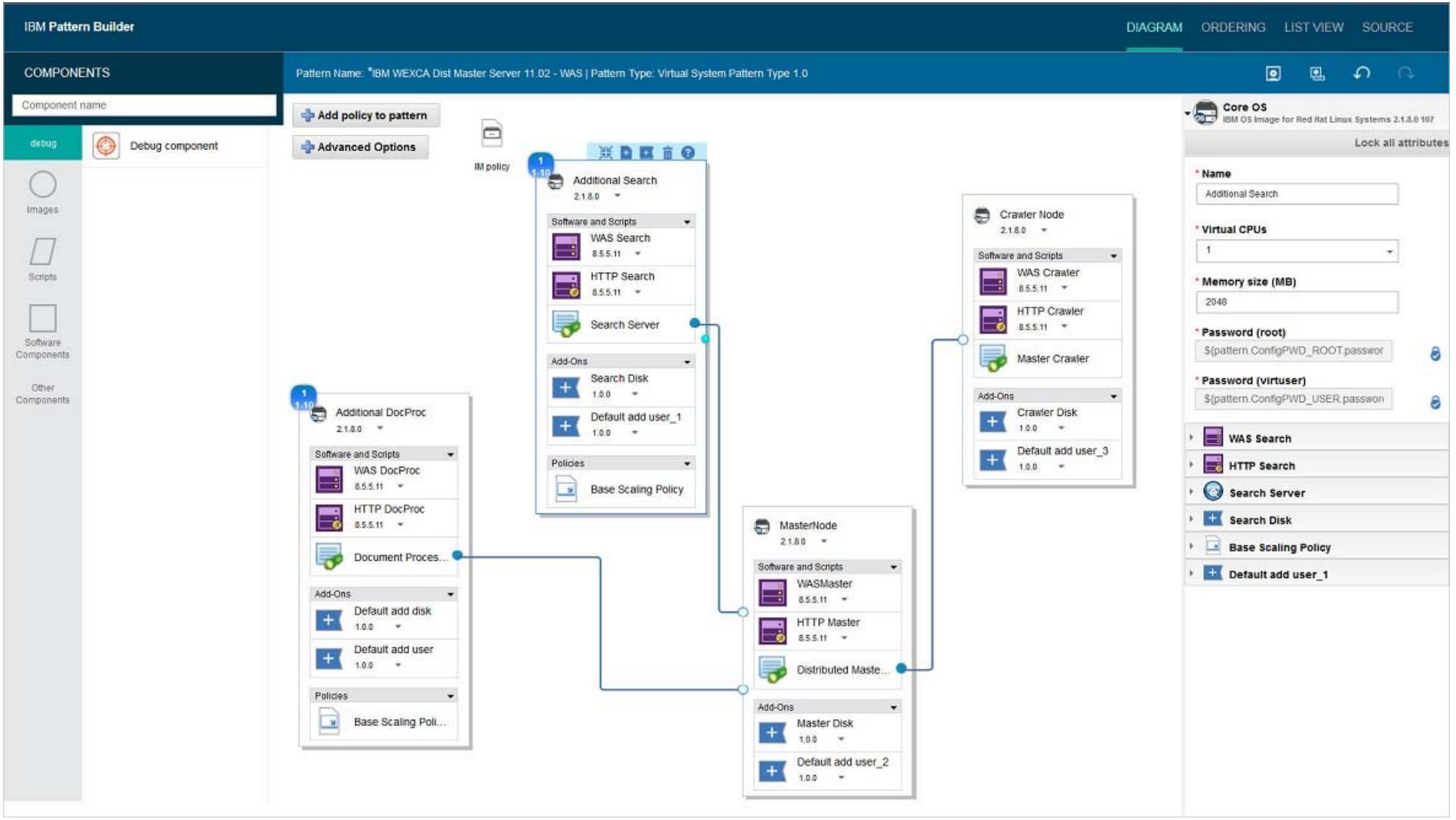
Cognitive Analytics = Digital Business + Digital Intelligence

IBM Watson Explorer

Analytics – Foundation for Digital Business
Watson technology – the Digital Intelligence for Cognitive Computing



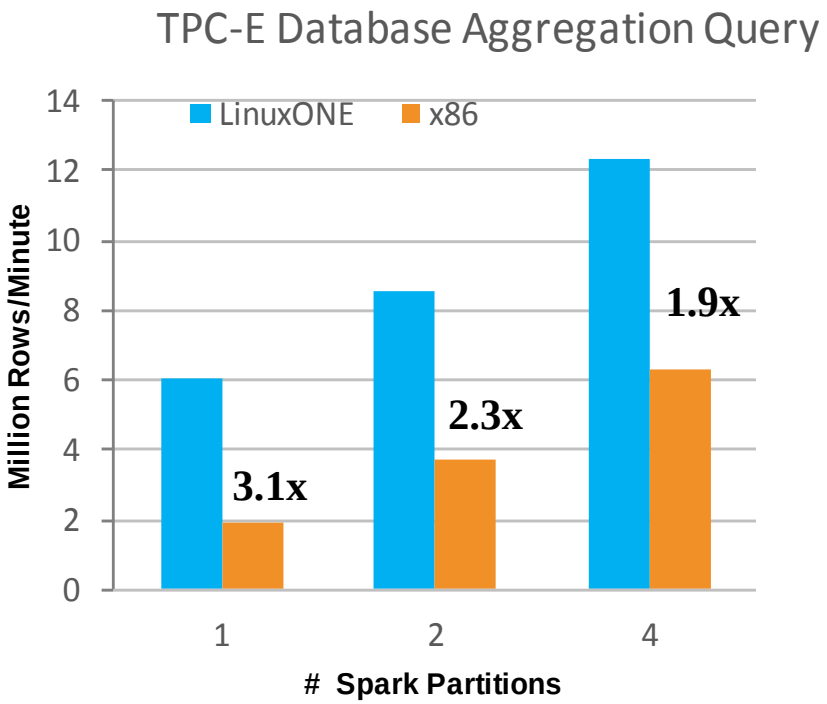
ibm.com/watson/developercloud/services-catalog.html



https://www.ibm.com/us-en/marketplace/watson-explorer-content-analytics?lnk=STW_US_MYIBME_C2_BLC&lnk2=learn_watson-explorer-content-analytics

Analytics workloads run faster on the Linux on z platform

- Test using independent OLTP brokerage database and analytical queries
  - Apache Spark with map-reduce script performs fetch, query and aggregation
    - 1 master and 4 worker JVMs
  - Oracle used for OLTP database, hosted on our platform
    - 348 M rows of brokerage trade data
- Spark co-located on our platform drove **up to 3x** more throughput than Spark running off platform on x86

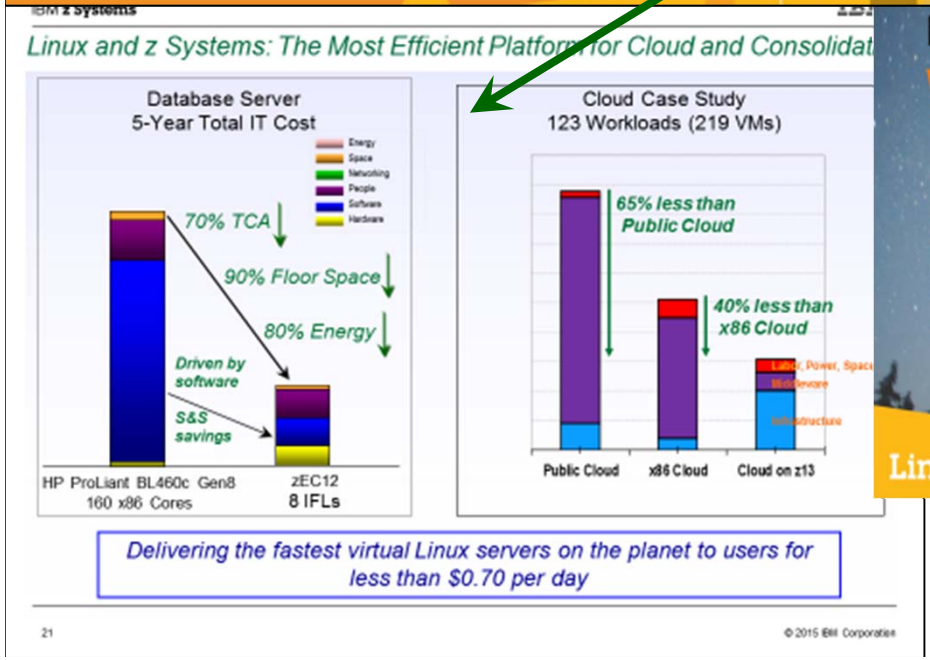


Test ran one LinuxONE guest, RHEL 6.5 on LinuxONE & z/VM 6.3 with 32 vCPUs (16 cores with SMT) and 512GB memory, Spark 1.5.0, IBM JDK1.7-SR3
Test ran on unvirtualized x86 with 16 cores on Intel(R) Xeon(R) CPU E5-2698 v3 @ 2.30GHz, 512GB memory running SLES 11.3, Spark 1.4.1, JDK1.8
TPC-E 20K scale on Oracle database V12
Scala map-reduce script (aggregation query) on TPC-E 20K scale Trade table using 1 master and 4 worker JVMs
Out-of-box, default parameters, no tuning

Linux on z Systems Technology: Not a "box" but "Linux Your Way"

"Linux Your Way" means a commitment to:

- ***An Open Source Ecosystem***
- ***Unmatched TCA with unmatched scale***
- ***The best performance***



Increased retail sales revenue through point-of-sale & suggest-sale insight

Business Challenge

How to improve customer service and satisfaction in order to drive greater revenue.

Technical Challenge

Existing analytic processes were unable to manage the analysis of historic and transaction data from Petrol's retail stores, service stations and home oil/gas businesses.

Solution

Implemented IBM DB2 Analytics Accelerator to support high performance queries and IBM SPSS to make real time, point of sale product recommendations.

"IBM provides us with tools that align with smarter commerce, enabling us to deliver the right message to the right person at the right time, to understand product affinities and intelligently drive the sale all in a customer centric way"

External Links: Case Study

Building an open & flexible strategic analytics platform for the future

Business Challenge

SDV is a German Bank providing retail services for over 100 years. They needed to find the right platform that will fulfill regulatory requirements and expand to provide high-end, real-time analytic solutions.

Technical Challenge

The existing infrastructure was challenged in meeting the fulfillment and regulatory requirement of BCBS 239 and was not capable of supporting the requirements of new projects.

Solution

Selecting the IBM LinuxONE System allows the bank to consolidate their analytic environment on one platform using DB2 on z Systems Linux. The platform provides an innovative, flexible infrastructure to expand their analytic capabilities to an open community-driven ecosystem provided by Linux.

Unleashing new growth and operational efficiency with an infrastructure transformation

Business Challenge

Growing fast, Brazilian credit union system Sicoob must cope with extra transactions at short notice. How could it combat the increased IT complexity that resulted without impacting service quality.

Technical Challenge

To keep pace with their rapid business growth Sicoob did not have the IT infrastructure to support reliable 24/7 service and mobile access for their customers.

Solution

Selecting IBM z Systems as the strategic platform. Sicoob migrated and consolidated member databases to IBM DB2 LUW with BLU Acceleration. They deployed IBM InfoSphere DataStage and IBM Cognos running in a Linux environment on the mainframe.

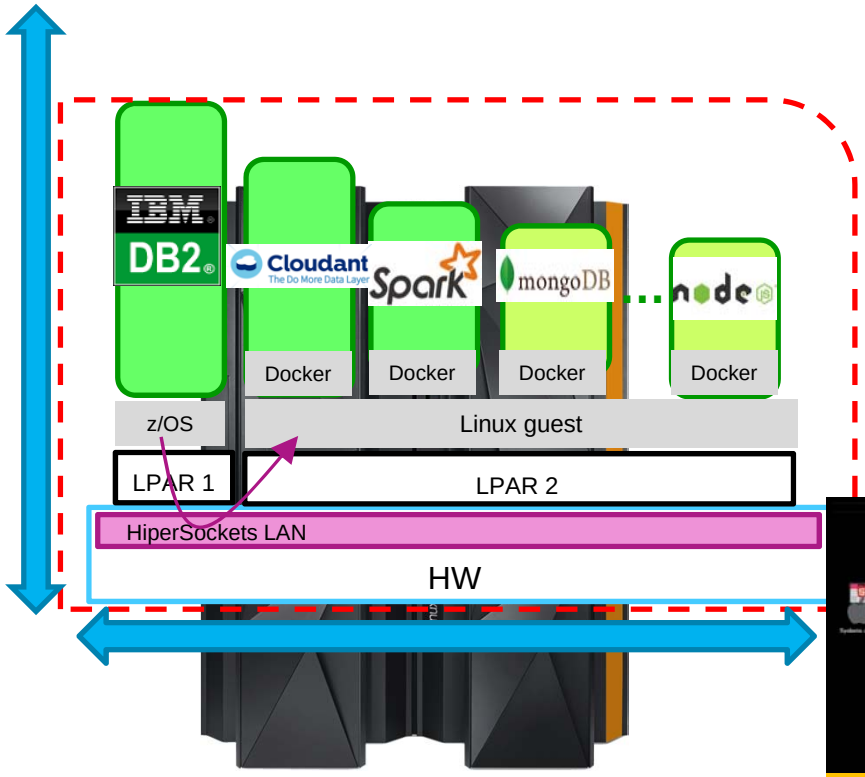
"IBM Challenges and opportunities have led us to restructure our technology infrastructure and adopt IBM System z technology, which guarantees greater stability and performance for our products and services. This facilitates our growth, by lowering the cost of maintenance and administration in the production environment, and by reducing power consumption in the data center.... "

– Denio Rodrigues, IT Executive

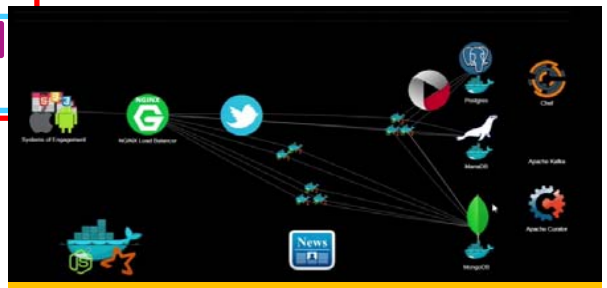
External Links: Case Study2, YouTube Video1, YouTube Video2

IBM Z = SOR+SOI+SOE in a Box

System Of Record (SOR), System Of Insight (SOI), System Of Engagement (SOE)



- Compose high-performance scalable applications
- Dynamically and seamlessly re-allocate resources between micro-services
- Provide right-time analytics
- Provide powerful engagement



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

Putting it all together – Open Source running LinuxONE and IBM z Systems Demo: "Scalable Financial Trading Analysis & Insights"

Input Data

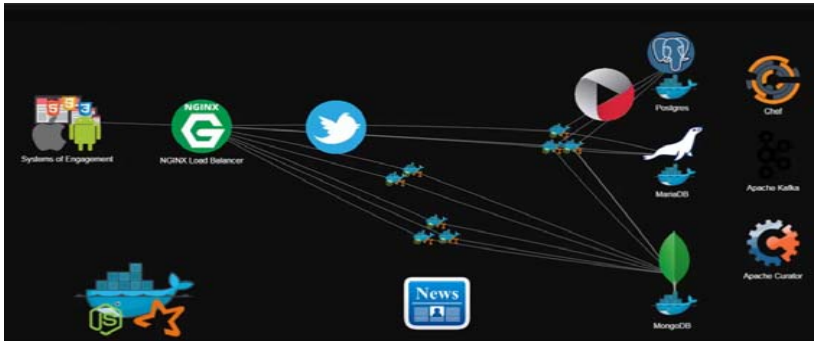
Historical S&P 500 Index

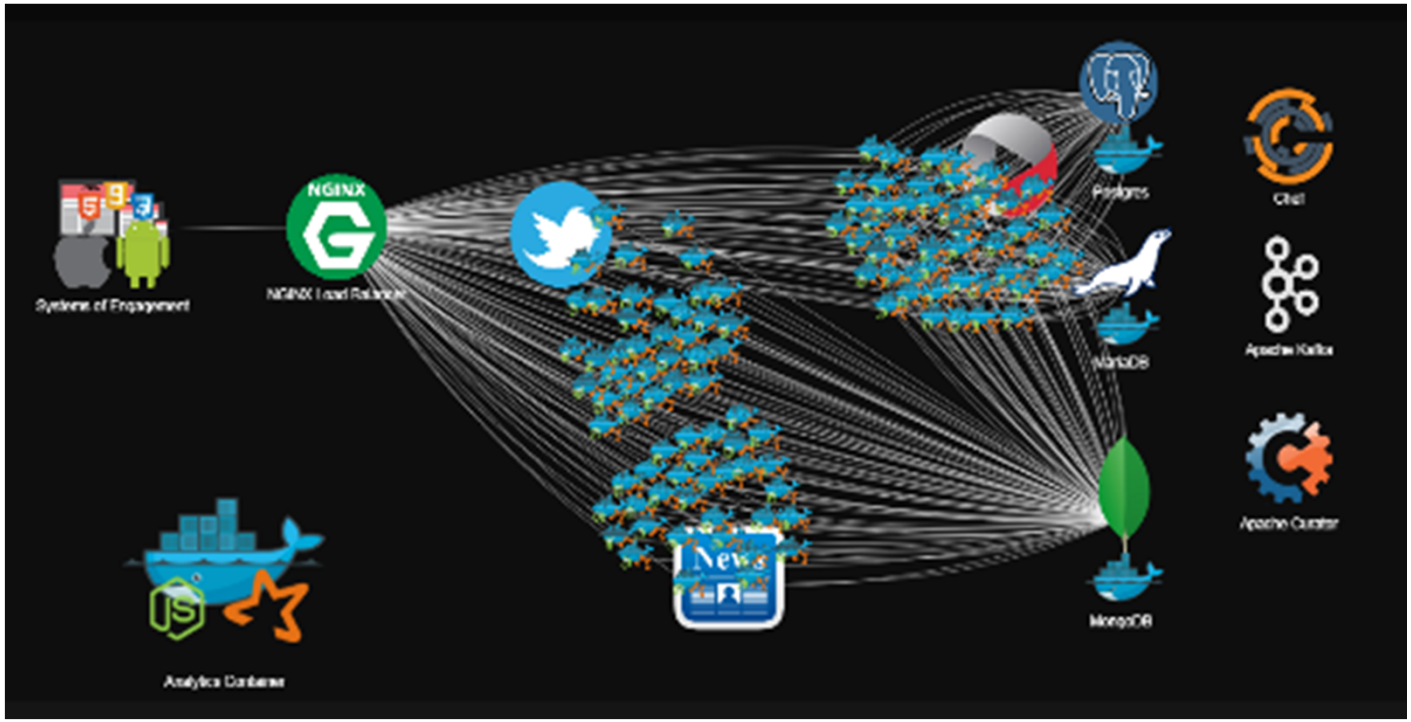
News Feed

Sentiment Analysis

Trade Transactions

Geospatial Analysis

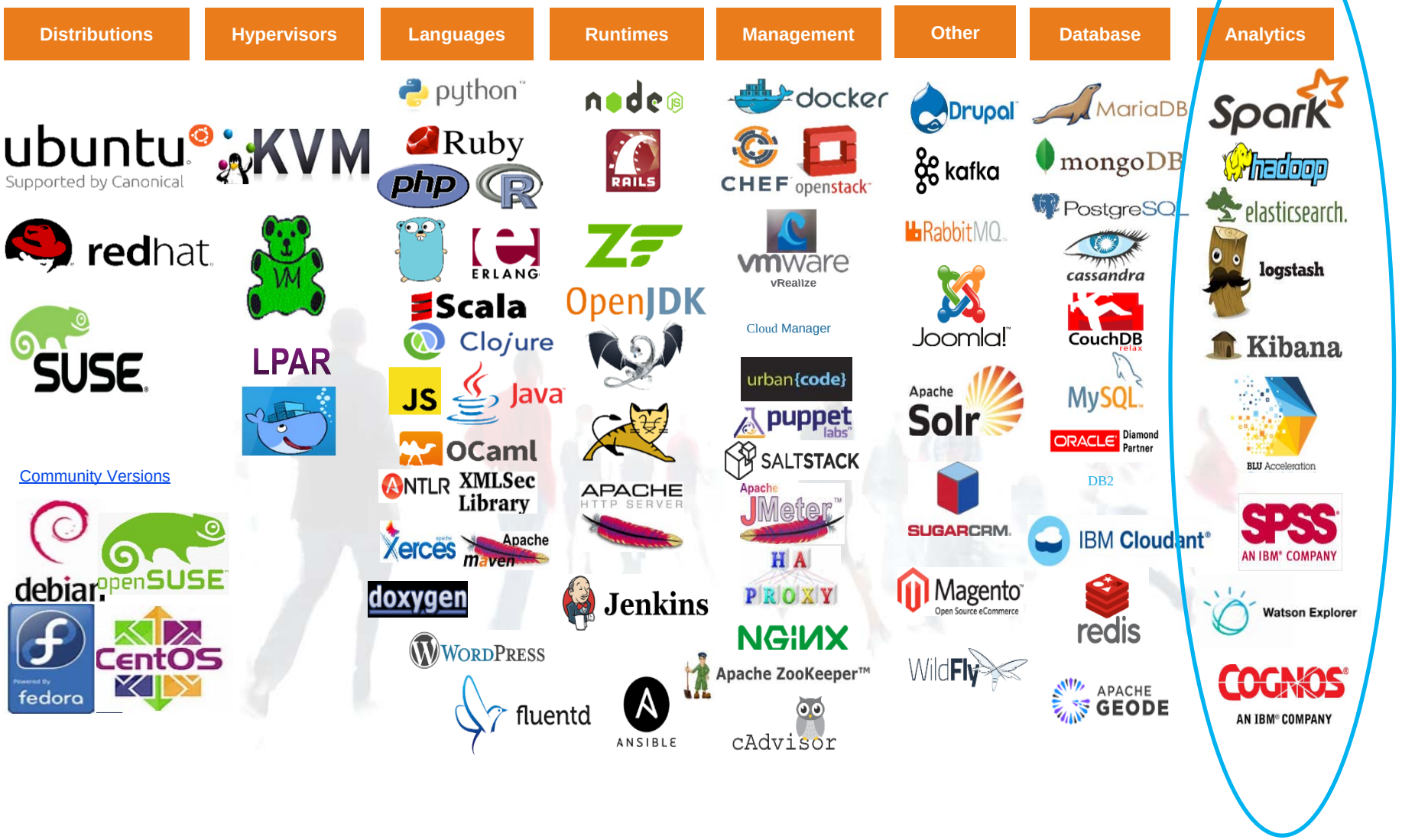




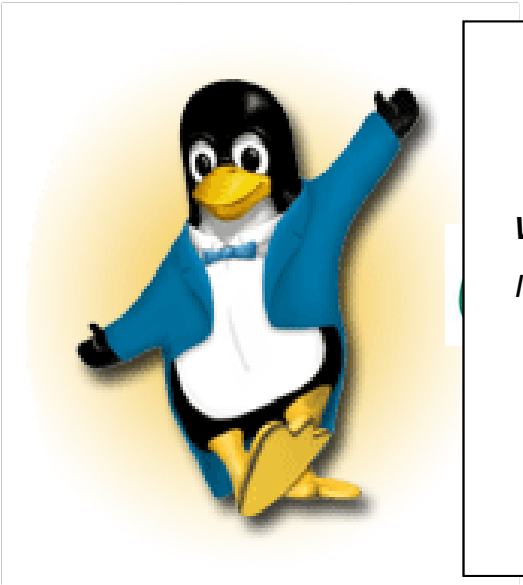
https://www.youtube.com/watch?v=VWBNoIwGEjo

Linux your Way - Greater flexibility and choice includes Analytics

Choose the distribution, runtime, hypervisor, database and analytics – it's the Linux you know and love with the openness, flexibility and agility you need for you business.



Questions?



Wilhelm Mild

IBM Executive IT Architect

The Open Group Open Certified Distinguished Architect

IBM

IBM Deutschland Research & Development GmbH
Schönaicher Strasse 220
71032 Böblingen, Germany

Office: +49 (0)7031-16-3796
wilhelm.mild@de.ibm.com

Trademarks

The following are trademarks of the International Business Machines Corporation in the United States and/or other countries.

Chiphopper	developerWorks*	FlashSystem	HyperSwap*	IMS	PR/SM	z/Architectur	z Systems
CICS*	DS8000*	GDPS*	IBM*	LinuxONE	Storwize*	e*	z/OS*
DB2*	ECKD	GPFS	Ibm.com	LinuxONE Emperor	XIV*	zEnterprise*	z/VSE*
DB2 Connect	FICON*	HiperSockets	IBM (logo)*	LinuxONE Rockhopper	z13, z14	z/OS*	z/VM*

* Registered trademarks of IBM Corporation

The following are trademarks or registered trademarks of other companies.

Adobe, the Adobe logo, PostScript, and the PostScript logo are either registered trademarks or trademarks of Adobe Systems Incorporated in the United States, and/or other countries.
Cell Broadband Engine is a trademark of Sony Computer Entertainment, Inc. in the United States, other countries, or both and is used under license therefrom.
Intel, Intel logo, Intel Inside, Intel Inside logo, Intel Centrino, Intel Centrino logo, Celeron, Intel Xeon, Intel SpeedStep, Itanium, and Pentium are trademarks or registered trademarks of Intel Corporation or its subsidiaries in the United States and other countries.
IT Infrastructure Library is a registered trademark of the Central Computer and Telecommunications Agency which is now part of the Office of Government Commerce.
ITIL is a registered trademark, and a registered community trademark of the Office of Government Commerce, and is registered in the U.S. Patent and Trademark Office.
Java and all Java based trademarks and logos are trademarks or registered trademarks of Oracle and/or its affiliates.
Linear Tape-Open, LTO, the LTO Logo, Ultrium, and the Ultrium logo are trademarks of HP, IBM Corp. and Quantum in the U.S. and Linux is a registered trademark of Linus Torvalds in the United States, other countries, or both. Microsoft, Windows, Windows NT, and the Windows logo are trademarks of Microsoft Corporation in the United States, other countries, or both. OpenStack is a trademark of OpenStack LLC. The OpenStack trademark policy is available on the OpenStack website.
TEALEAF is a registered trademark of Tealeaf, an IBM Company.
Windows Server and the Windows logo are trademarks of the Microsoft group of countries.
Worklight is a trademark or registered trademark of Worklight, an IBM Company.
UNIX is a registered trademark of The Open Group in the United States and other countries
Docker is a registered trademarks of Docker, Inc. in the United States and/or other countries
.

* Other product and service names might be trademarks of IBM or other companies.

Notes:

Performance is in Internal Throughput Rate (ITR) ratio based on measurements and projections using standard IBM benchmarks in a controlled environment. The actual throughput that any user will experience will vary depending upon considerations such as the amount of multiprogramming in the user's job stream, the I/O configuration, the storage configuration, and the workload processed. Therefore, no assurance can be given that an individual user will achieve throughput improvements equivalent to the performance ratios stated here. IBM hardware products are manufactured from new parts, or new and serviceable used parts. Regardless, our warranty terms apply.

All customer examples cited or described in this presentation are presented as illustrations of the manner in which some customers have used IBM products and the results they may have achieved. Actual environmental costs and performance characteristics will vary depending on individual customer configurations and conditions.

This publication was produced in the United States. IBM may not offer the products, services or features discussed in this document in other countries, and the information may be subject to change without notice. Consult your local IBM business contact for information on the product or services available in your area. All statements regarding IBM's future direction and intent are subject to change or withdrawal without notice, and represent goals and objectives only.

Information about non-IBM products is obtained from the manufacturers of those products or their published announcements. IBM has not tested those products and cannot confirm the performance, compatibility, or any other claims related to non-IBM products. Questions on the capabilities of non-IBM products should be addressed to the suppliers of those products.

Prices subject to change without notice. Contact your IBM representative or Business Partner for the most current pricing in your geography.

This information provides only general descriptions of the types and portions of workloads that are eligible for execution on Specialty Engines (e.g, zIIPs, zAAPs, and IFLs) ("SEs"). IBM authorizes customers to use IBM SE only to execute the processing of Eligible Workloads of specific Programs expressly authorized by IBM as specified in the "Authorized Use Table for IBM Machines" provided at www.ibm.com/systems/support/machine_warranties/machine_code/aut.html ("AUT"). No other workload processing is authorized for execution on an SE. IBM offers SE at a lower price than General Processors/Central Processors because customers are authorized to use SEs only to process certain types and/or amounts of workloads as specified by IBM in the AUT.

Notices and Disclaimers

U.S. Government Users Restricted Rights - Use, duplication or disclosure restricted by GSA ADP Schedule Contract with IBM.

Information in these presentations (including information relating to products that have not yet been announced by IBM) has been reviewed for accuracy as of the date of initial publication and could include unintentional technical or typographical errors. IBM shall have no responsibility to update this information. THIS DOCUMENT IS DISTRIBUTED "AS IS" WITHOUT ANY WARRANTY, EITHER EXPRESS OR IMPLIED. IN NO EVENT SHALL IBM BE LIABLE FOR ANY DAMAGE ARISING FROM THE USE OF THIS INFORMATION, INCLUDING BUT NOT LIMITED TO, LOSS OF DATA, BUSINESS INTERRUPTION, LOSS OF PROFIT OR LOSS OF OPPORTUNITY. IBM products and services are warranted according to the terms and conditions of the agreements under which they are provided.

Any statements regarding IBM's future direction, intent or product plans are subject to change or withdrawal without notice.

Performance data contained herein was generally obtained in a controlled, isolated environments. Customer examples are presented as illustrations of how those customers have used IBM products and the results they may have achieved. Actual performance, cost, savings or other results in other operating environments may vary.

References in this document to IBM products, programs, or services does not imply that IBM intends to make such products, programs or services available in all countries in which IBM operates or does business.

Workshops, sessions and associated materials may have been prepared by independent session speakers, and do not necessarily reflect the views of IBM. All materials and discussions are provided for informational purposes only, and are neither intended to, nor shall constitute legal or other guidance or advice to any individual participant or their specific situation.

It is the customer's responsibility to insure its own compliance with legal requirements and to obtain advice of competent legal counsel as to the identification and interpretation of any relevant laws and regulatory requirements that may affect the customer's business and any actions the customer may need to take to comply with such laws. IBM does not provide legal advice or represent or warrant that its services or products will ensure that the customer is in compliance with any law.

Notices and Disclaimers (con’t.)

Information concerning non-IBM products was obtained from the suppliers of those products, their published announcements or other publicly available sources. IBM has not tested those products in connection with this publication and cannot confirm the accuracy of performance, compatibility or any other claims related to non-IBM products. Questions on the capabilities of non-IBM products should be addressed to the suppliers of those products. IBM does not warrant the quality of any third-party products, or the ability of any such third-party products to interoperate with IBM’s products. IBM EXPRESSLY DISCLAIMS ALL WARRANTIES, EXPRESSED OR IMPLIED, INCLUDING BUT NOT LIMITED TO, THE IMPLIED WARRANTIES OF MERCHANTABILITY AND FITNESS FOR A PARTICULAR PURPOSE.

The provision of the information contained herein is not intended to, and does not, grant any right or license under any IBM patents, copyrights, trademarks or other intellectual property right.

•IBM, the IBM logo, ibm.com, Bluemix, Blueworks Live, CICS, Clearcase, DOORS®, Enterprise Document Management System™, Global Business Services ®, Global Technology Services ®, Information on Demand, ILOG, Maximo®, MQIntegrator®, MQSeries®, Netcool®, OMEGAMON, OpenPower, PureAnalytics™, PureApplication®, pureCluster™, PureCoverage®, PureData®, PureExperience®, PureFlex®, pureQuery®, pureScale®, PureSystems®, QRadar®, Rational®, Rhapsody®, SoDA, SPSS, StoredIQ, Tivoli®, Trusteer®, urban{code}®, Watson, WebSphere®, Worklight®, X-Force® and System z® Z/OS, are trademarks of International Business Machines Corporation, registered in many jurisdictions worldwide. Other product and service names might be trademarks of IBM or other companies. A current list of IBM trademarks is available on the Web at "Copyright and trademark information" at: www.ibm.com/legal/copytrade.shtml.