Neue Wege zur Speicherung, Verarbeitung und Analyse von IoT-Daten

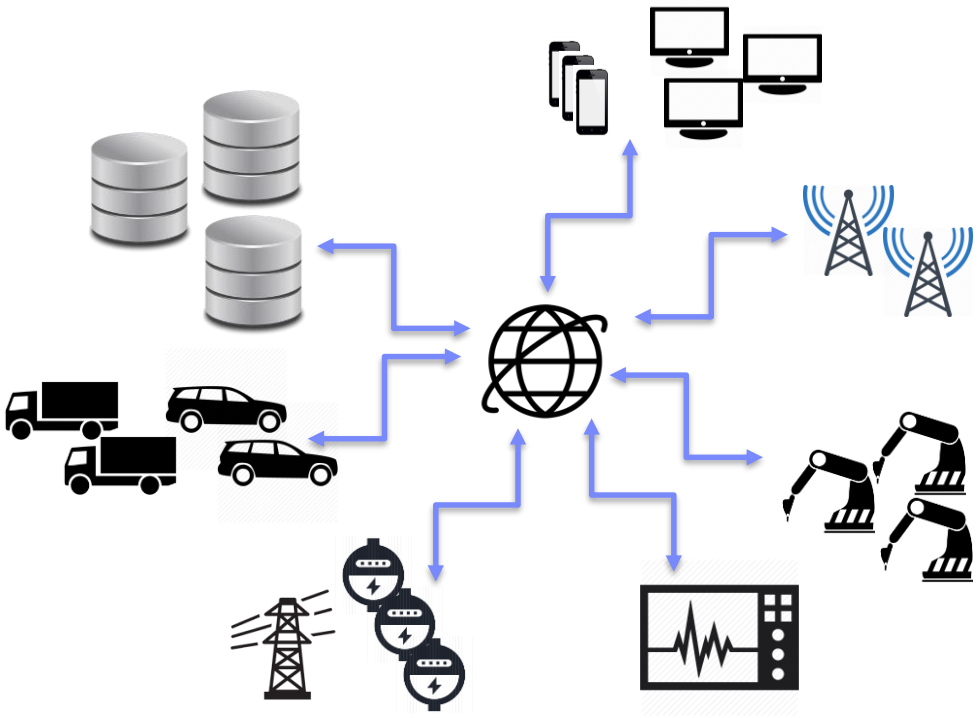
Andreas Weininger, IBM
Andreas.Weininger@de.ibm.com
@aweininger

Agenda

- Neue Anforderungen durch IoT
- Architektur zur Lösung
- Technologien:
  - Hochgeschwindigkeitsdatenverarbeitung
  - Datenvirtualisierung

Neue Anforderungen durch IoT

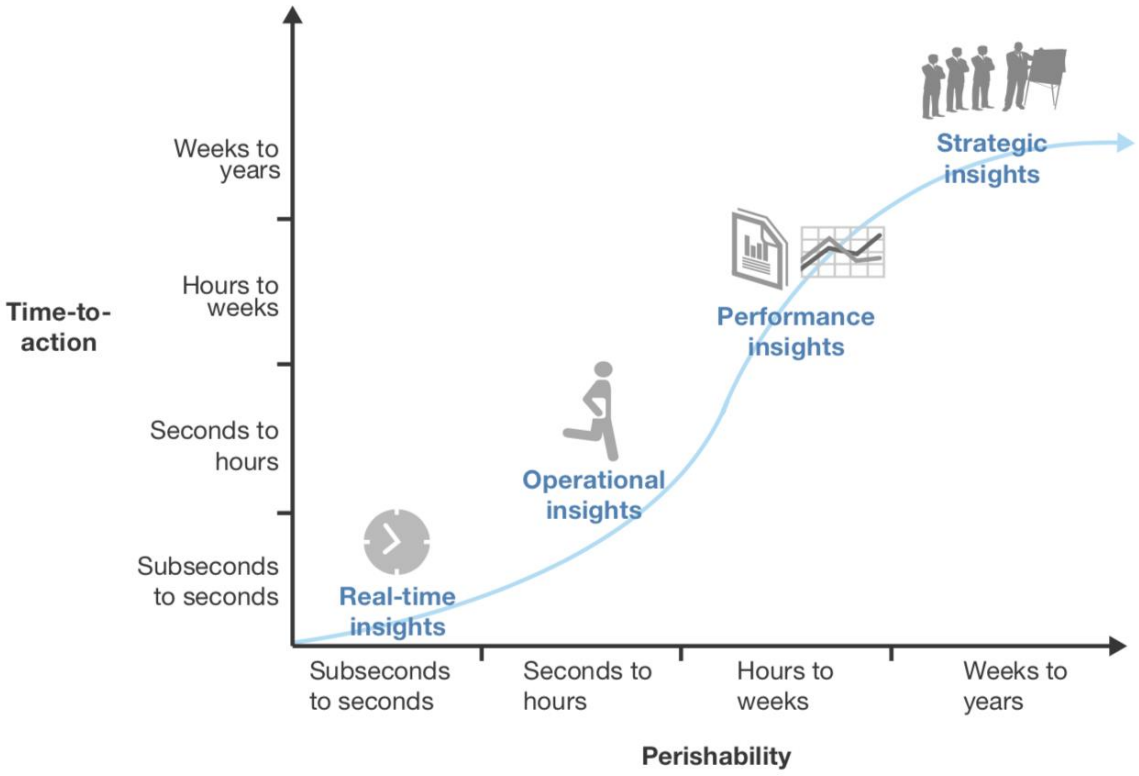
Daten sind überall und immer heterogener



Perishable Insights: Insights That Are Actionable

Source: Forrester Perishable Insights-Stop Wasting Money on Unactionable Insights Report 2016

- Being able to act on insights in real-time can provide maximum differentiation and value to businesses.
- To act on real time events, businesses need an Application Architecture that can ingest, analyze and act upon events as they happen in sub-second to seconds response time.
- Machine Learning combined with Reactive Programing frame works enable insights to be actioned automatically and stay current so business can react to events in sub-second timer frames

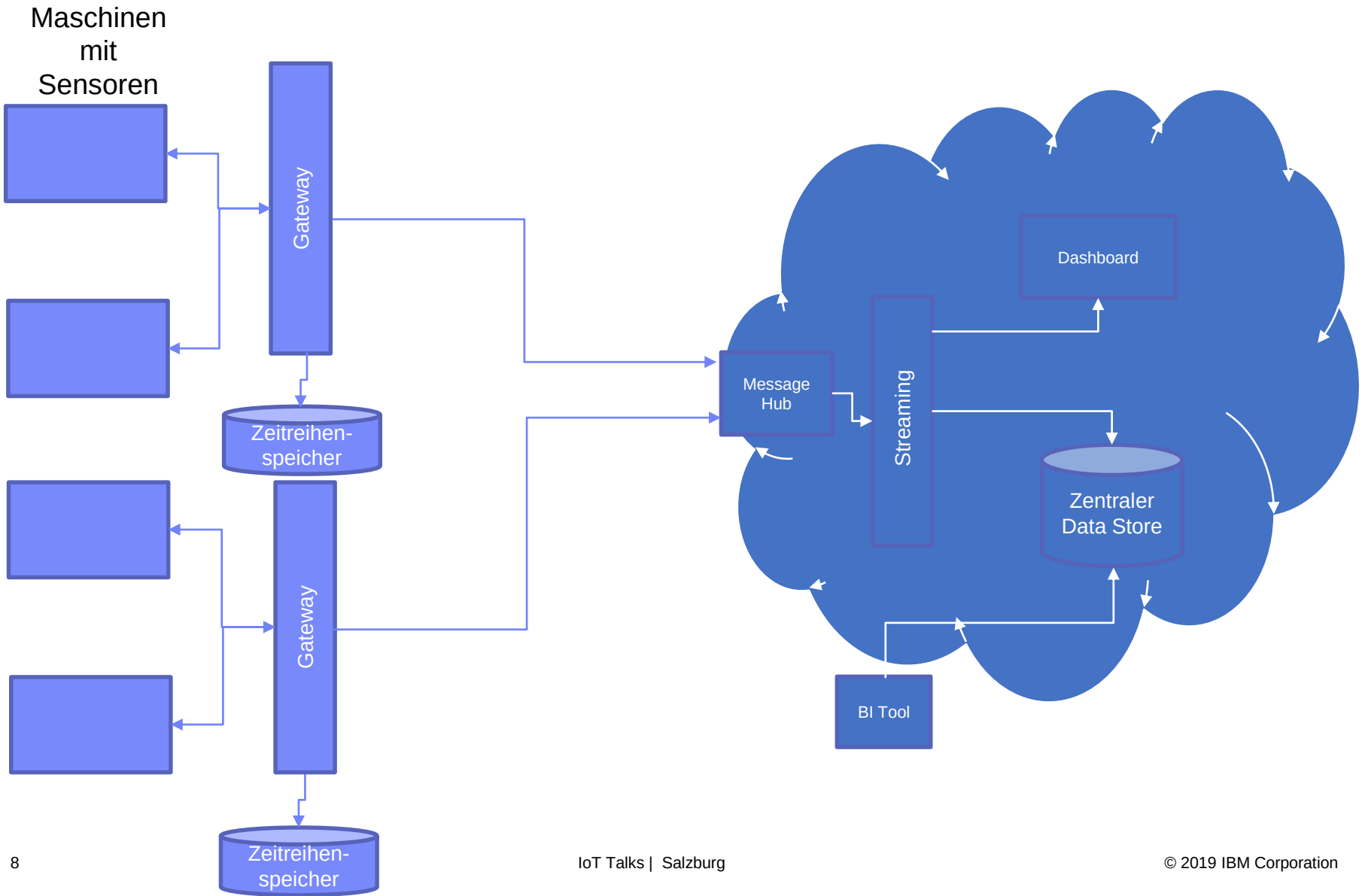


Was hat sich geändert?

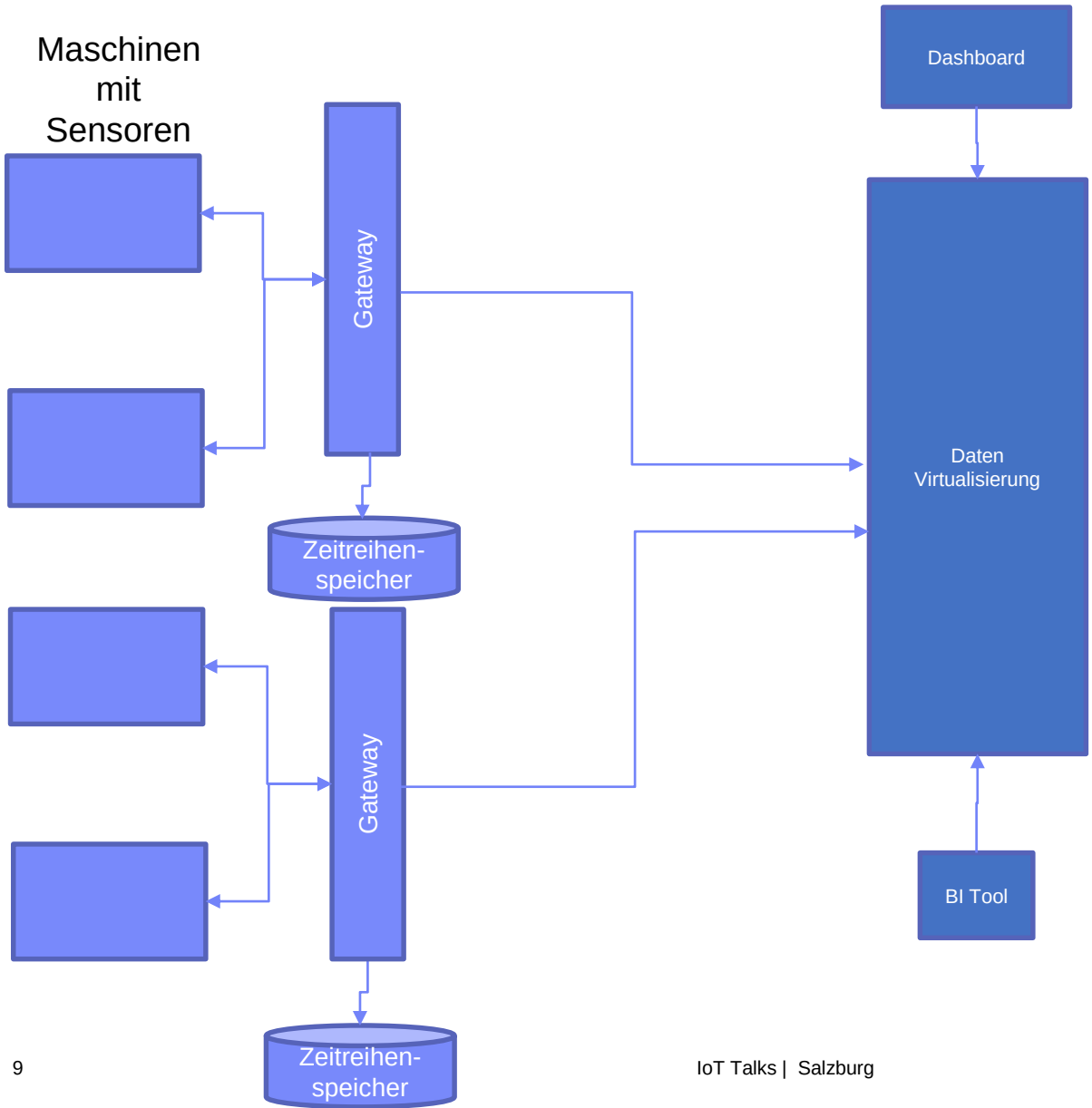
- IoT bietet neue Anforderungen an die Speicherung, Verarbeitung und Analyse von Daten:
  - Extrem große Datenmengen grosse Datenmengen fallen in sehr kurzer Zeit an
  - Maximale Zeitspanne zwischen Erzeugung der Daten und ihrer (evtl. komplexen) Auswertung eventuell gering
  - Von Sensoren generierte Datenmenge kann die verfügbare Bandbreite der Anbindung zum zentralen Auswertungssystem überschreiten.

Architektur zur Lösung

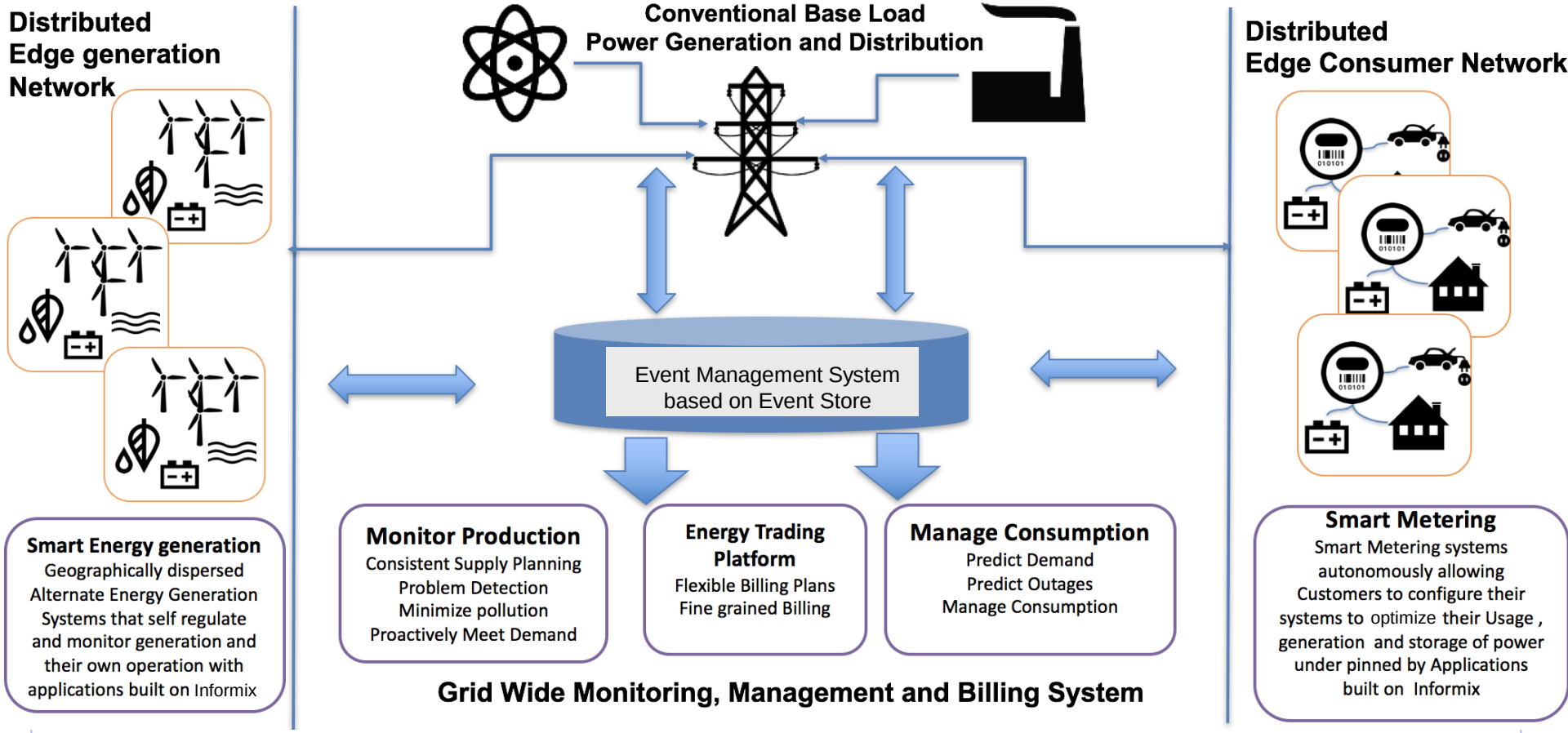
Architekturbeispiel Industrie 4.0



Architekturbeispiel Industrie 4.0



Architekturbeispiel Energie



Technologien

Hochgeschwindigkeitsdatenverarbeitung in Echtzeit

Beispiel: Connected Cars

Telematische Information:

- Geschwindigkeit, Schaltung, ABS, Airbag, Reifendruck, Lokation
- Mehr als 1500 Messwerte, einige in Millisekundenintervallen
- Wetter, locale Ereignisse
- Lokation: Verkehr
- Demographische Information, Fahrerverhalten

Möglichkeiten:

- Verständnis wie Kunden Produkt nutzen
- Wartungsplanung
- Geo-fenced Event Notification

Beispiel: Handel

Six Billion Mobile Phone Subscribers

Retail

iBeacon-type technologies

- Brick-and-mortar click stream analysis

Micro segmentation marketing

- Customer location and time opportunities

IBM Db2 Event Store Characteristics

① Lightning Fast Ingest[1]

- 1 Million inserts per second per node
- Ingest rate scales linearly with added nodes
- Memory-optimized indexing for fast lookups

② Real-time Analytics

- Real-time analytics over ALL ingested data
- OLAP queries faster than SparkSQL
- Integrated machine learning capabilities

③ Highly Available Atomic Transactions

- Replication for HA
- Configurable consistency
- Full insert Capabilities

④ Built for Open Hybrid Cloud

- Query elasticity through Spark
- Writes to shared storage in Parquet format
- No vendor lock-in

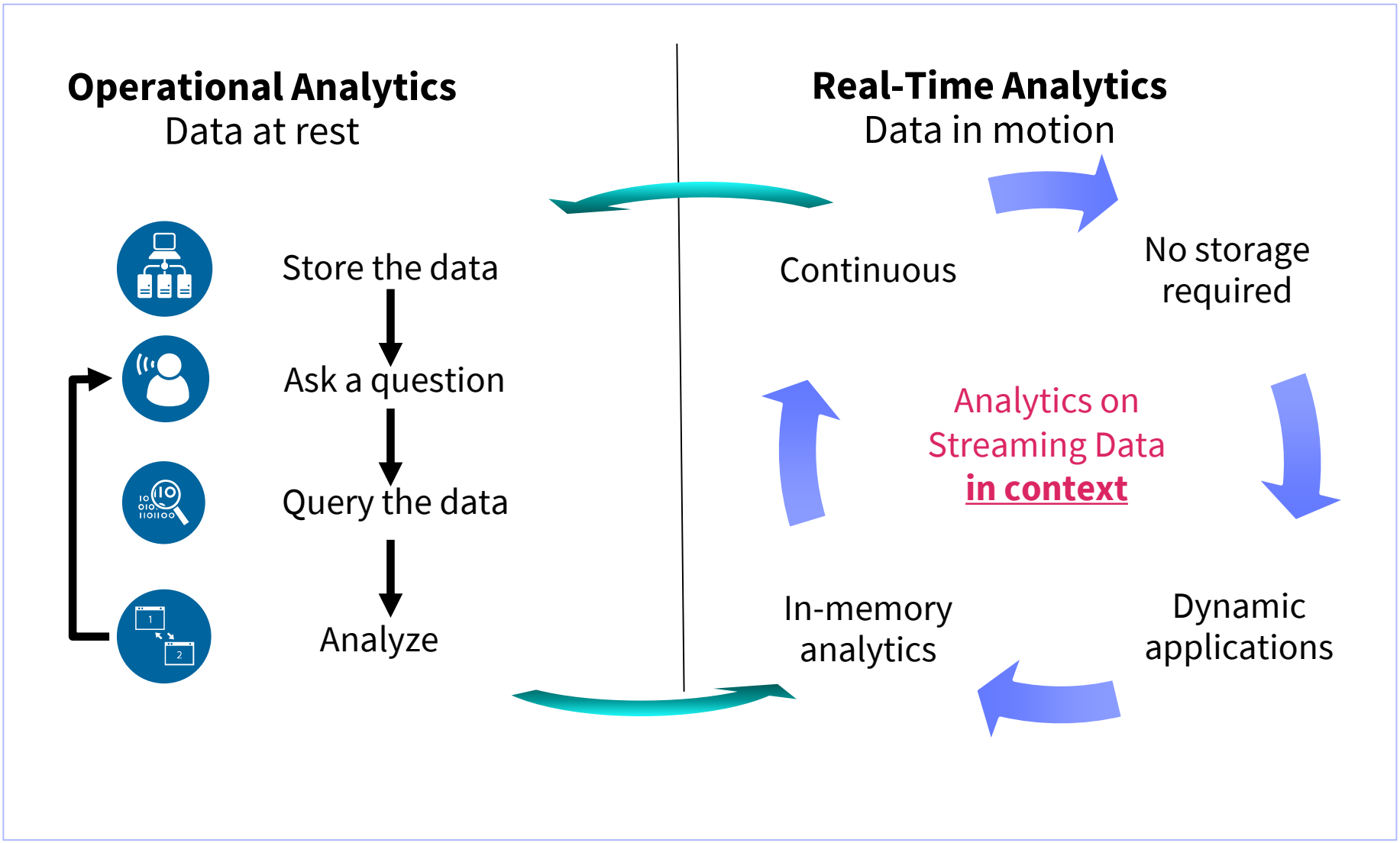
[1]Performance is based on measurements and projections using standard IBM benchmarks in a controlled environment. The actual throughput or performance that any user will experience will vary depending upon many factors, including considerations such as the amount of multiprogramming in the user's job stream, the I/O configuration, the storage configuration, and the workload processed. Therefore, no assurance can be given that an individual user will achieve results similar to those stated here.

Why Not Hadoop?

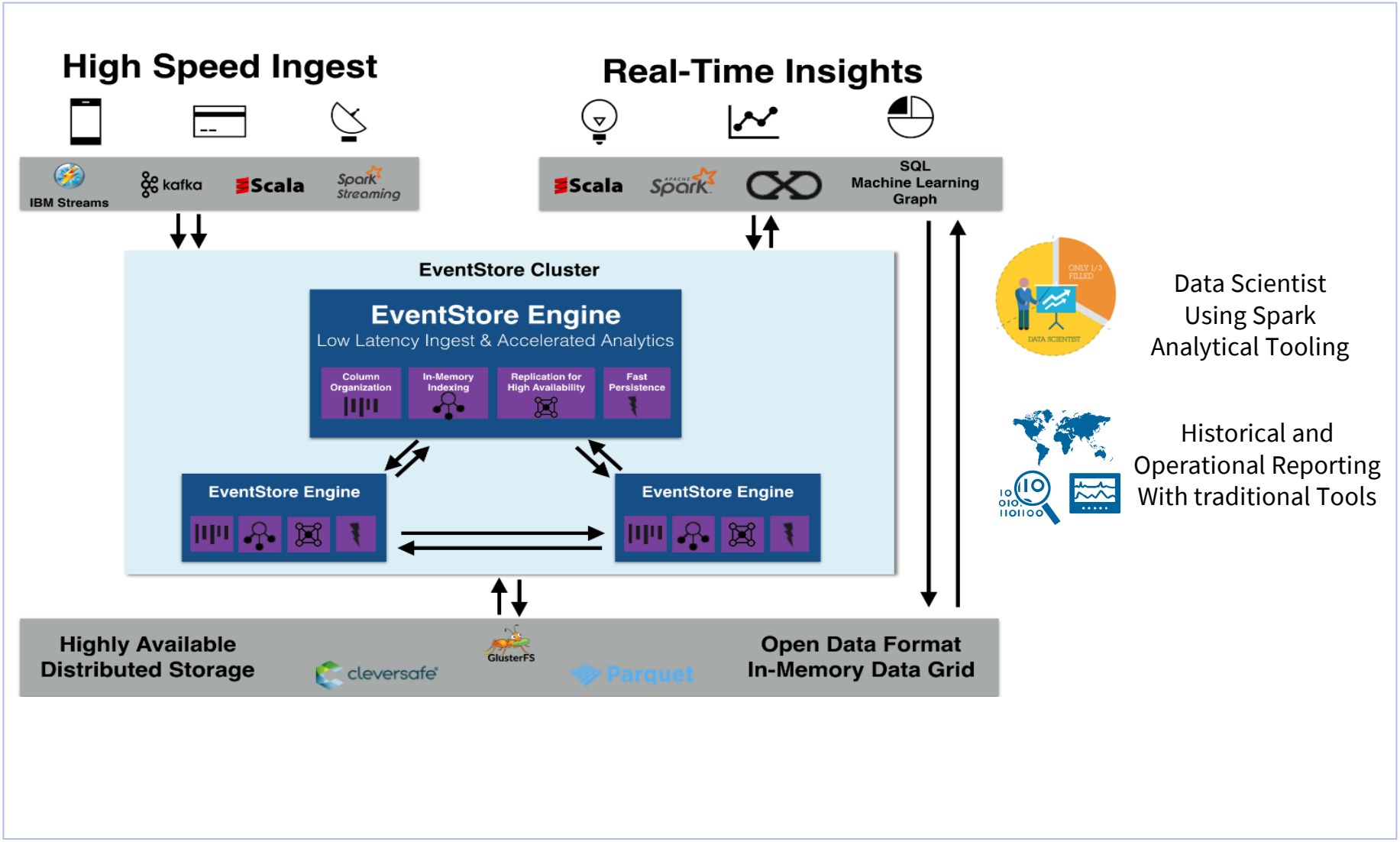
Hadoop is not designed for transactional systems
Hadoop is designed for large batch ingests

Db2 Event Store is a high-performance hybrid relational/analytics system
Db2 Event Store **continuously optimizes storage**
for high performance access

Real-Time Analytics: Db2 Event Store and IBM Streams

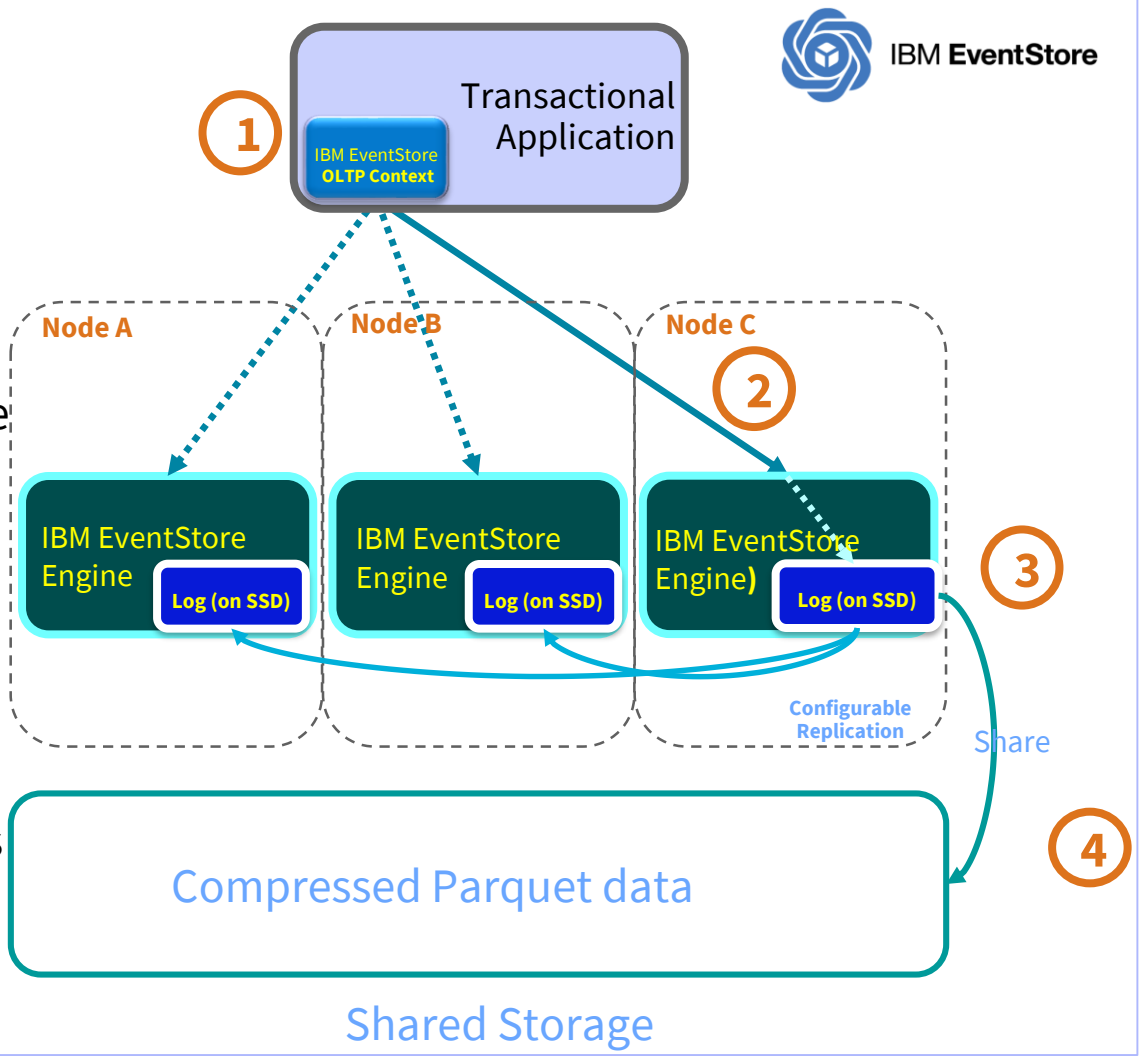


IBM Db2 Event Store System



IBM EventStore Ingest

1. Ingest occurs using Scala API or streaming sink
   - Scala API accessible through Spark-friendly Context
2. Batches of rows are formed and sent asynchronously from client to the appropriate EventStore nodes
3. Rows are placed in the queryable log, replicated to replica nodes, and reply is sent to client
4. After some time, data in logs is formed into Parquet blocks and written to the storage layer

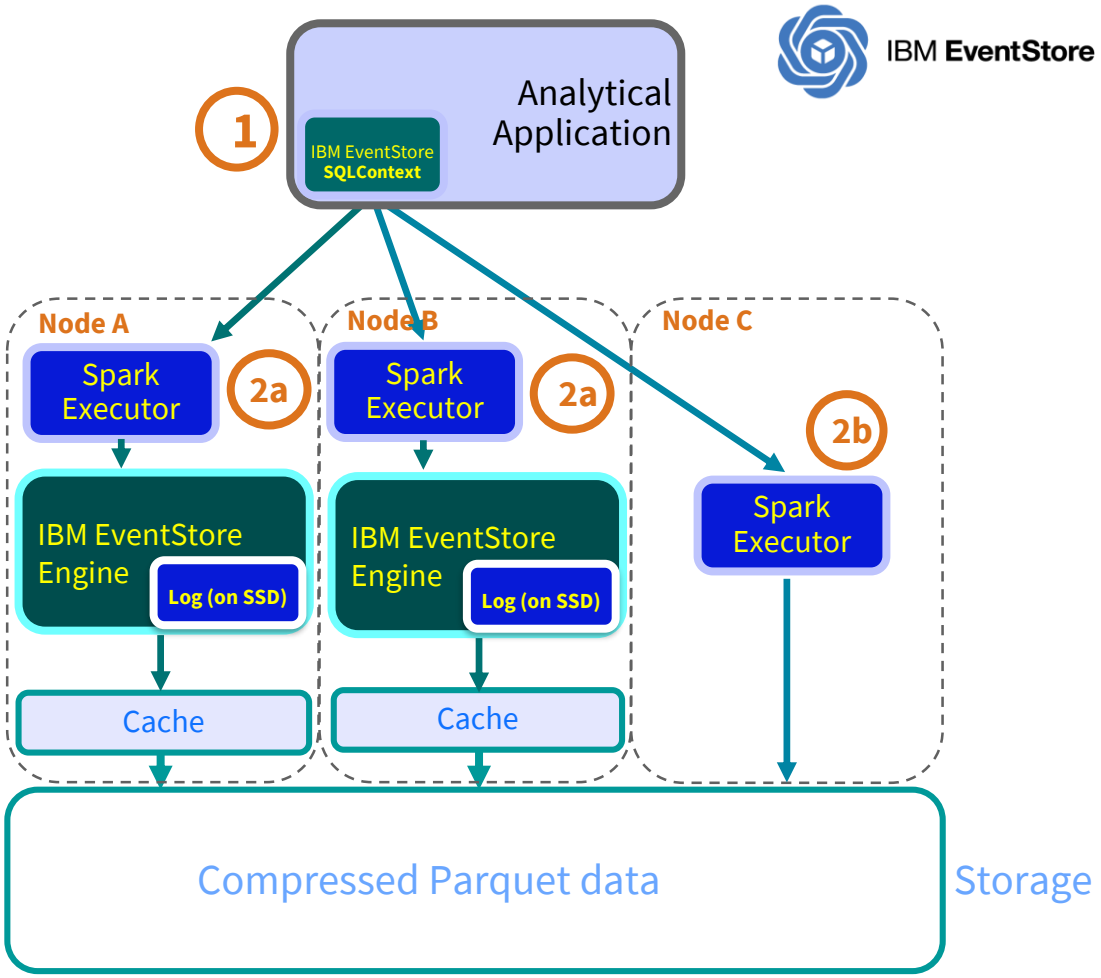


IBM EventStore Analytics

1. Analytics queries leverage the EventStore SQLContext which extents the Spark SQLContext
2. Queries are sent to either EventStore nodes, or vanilla Spark nodes depending on performance requirements and whether most recent (just inserted) data is required

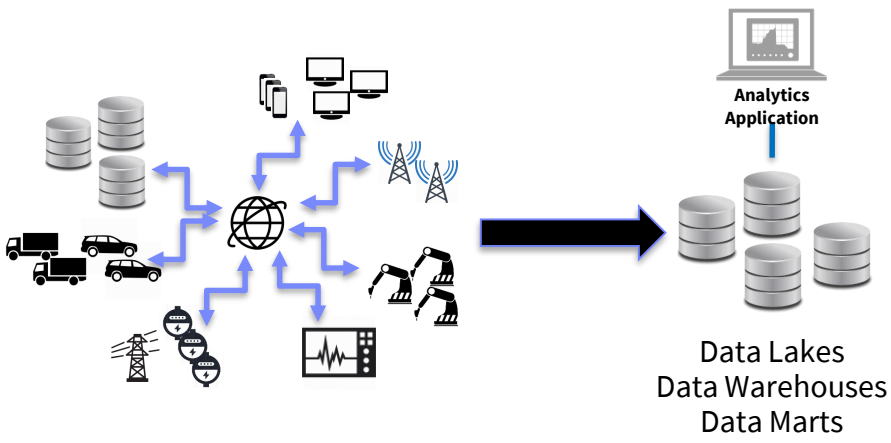
   a) Query is sent to EventStore nodes to retrieve most recent data and combine with shared data in cache or in storage layer

   b) Query is sent to Spark node(s) to read data all but most recent data from storage



Datenvirtualisierung

Performing Analytics Today



Costly and Complex

High Latency

Does not meet Business need

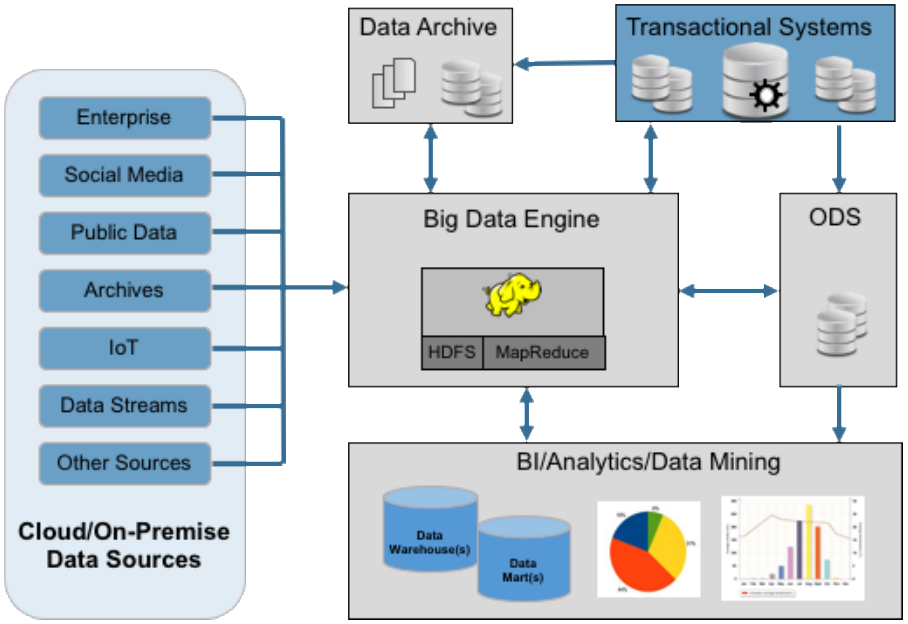
Compute resources at source not utilized

Error prone, data integrity challenges

Applications expect homogeneity

Not all data needs to be moved or copied

Resulting Data Architectures



Numerous ETLs

Unnecessary duplication, replication

Data governance issues accelerating

A new approach to Data Virtualization

Now in beta trial. Coming to ICP for Data in November

1 Query anything, anywhere.

Query **many heterogenous data sources as one** across cloud, on-premise and mobile with advanced analytics using the most popular languages and tools

2 Extreme simplicity.

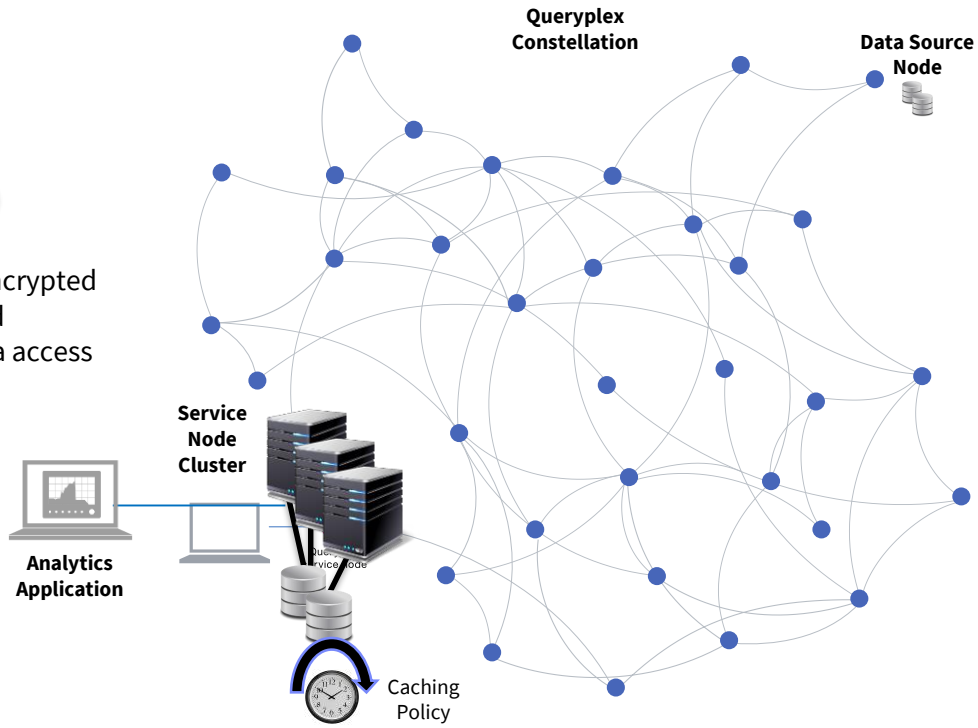
Automatically discover, and connect **few to many devices and data stores** into a single self balancing constellation. Avoid the complexity of centralized copies. Data only persists at the source.

3 Execution speedup.

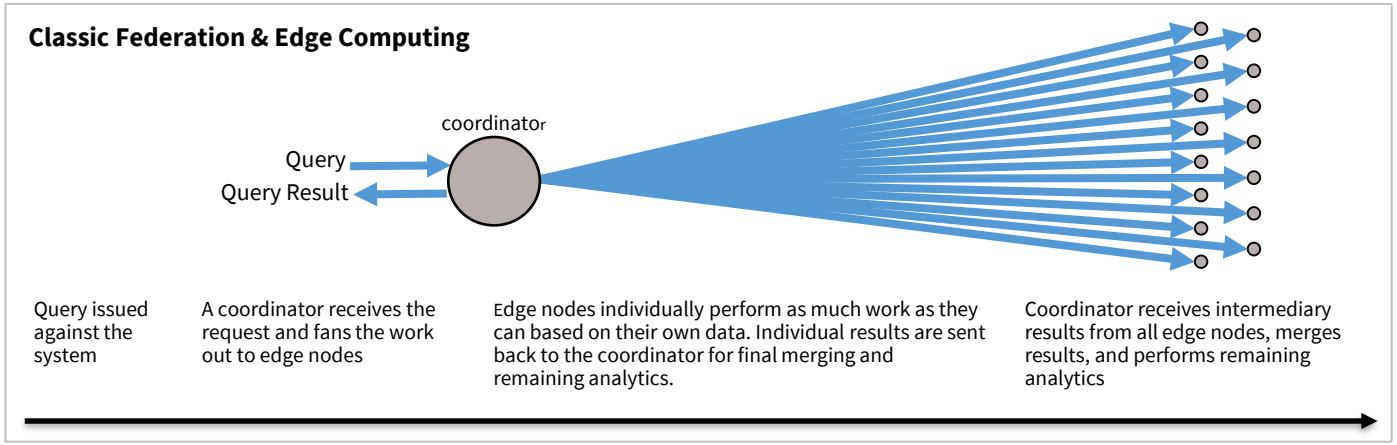
Many times acceleration using the power of every device to compute and aggregate results.

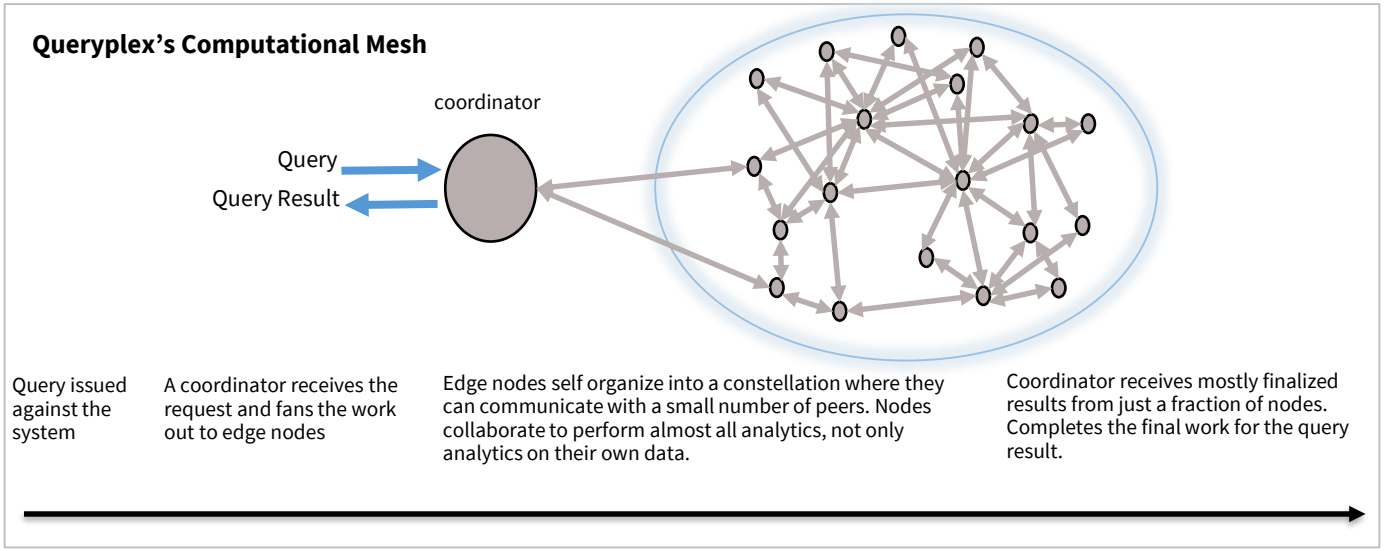
4 Security.

Fully secure and encrypted communication and preservation of data access rights at source.



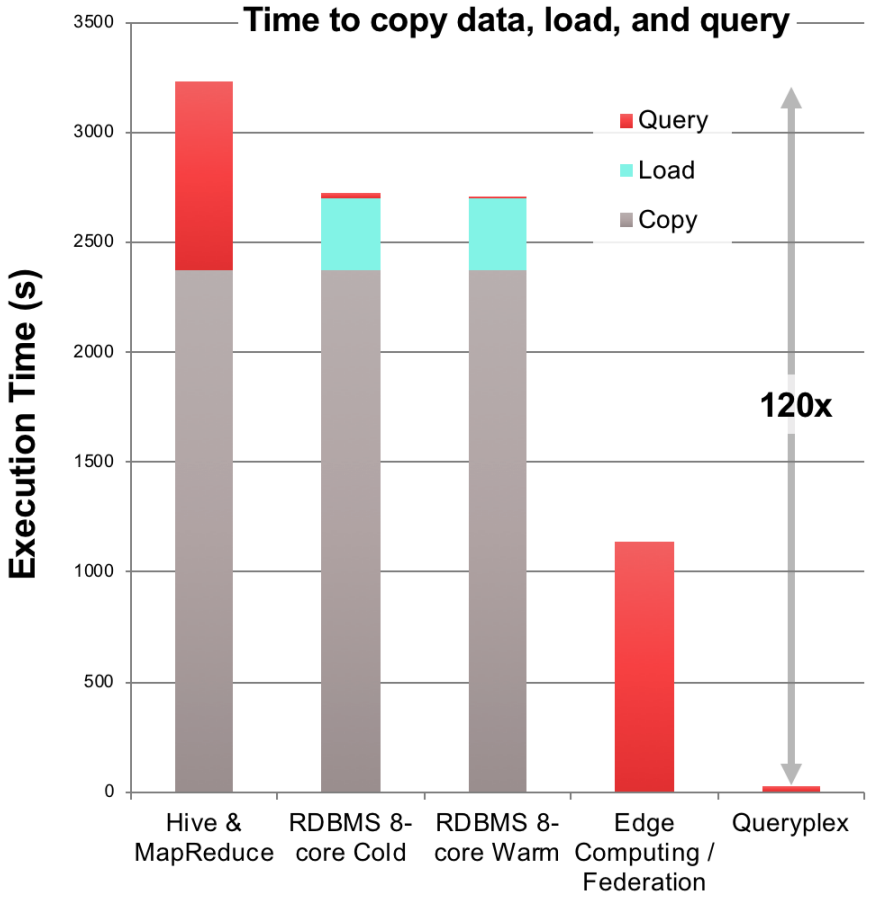
Federation vs. Queryplex Datenvirtualisierung





Early Performance Benchmark – Query with Copy & Load

- Query includes six aggregates and grouping on 2.5 years of data, with 2,500 data points per day.
  - San Jose: Head node
  - Toronto: Edge devices
- RDBMS and Hive tests run on Intel CPU, roughly 4x faster than ARM.
- Edge/Federation and Queryplex data resides on MySQL databases inside of 36 Raspberry Pi devices, with ARM processors.
  - San Jose: Head node
  - Toronto: Edge devices



Fragen?

European IBM Informix Days 2019 im Watson IoT Center in München am 3. und 4. Juni

Vorträge vom Labor zu Informix (Zeitreihendatenbank)
Führung IoT Center

Anmeldung unter:
https://ibm.biz/InformixDays2019

Notices and disclaimers

Notices and disclaimers continued

- 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, Aspera®, Bluemix, Blueworks Live, CICS, Clearcase, Cognos®, DOORS®, Emptoris®, Enterprise Document Management System™, FASP®, FileNet®, Global Business Services®,
Global Technology Services®, IBM ExperienceOne™, IBM SmartCloud®, IBM Social Business®, Information on Demand, ILOG, Maximo®, MQIntegrator®, MQSeries®, Netcool®, OMEGAMON, OpenPower, PureAnalytics™, PureApplication®, pureCluster™, PureCoverage®, PureData®, PureExperience®, PureFlex®, pureQuery®, pureScale®, PureSystems®, QRadar®, Rational®, Rhapsody®, Smarter Commerce®, SoDA, SPSS, Sterling Commerce®, StoredIQ, Tealeaf®, Tivoli® Trusteer®, Unica®, 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.