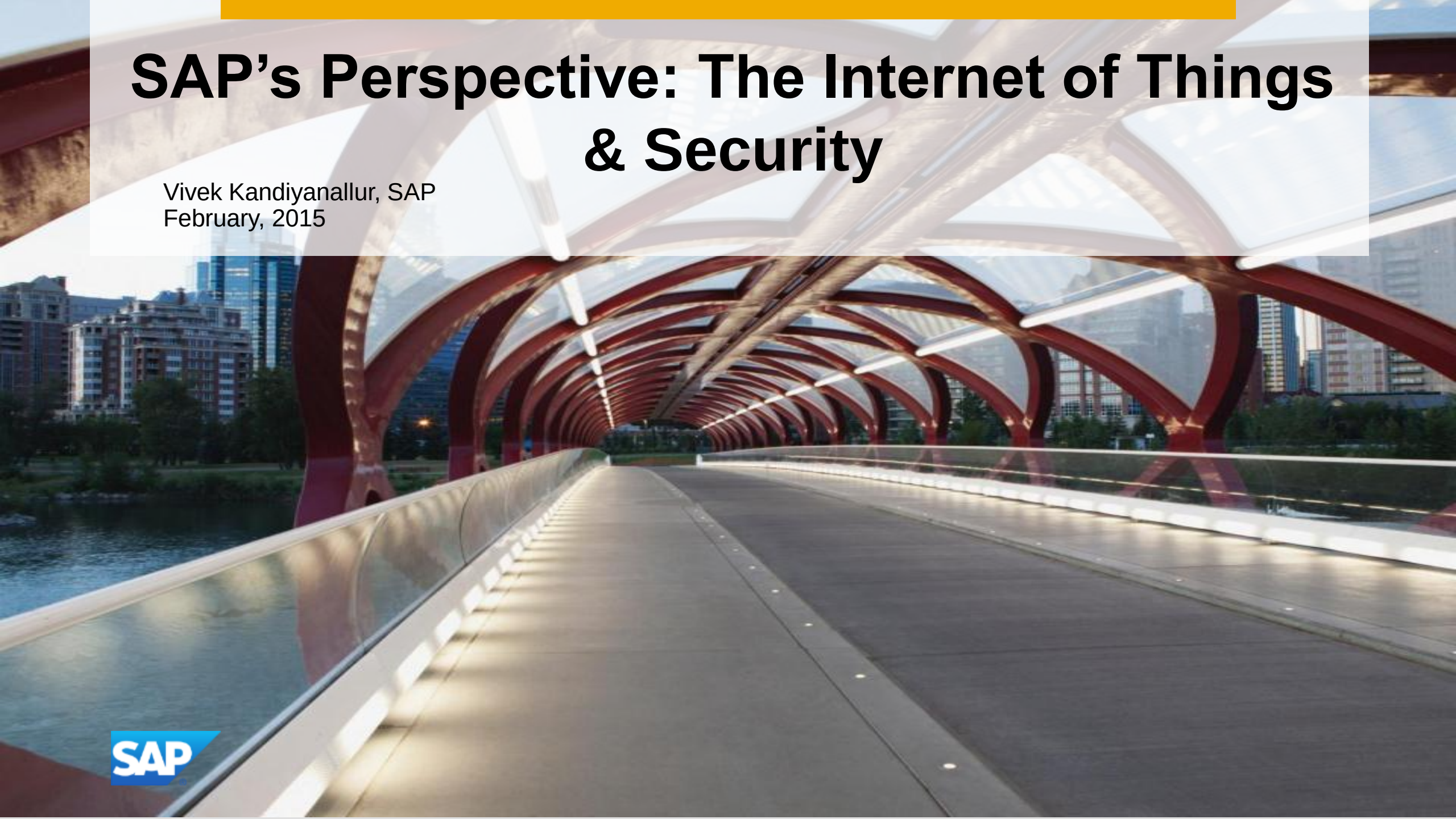
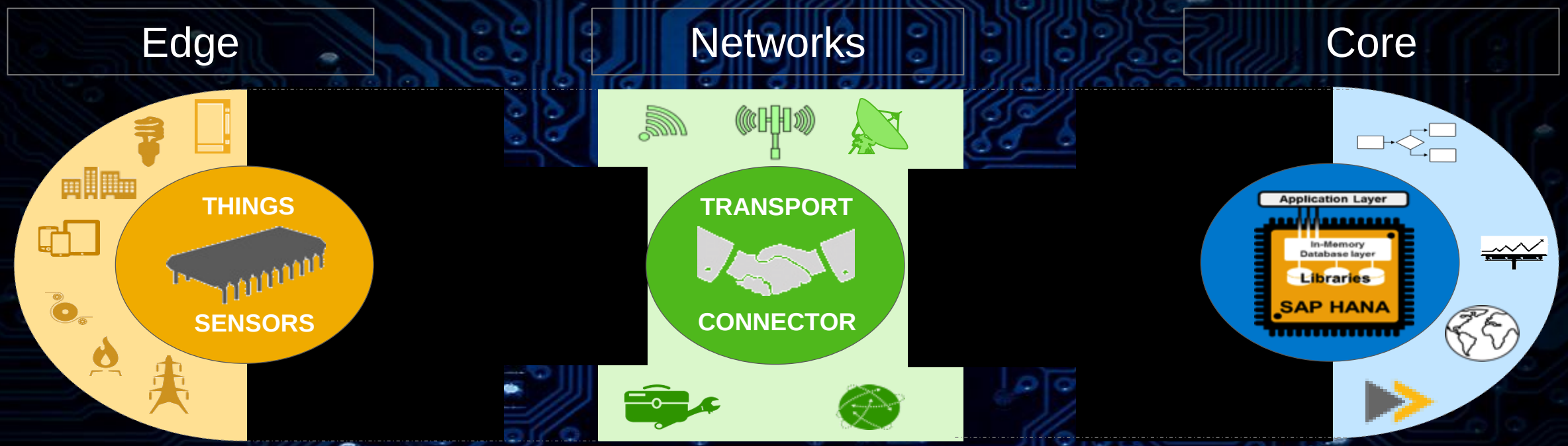
SAP's Perspective: The Internet of Things & Security

Vivek Kandiyanallur, SAP
February, 2015



Deconstructing the Internet of Things



10's of billions of connected things

Private and public networks

Big Data High-performance compute infrastructure

SAP Plant Connectivity

SAP SQL Anywhere | Ultralite

SAP Event Stream Processor (Edge)

SAP Manufacturing Intelligence & Infrastructure

SAP SQL Anywhere MobiLink

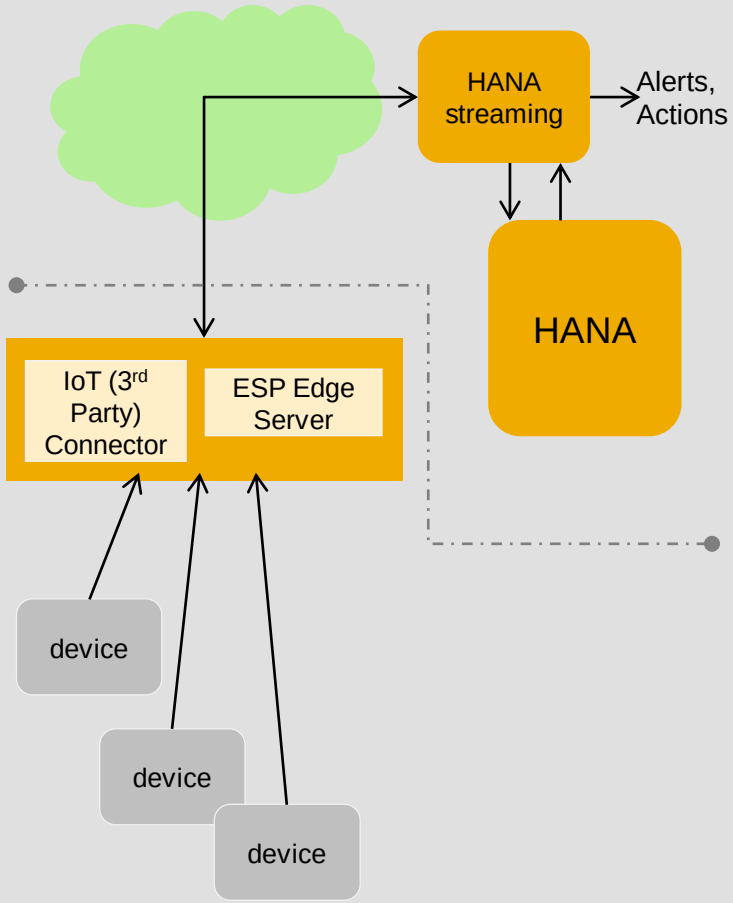
SAP Event Stream Processor | Smart Streaming

SAP HANA Platform

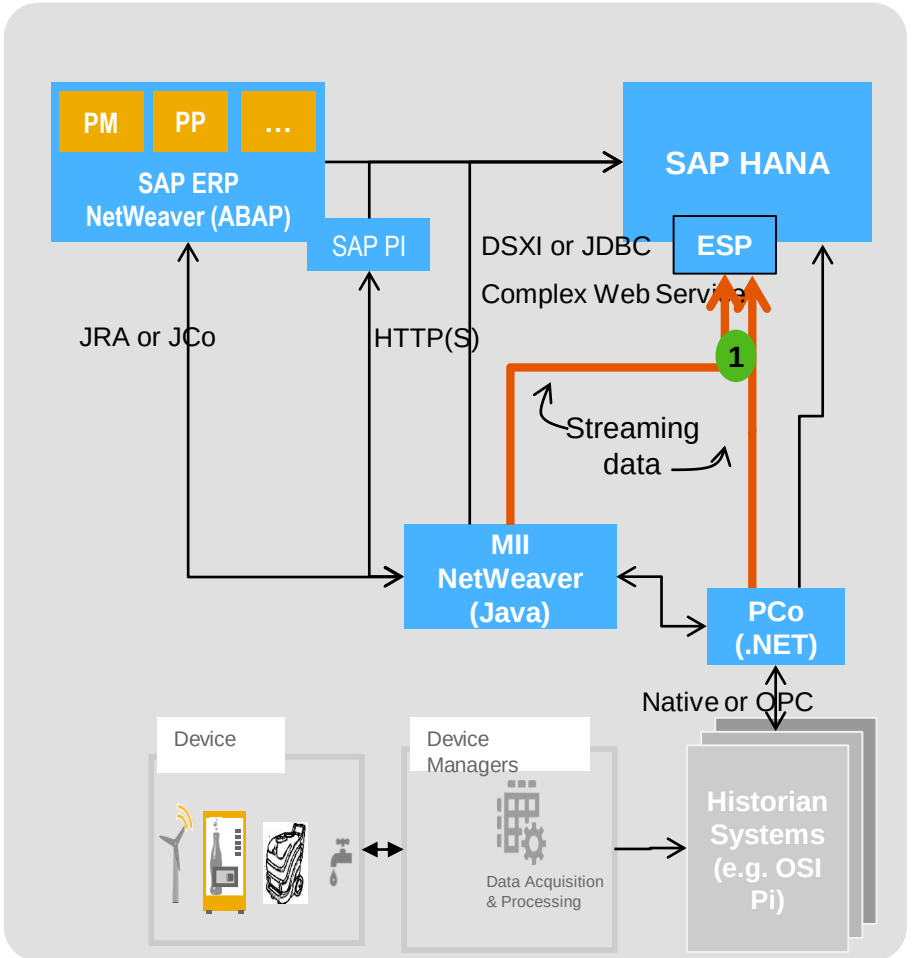
Partner Value Chain : Important contributors to delivering device data .

Illustrative Solution Recipes

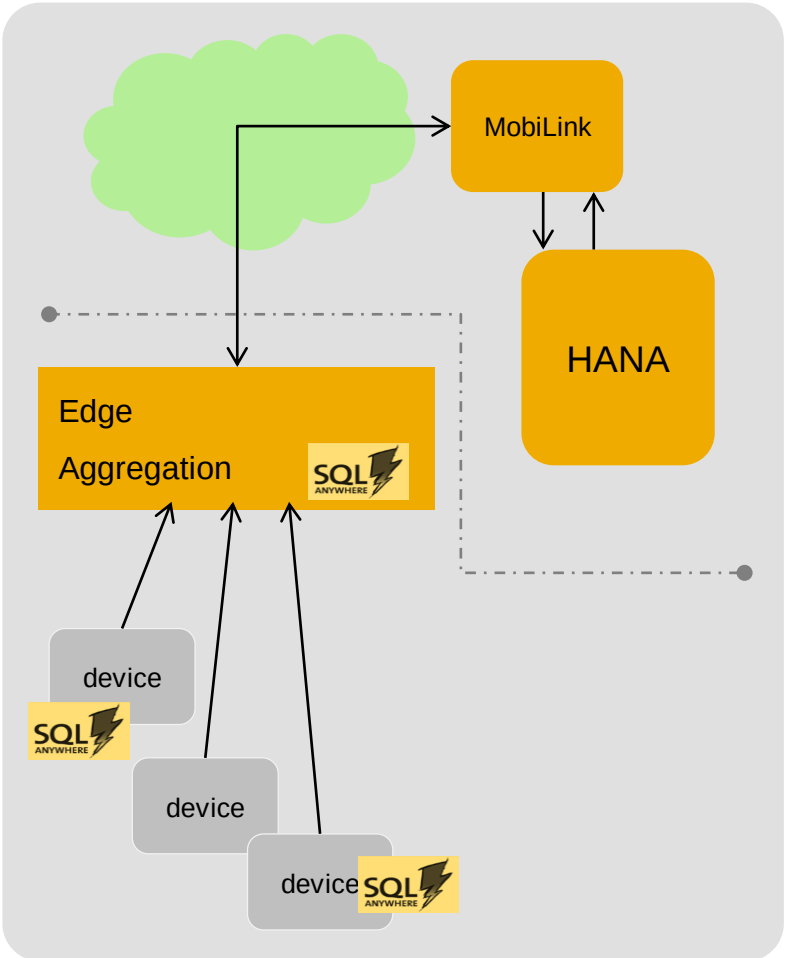
IoT Connector | Streaming



IoT Industrial



IoT Always Available



Agenda

HANA Platform

HANA Smart Streaming

SQL Anywhere

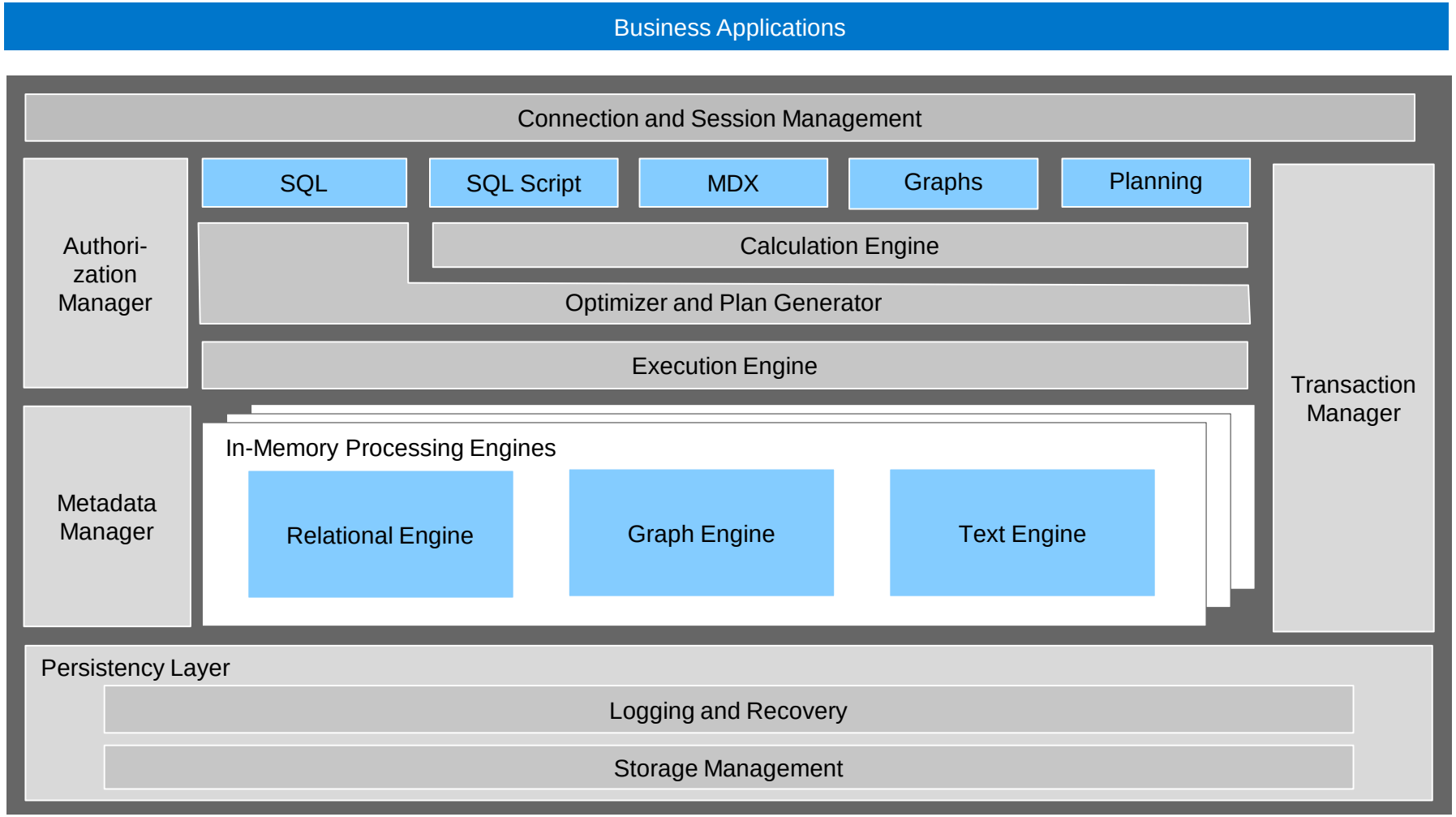
HANA Cloud Platform

IoT Solutions

IoT Security & Research

SAP HANA Database

Multi-Engine for Different Application Needs



SAP HANA Technology & Features
Combined in one DBMS Platform

In-memory DBMS
- Exploit SSD/disk for spilling, aging/archiving, durability/fault-tolerance

Standard RDBMS features
- SQL, stored procedures
- ACID, MVCC with snapshot isolation, logging and recovery

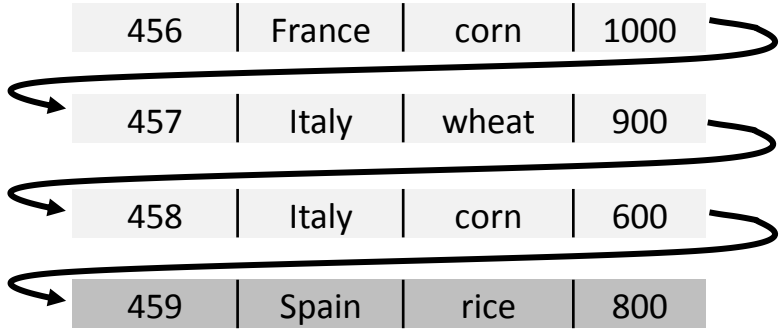
Focus on column store
- Late materialization and decompression
- Row store capability, e.g. for system catalogs

High Performance
- Efficient compression techniques
- Parallelization at multiple levels
- Scanning operations co-optimized with hardware

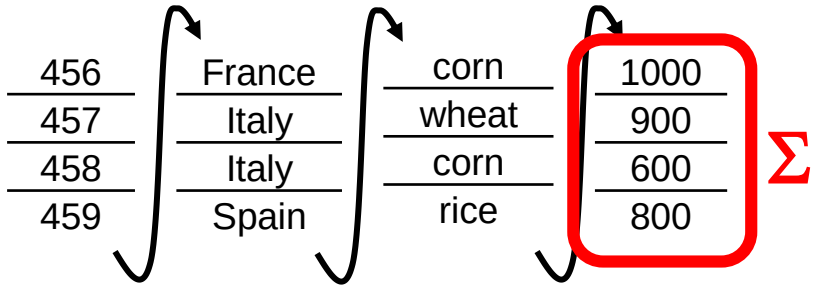
Reduced TCO and administration
- Avoid indexes, aggregates and materialized views, with exceptions (like primary key indexes)

In-Memory Computing – Data Structures

Order	Country	Product	Sales
456	France	corn	1000
457	Italy	wheat	900
458	Italy	corn	600
459	Spain	rice	800



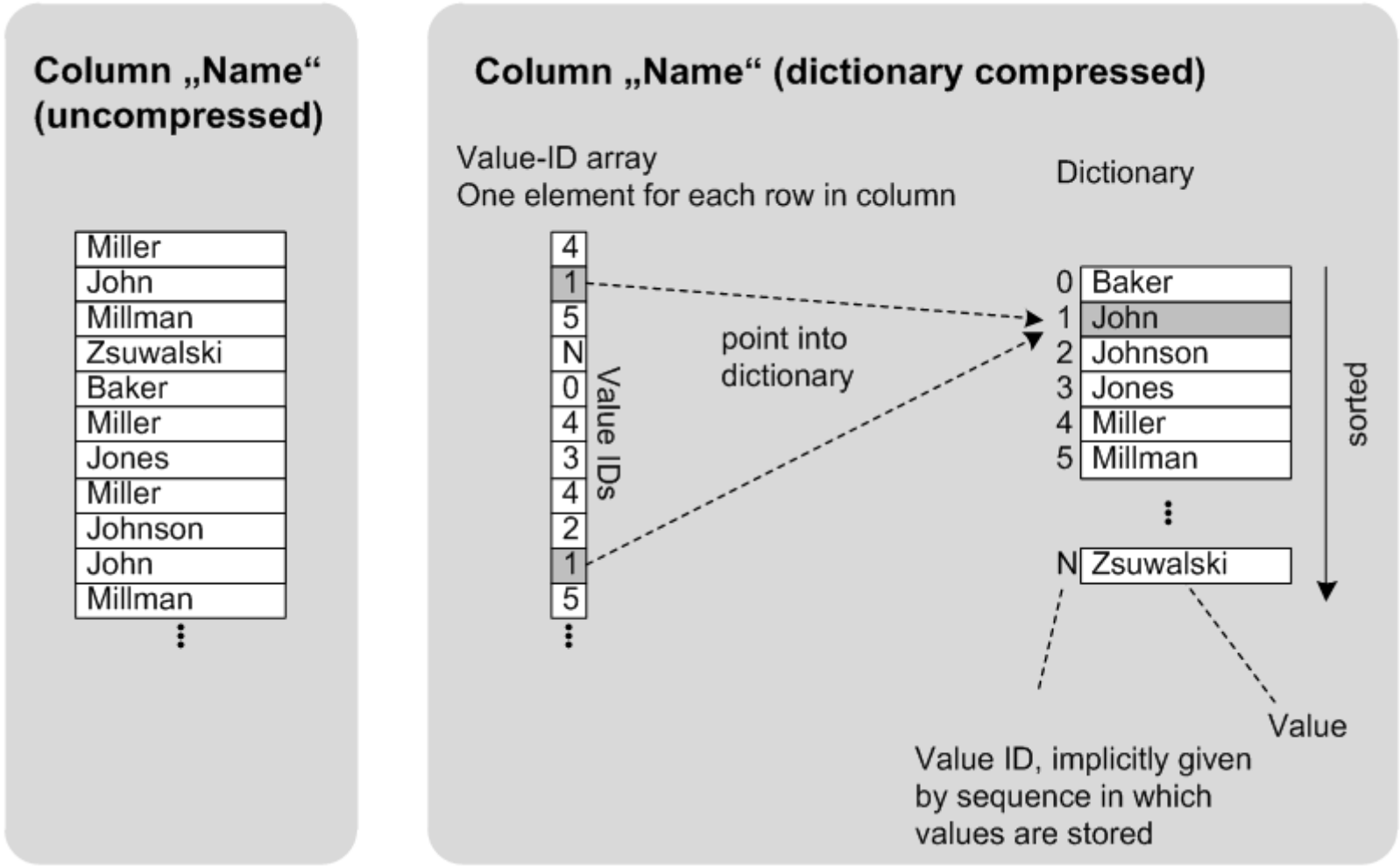
Typical Database



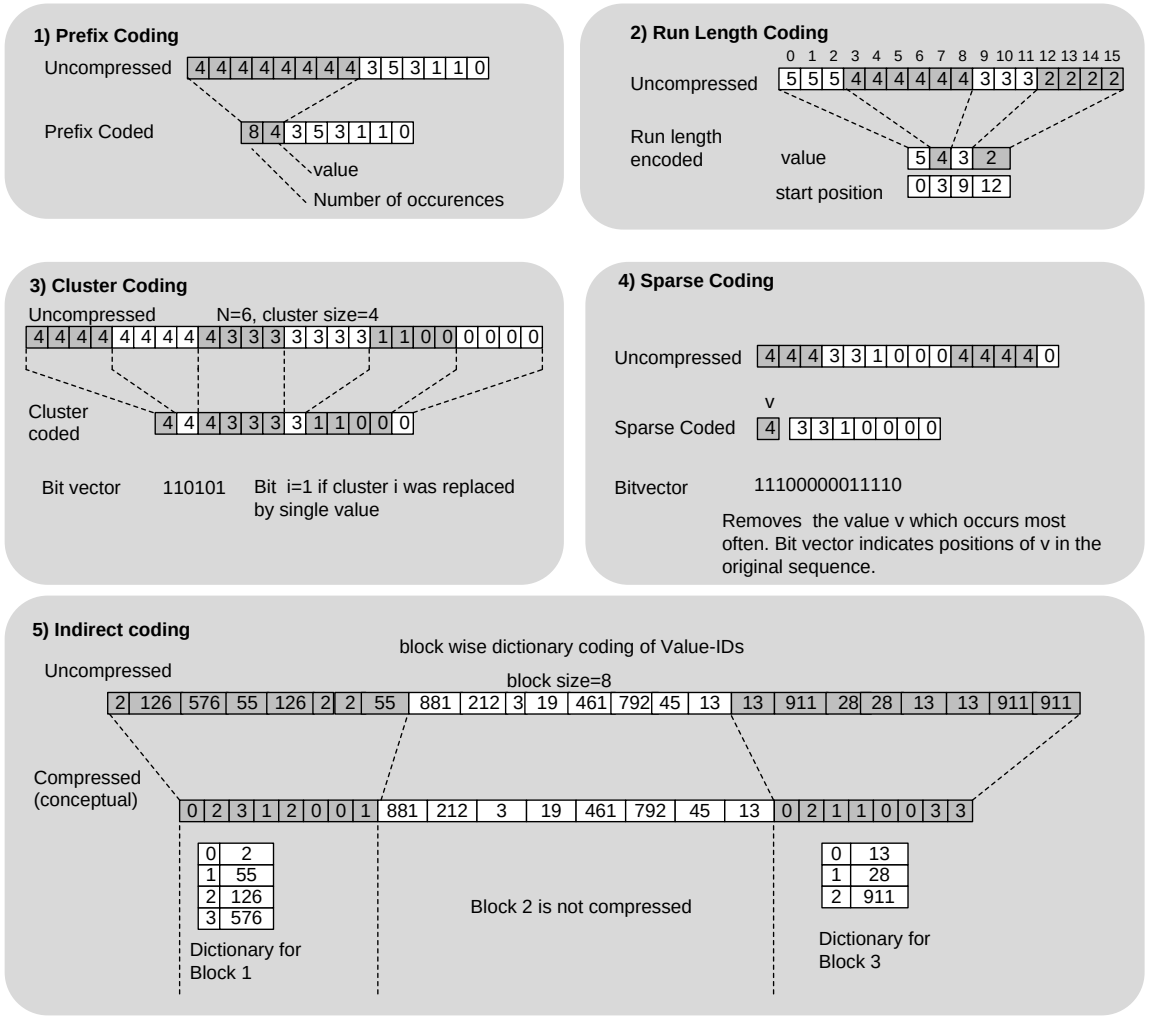
SAP HANA: column order

```
SELECT  Country, SUM(sales) FROM SalesOrders
WHERE Product = 'corn'
GROUP  BY  Country
```

SAP HANA: Dictionary Compression



Additional Compression Technologies



SAP HANA Column Store

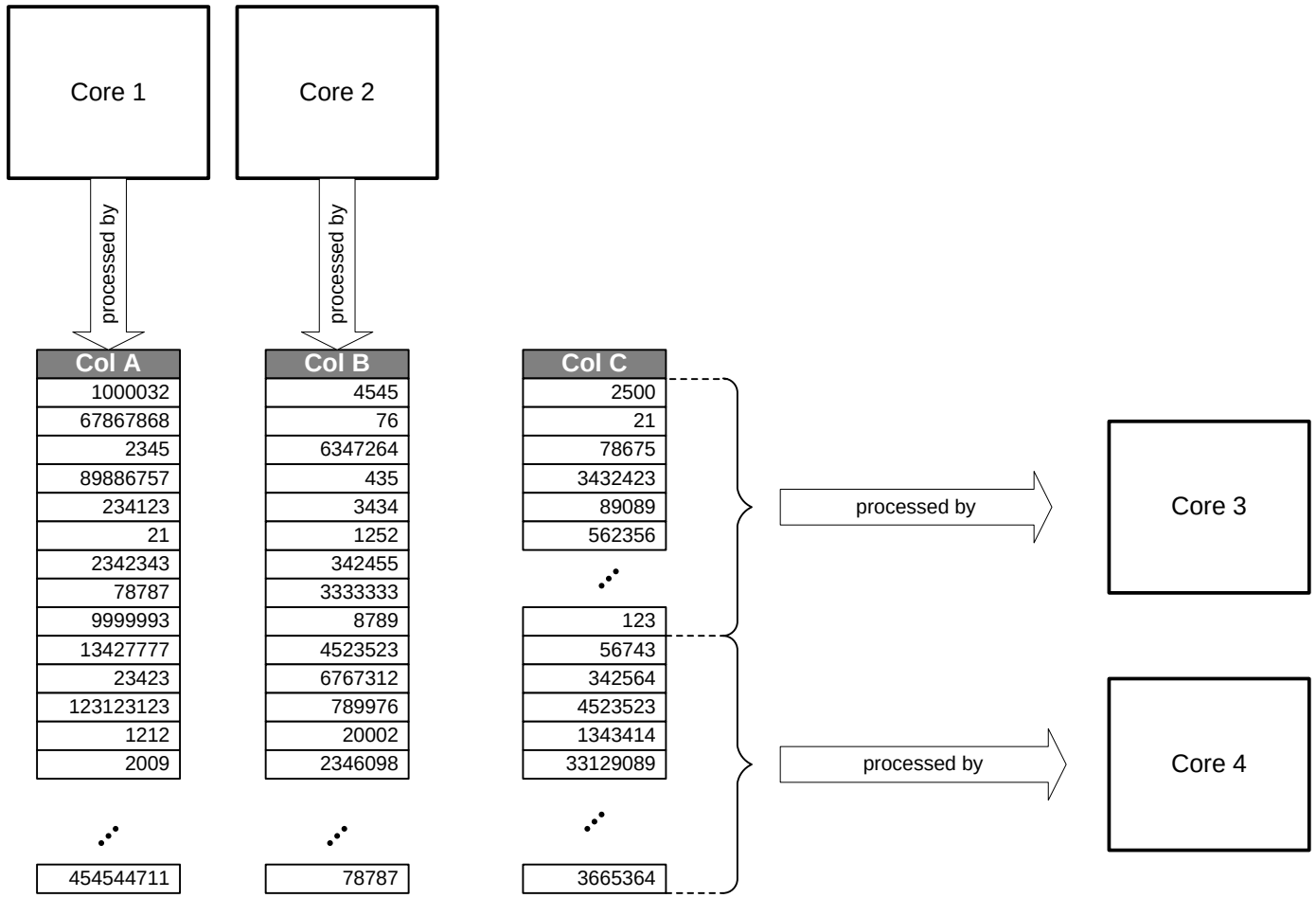
Column Main: Read-optimized store for immutable data

- High data compression
- Efficient compression methods (dictionary and run-length, cluster, prefix, etc.)
  - Dictionary values for main are sorted in same order as data
- Heuristic algorithm orders data to maximize secondary compression of columns
- Compression works well, speeding up operations on columns (~ factor 10)

Column Delta: Write-optimized store for inserts, updates and deletes

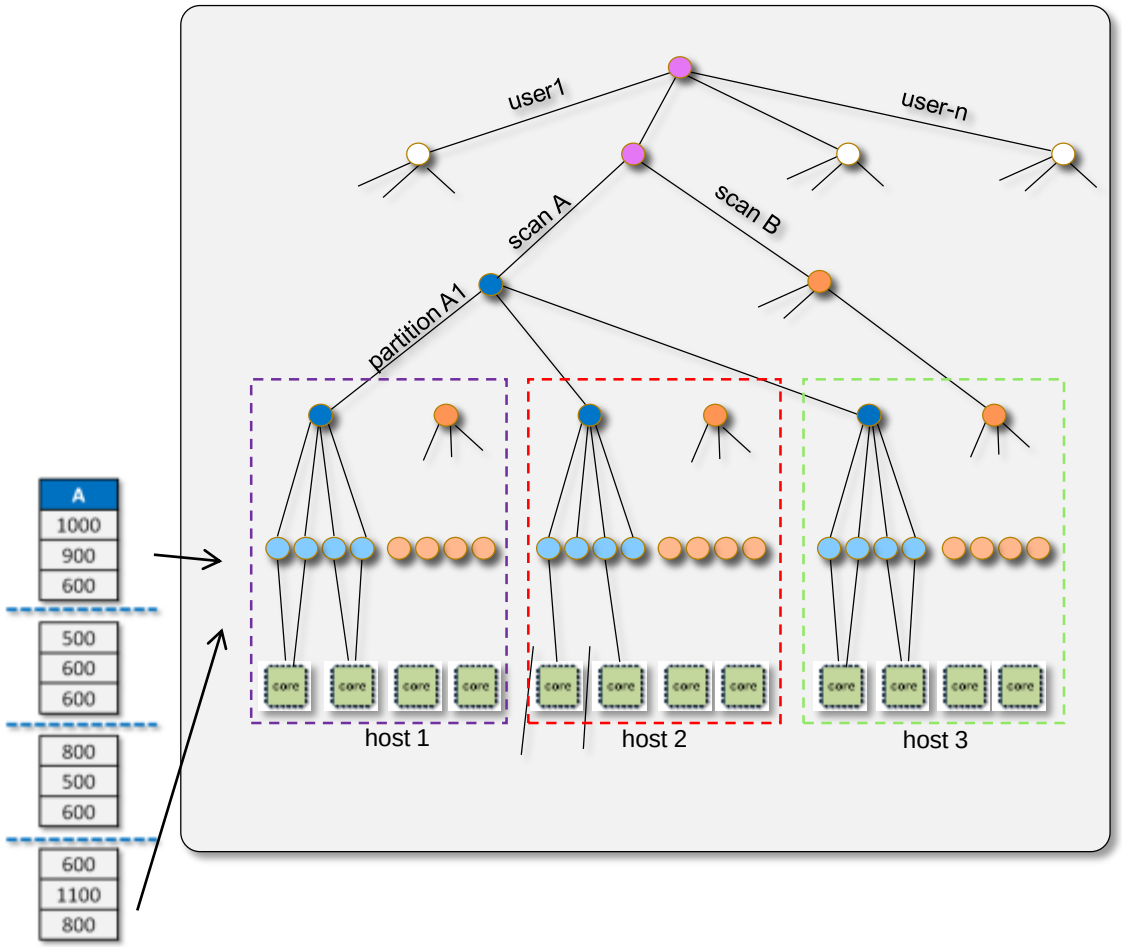
- Less compression of data
- Data is appended to delta to optimize write performance
- Unsorted dictionary on delta helps speed write performance
- Delta is merged with main periodically, or when thresholds are exceeded
  - Delta merge for a table partition is done on-line, in background
  - Enables highly efficient scan of Main again

SAP HANA: Multi-Core Parallelization



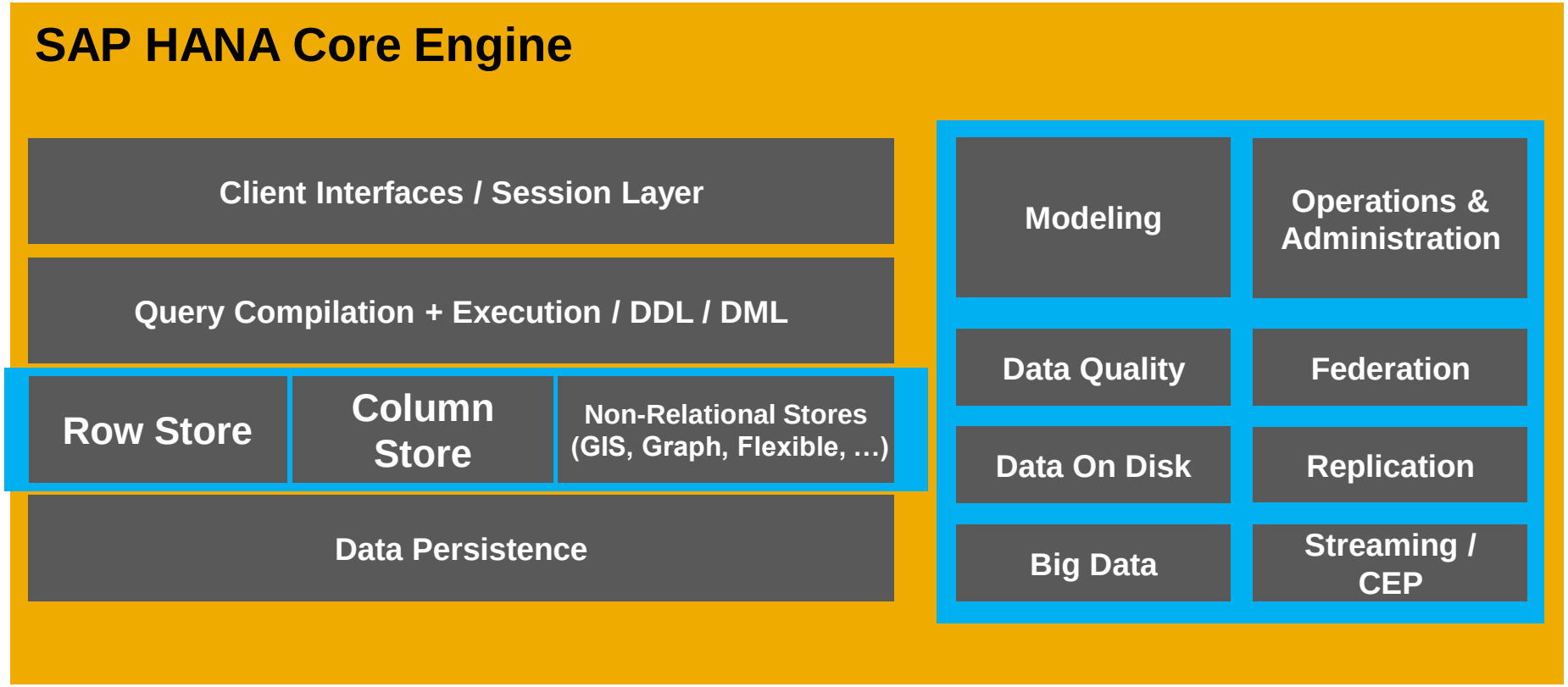
Parallelization at All Levels

- Multiple user sessions
- Concurrent operations within a query (… T1.A … T2.B…)
- Data partitioning on one or more hosts
- Horizontal segmentation, concurrent aggregation
- Multi-threading at Intel processor core level
- Vector processing



HANA Core Platform

ONE platform for simple and efficient data processing



Agenda

HANA Platform

HANA Smart Data Streaming

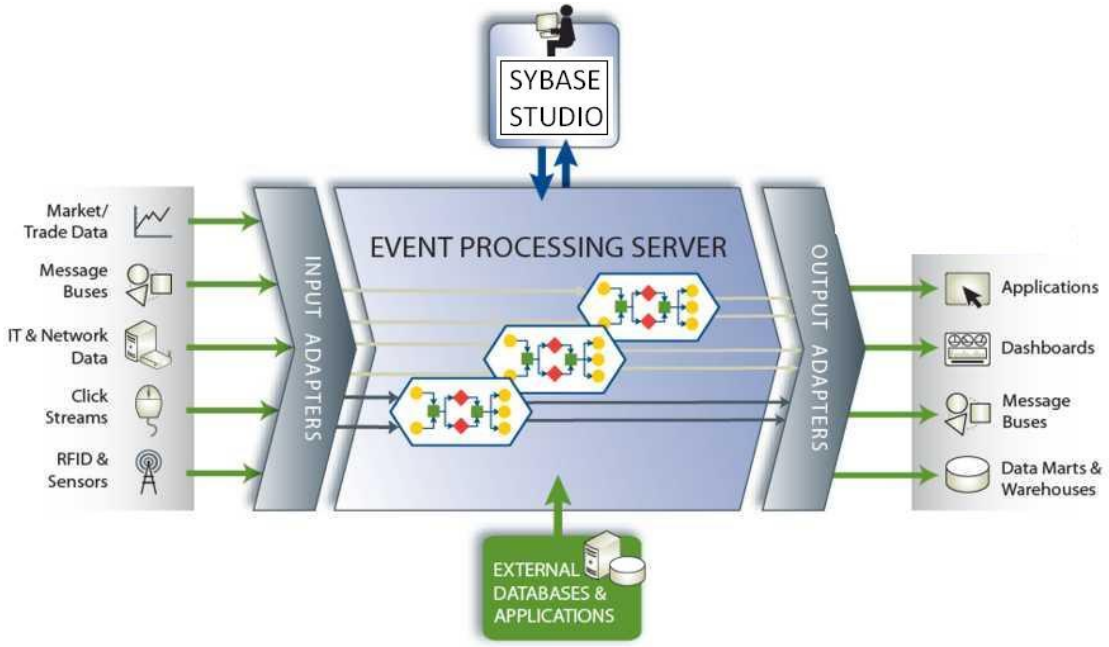
SQL Anywhere

HANA Cloud Platform

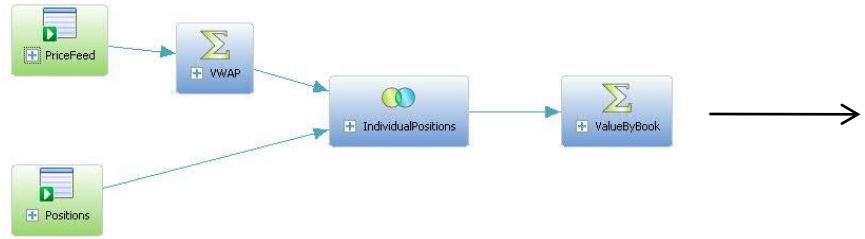
IoT Solutions

IoT Security & Research

Event stream processing uses continuous queries



Continuous Queries



Step 1:

Define the continuous queries and the dataflow

Step 2:

Wait for data to arrive. As it arrives, it flows through the continuous queries to produce immediate results

Smart data streaming extends the capabilities of the SAP HANA Platform

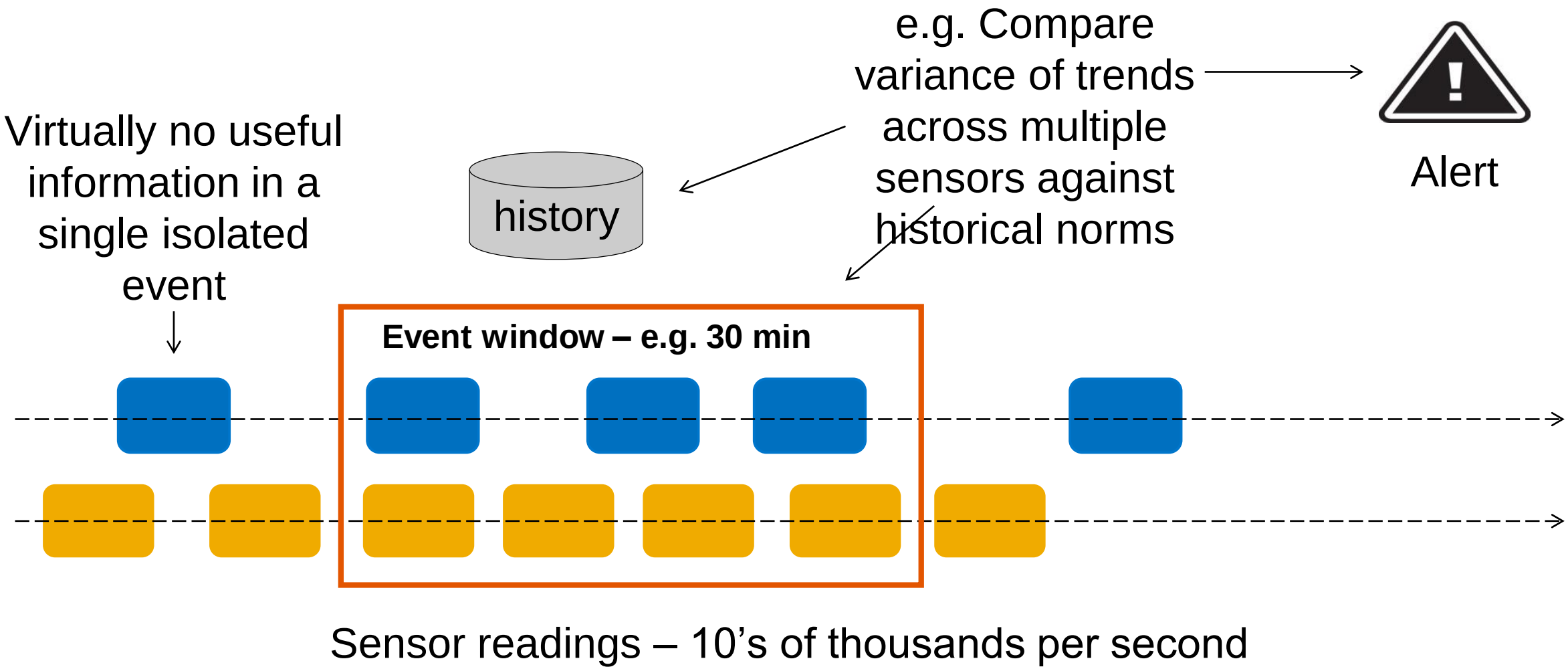
Stream capture

- **Capture data arriving as individual events – at potentially high speeds**
  - Hundreds of thousands or millions of events per second
  - Micro-batching and parallel processing to optimize load speeds
- **Capture events that are published from streaming sources**
  - e.g. message bus
- **Filter, transform or enrich the data on the way in**
  - Capture only the data you want, in the form you need it
- **Prioritize data**
  - Capture high value data in HANA and direct other data into Hadoop

Continuous analysis, Immediate Response

- **Monitor incoming event streams**
  - Watch for trends or patterns
  - Monitor correlations
  - Detect missing events
  - Continuously update and monitor aggregate statistics
- **Generate alerts, notifications**
- **Initiate immediate response**

Complex Event Processing extracts insight from events



CCL: Continuous Computation Language – Some examples

Input Stream:

```
CREATE INPUT STREAM EventsIn

SCHEMA (

ID INTEGER ,

Value integer ,

TS msdate )
```

Filter:

```
CREATE OUTPUT STREAM Filter1

AS SELECT *

FROM EventsIn

WHERE EventsIn.Value > 28
;
```

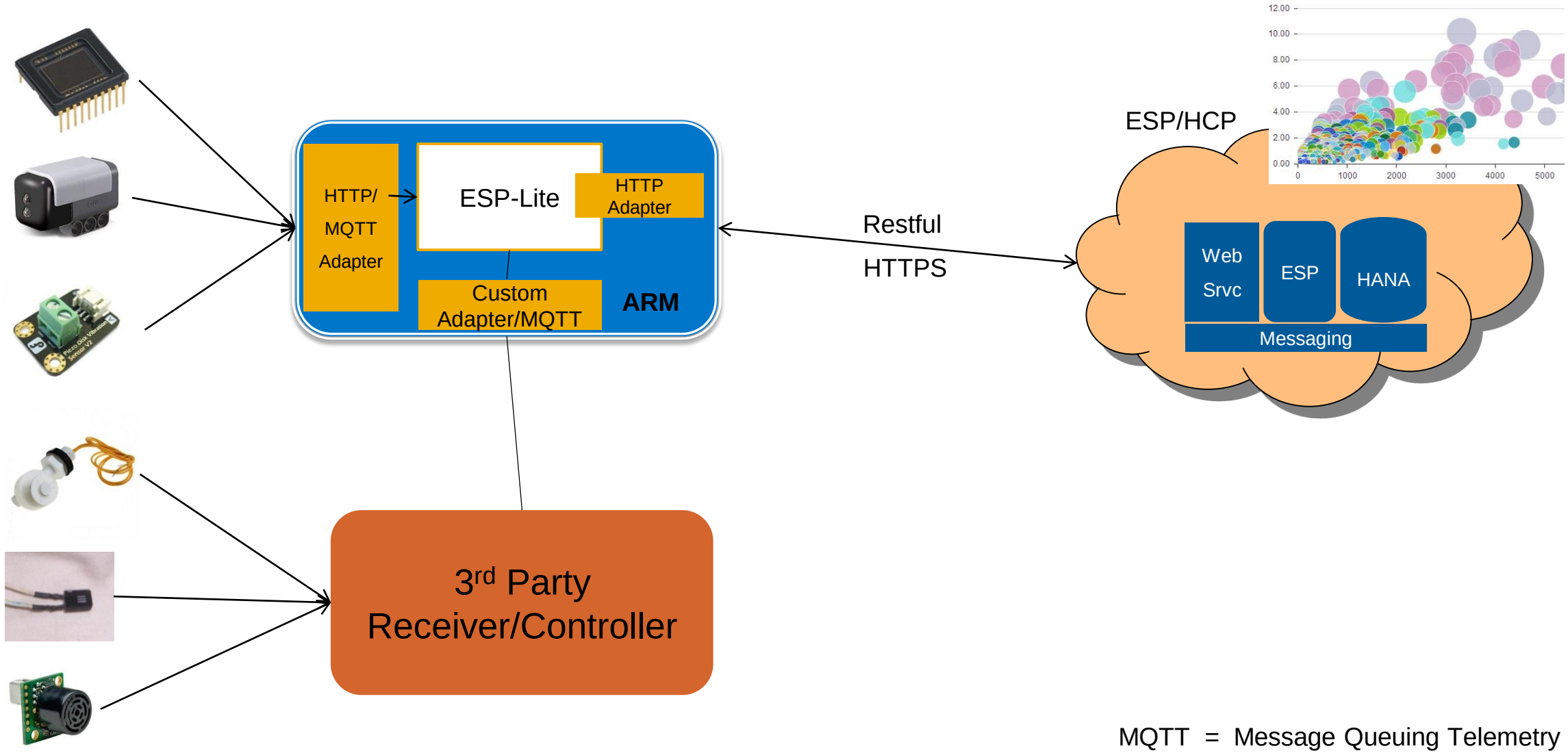
Aggregate:

```
CREATE OUTPUT WINDOW MovAvg

PRIMARY KEY DEDUCED

AS SELECT EventsIn

.ID ID ,

avg ( EventsIn.Value ) Value ,

EventsIn.TS TS

FROM EventsIn KEEP 3 MIN

GROUP BY EventsIn.ID ;
```

Join:

```
CREATE OUTPUT STREAM Join1

AS SELECT

Dev2.ID ID ,

Dev2.Temp Temp ,

MovAvg.Value Value ,

MovAvg.TS TS

FROM Dev2 INNER JOIN MovAvg

ON Dev2.ID = MovAvg.ID ;
```

IoT – ESP Lite on the Edge & Scalable Cloud Platform



MQTT = Message Queuing Telemetry Transport

Agenda

HANA Platform

HANA Smart Streaming

SQL Anywhere

HANA Cloud Platform

IoT Solutions

IoT Security & Research

SQL Anywhere and MobiLink Overview

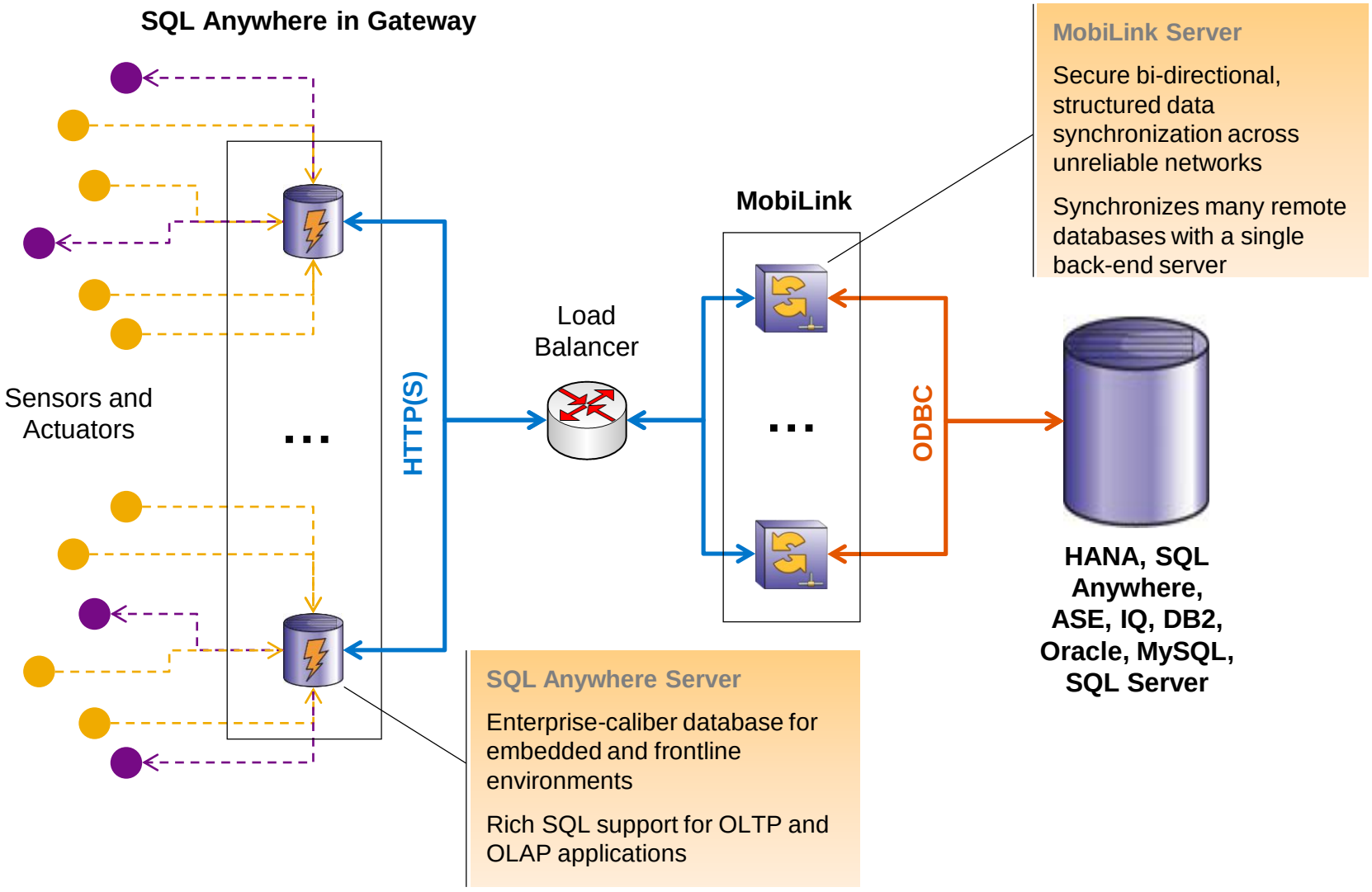
SQL Anywhere

- Highly embeddable (~20 MB) enterprise caliber database with rich feature set (complex SQL, spatial, full text search)
- Broad platform support including Linux ARM (e.g. Raspberry Pi) and x86 (Intel Edison)

MobiLink

- Session based, bi-directional synchronization between multiple remote SQL Anywhere databases and a back-end database
- Can optionally be horizontally scaled behind load balancer to support large installations (> 100,000 remote databases)
- Support for slow, unreliable and intermittently connected networks

SQL Anywhere and MobiLink Architecture



SQL Anywhere and MobiLink are Complementary

MobiLink is *not* real-time, it is on-demand session-based synchronization (e.g. when connection becomes available)

- Synchronization frequency is determined by remote application and can range from several times an hours, to once a month or less

SQL Anywhere and MobiLink are useful for environments that require:

- Complex data processing and analytics capabilities at the edge/gateway
- Durable storage of complex data for eventual transmission to core/cloud when connection is available
- Secure, bi-directional synchronization of structured including sending metadata/reference data from core to gateway
- An enterprise caliber database for applications/control logic running on the gateway

SQL Anywhere provides a complementary data plane to other real-time data collection technologies (e.g. streaming)

Agenda

HANA Platform

HANA Smart Streaming

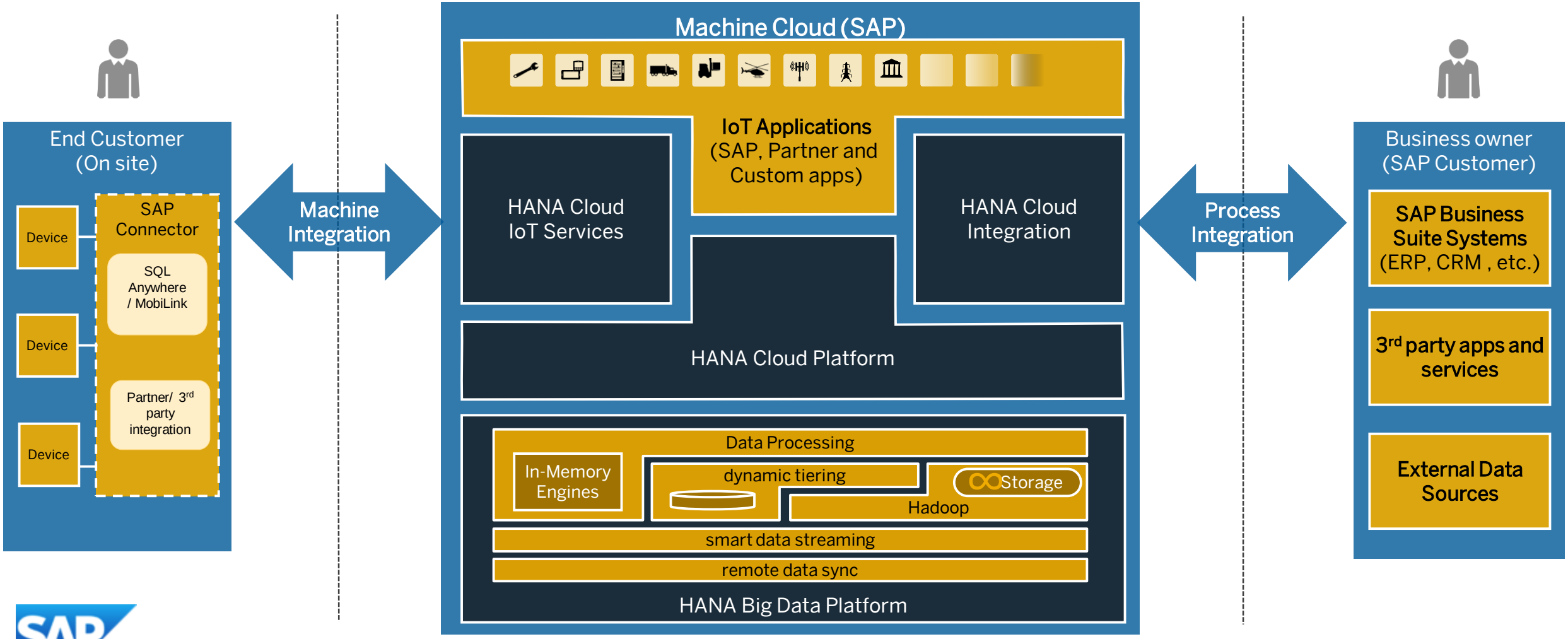
SQL Anywhere

HANA Cloud Platform

IoT Solutions

IoT Security & Research

SAP IoT Offering



Agenda

HANA Platform

HANA Smart Streaming

SQL Anywhere

HANA Cloud Platform

IoT Solutions

IoT Security & Research

Enabling the Internet of Things with SAP solutions

IoT Predictive Maintenance & Service

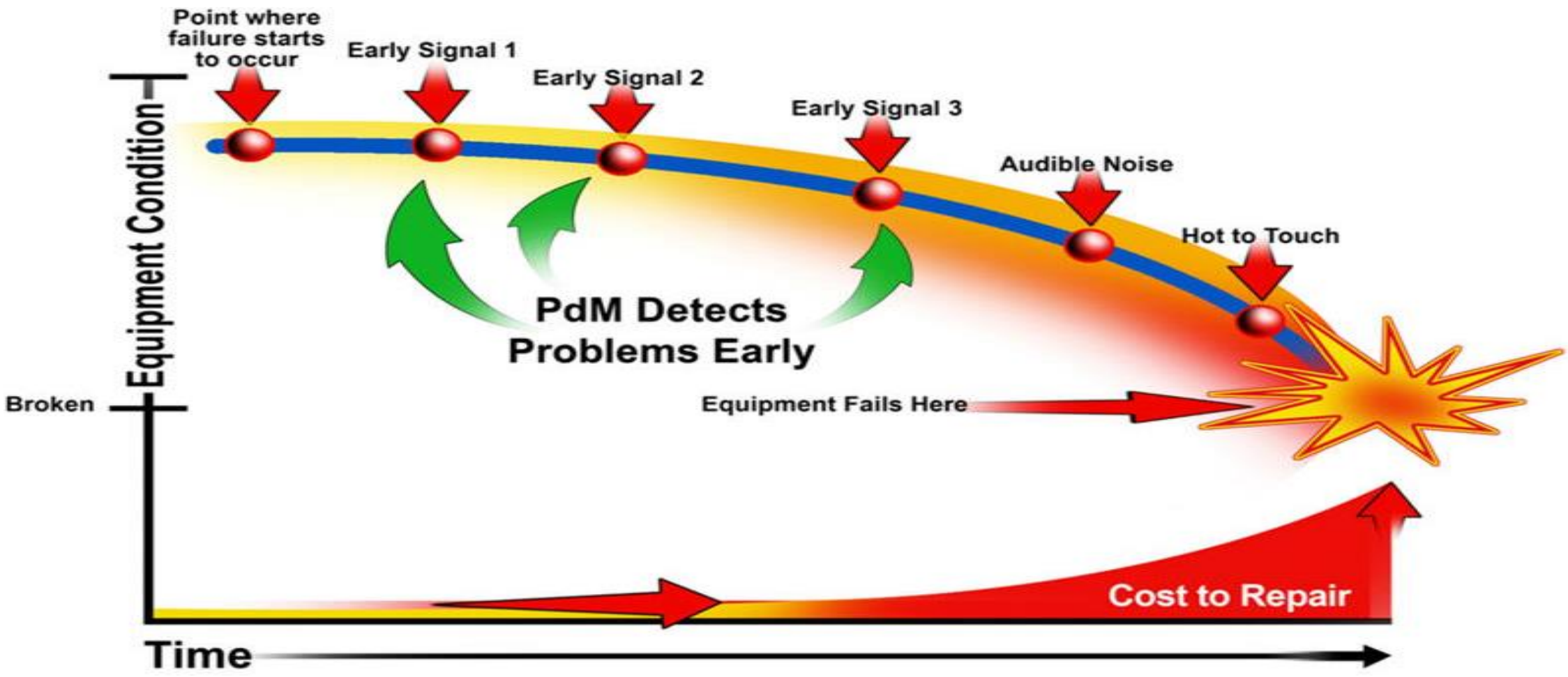
IoT Connected Manufacturing

IoT Connected Logistics

IoT Augmented Reality

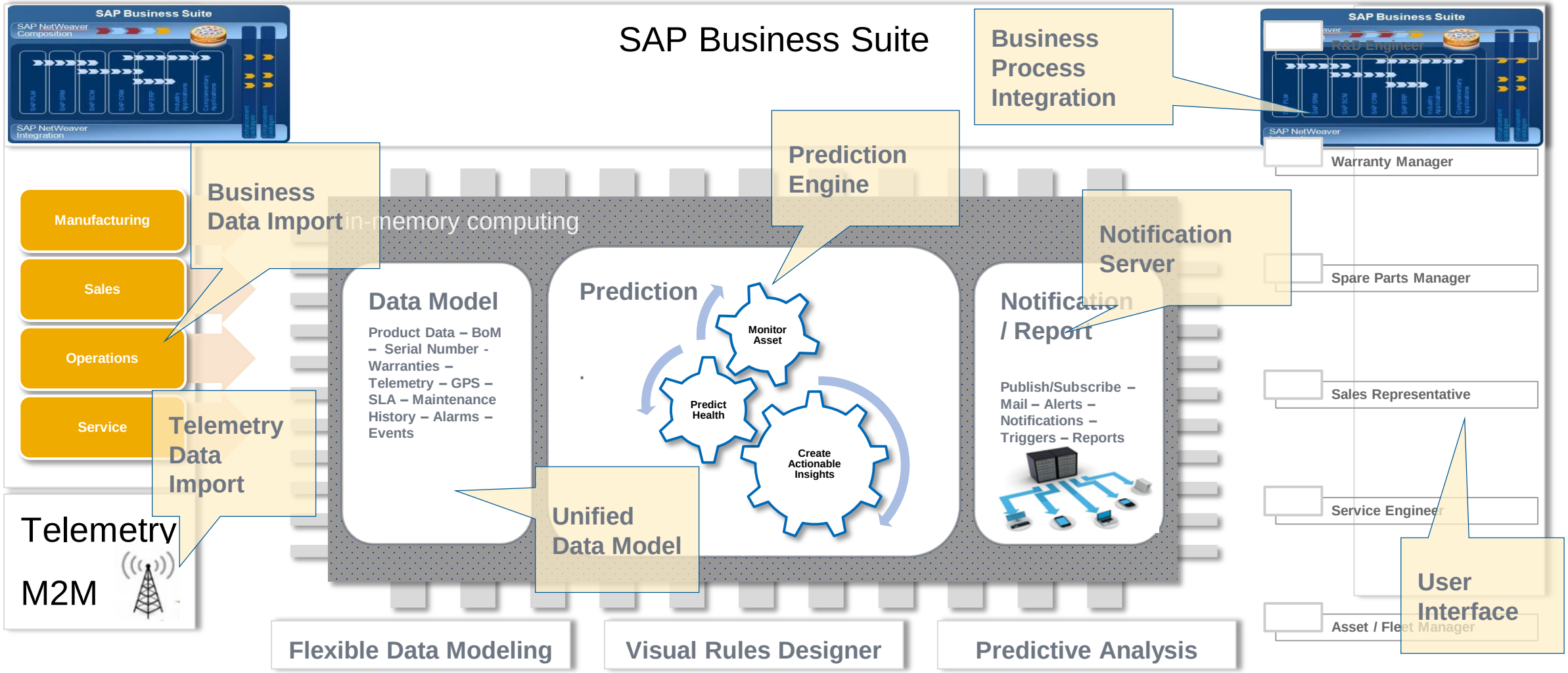
IoT Innovation Bundle

Predictive Maintenance



Graph from http://reliabilitycenteredenergymanagement.com/wp-content/uploads/2011/10/P-F-Curve.jpg

Building Blocks



SAP Predictive Maintenance and Service Process

Increase effectiveness

Effectiveness is the capability of producing a desired result

Predictive Maintenance and Service

IT / OT Connectivity → Condition Monitoring Remote Service → Fault Pattern Recognition → Machine Health Prediction

Create Maintenance or Service Order → Schedule Order → Execute Order on mobile device → Visual Support

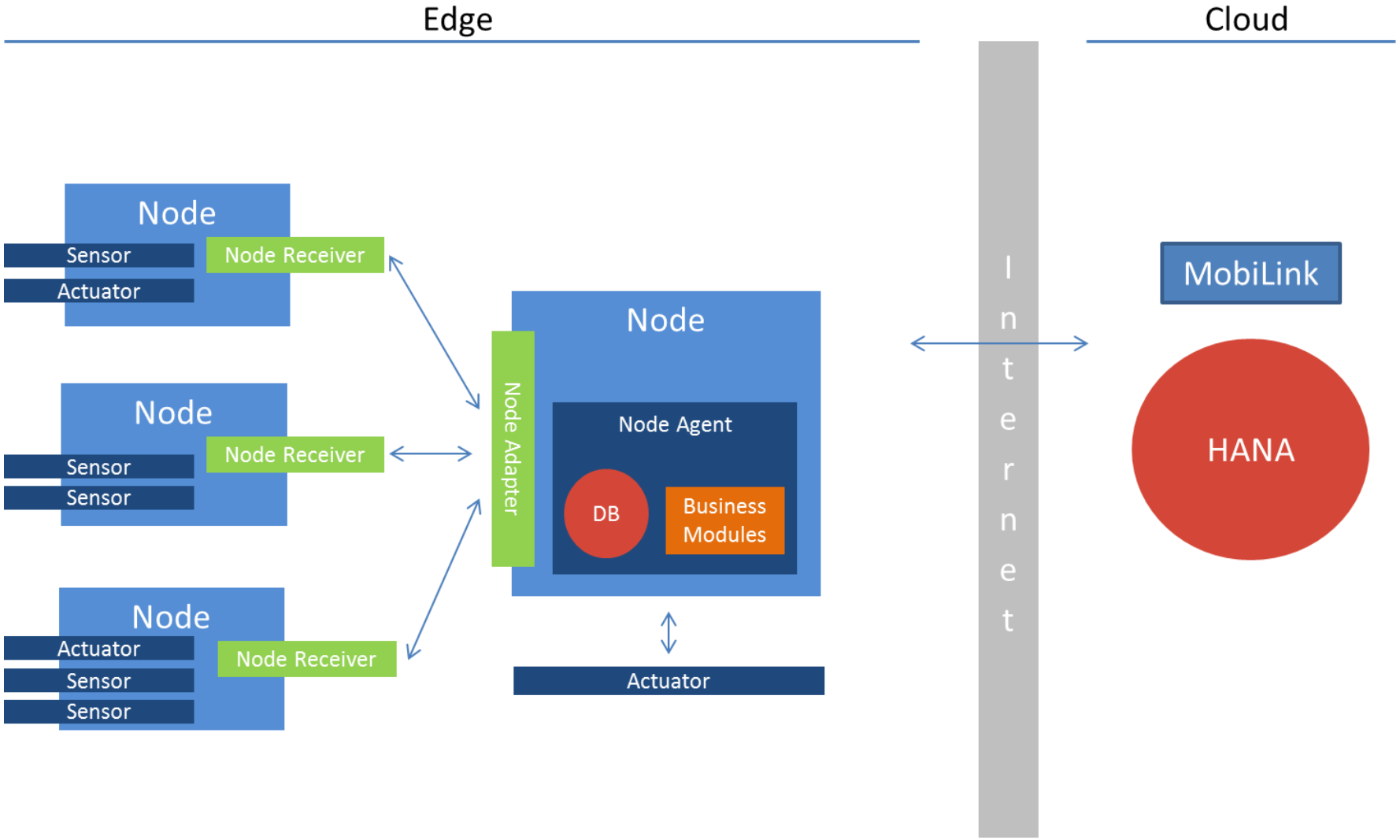
Increase efficiency

Time, effort or cost is well used for the intended task or purpose

SAP ERP

Non-SAP Apps

Smart Green Roof



Agenda

HANA Platform

HANA Smart Streaming

SQL Anywhere

HANA Cloud Platform

IoT Solutions

IoT Security & Research

Sensor devices and Wireless networks

- CPU, RAM, battery restriction
- No direct application of traditional processing or security approaches
- Unreliability of Sensor Devices
- Unreliability of Wireless Sensor Networks
- Large diversity of sensors
- Lack of inter operability with business applications

Past and present security research projects

- **Context aware security policies – Extension to XACML for service discovery in ubiquitous networks**
- **Secure handshake – A protocol that allows user's to mutually verify another's property without revealing their identity.**
- **Trust assessment of sensor data**
- **Privacy preserving for asset tracking in supply chain**
- **Secure alerting in supply chain**
- **Secure exchange of RFID tracking data**
- **Privacy in cyber physical systems**
- **Multi-tenancy of sensors used in an office building**
- **Predictive analytics for pipeline integrity**

Foreseen security challenges – Predictive analytics for Pipleline Integrity

Need for sensor data anonymisation

End to end security

Efficient aggregation on encrypted data

Full or partial homomorphic encryption for sensor devices

Efficient secure alerting on encrypted data

Order Preserving Encryption for sensor devices

Efficient and scalable security solution for big sensor network (million of devices)

secure event stream processing

Deal with untrusted gateways

Key management of encryption key

Secure storage of cryptographic material on sensor

Key distribution on sensor

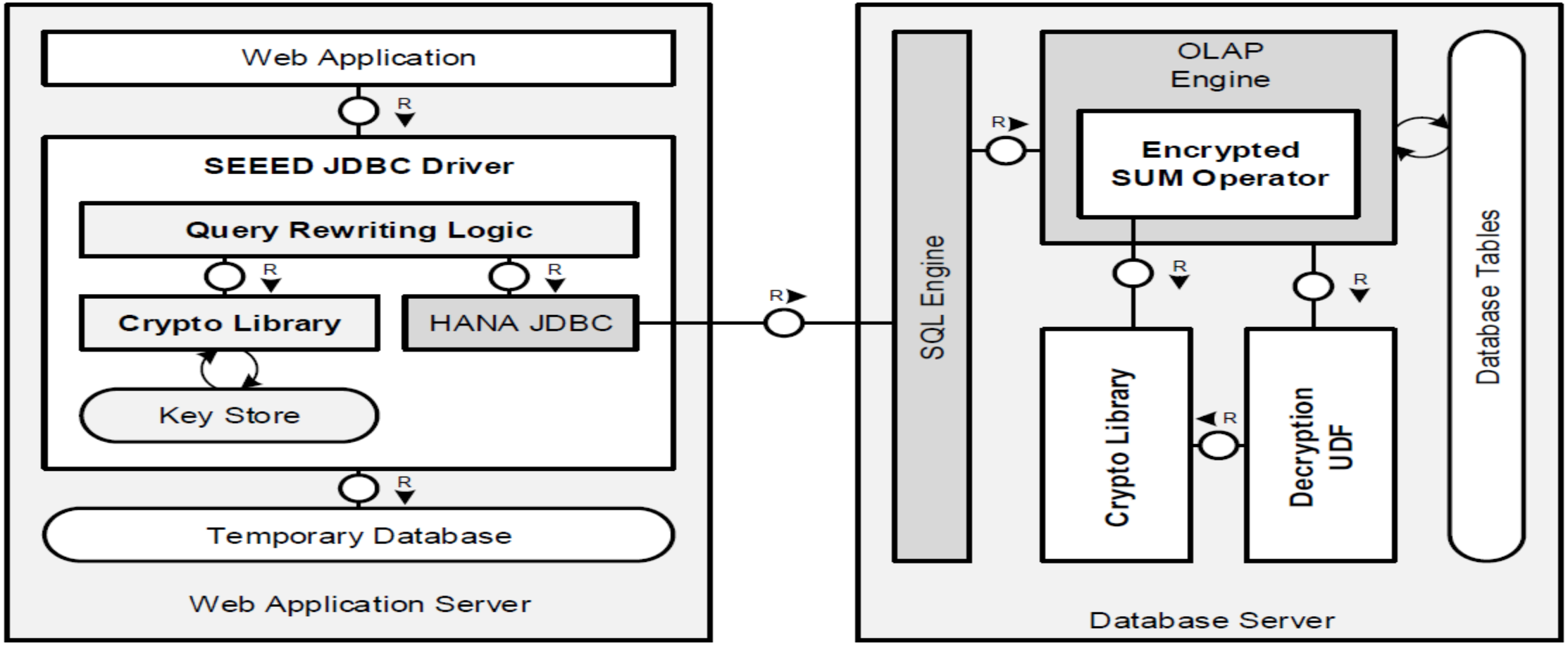
Key revocation

Sensor device authentication and identity management

Infrastructure, Cloud, Applications – Past and present security projects

- **Privacy-Preserving Benchmarking in the Cloud**
- **Secure collaborative supply chain management**
- **Resilient reputation systems**
- **Searching over encrypted data**

Searchable over encrypted data (SEEED)



Performance measurement

Test Case		SEEED	Plain	Impact
Server-Side Only	Exact Search	2.0	1.7	1.2 x
	Equi-Join	49.7	33.3	1.5 x
	Grouping with Aggregation (Sum)	674.1	57.8	11.7 x
Incl. Client-Side	Order by Aggregate (Sum)	870.1	56.3	15.4 x
TPCH	Q4	2,402	235	10.2 x
	Q5	1,373	207	6.6 x

Challenges

- **Improving the performance of aggregation queries**
- **Reducing the number of columns with lower encryption schemes**
- **Reduce the time taken to initially encrypt the database**
- **Re-encrypting the database due to lost/stolen master keys without taking the database offline**

SAP HANA Collaborative Research

Research overview: http://scn.sap.com/docs/DOC-27051

- Publications: http://scn.sap.com/docs/DOC-26787
- Academic partners: http://scn.sap.com/docs/DOC-26786
- Students and alumni: http://scn.sap.com/docs/DOC-26824

University collaborators at PhD level include:

- TU Dresden
- University of Mannheim
- TU München
- ETH Zürich
- EPFL
- HPI
- DHBW Mannheim
- TU Ilmenau
- TU Karlsruhe
- University of Heidelberg
- University of Toronto
- University of Waterloo
- More including conversations with others in-progress

vivek.kandiyanallur@sap.com

anil.goel@sap.com

@thedataneer