SAP’s Perspective: The Internet of Things & Security

Vivek Kandiyananallur, SAP
February, 2015
Deconstructing the Internet of Things

**Edge**
- THINGS
- SENSORS

10’s of billions of connected things

**Networks**
- TRANSPORT
- CONNECTOR

Private and public networks

**Core**
- Big Data High-performance compute infrastructure

**Partner Value Chain**: Important contributors to delivering device data.

- 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
Illustrative Solution Recipes

**IoT Connector | Streaming**

- HANA streaming
- ESP Edge Server
- Alerts, Actions
- IoT (3rd Party) Connector
- Device

**IoT Industrial**

- PM
- PP
- ... NetWeaver (ABAP)
- SAP PI
- DSXI or JDBC
- Complex Web Service
- JRA or JCo
- HTTP(S)
- Streaming data
- ESP
- MII NetWeaver (Java)
- PCo (.NET)
- Native or OPC
- Historian Systems (e.g. OSI PI)
- Device Managers
- Data Acquisition & Processing

**IoT Always Available**

- MobiLink
- Edge Aggregation
- Device
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

<table>
<thead>
<tr>
<th>Order</th>
<th>Country</th>
<th>Product</th>
<th>Sales</th>
</tr>
</thead>
<tbody>
<tr>
<td>456</td>
<td>France</td>
<td>corn</td>
<td>1000</td>
</tr>
<tr>
<td>457</td>
<td>Italy</td>
<td>wheat</td>
<td>900</td>
</tr>
<tr>
<td>458</td>
<td>Italy</td>
<td>corn</td>
<td>600</td>
</tr>
<tr>
<td>459</td>
<td>Spain</td>
<td>rice</td>
<td>800</td>
</tr>
</tbody>
</table>

Typical Database

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

SAP HANA: column order
SAP HANA: Dictionary Compression

Column „Name“ (uncompressed)

<table>
<thead>
<tr>
<th>Value IDs</th>
<th>Miller</th>
<th>John</th>
<th>Millman</th>
<th>Zsuwalski</th>
<th>Baker</th>
<th>Miller</th>
<th>Jones</th>
<th>Johnson</th>
<th>John</th>
<th>Millman</th>
</tr>
</thead>
<tbody>
<tr>
<td>Value</td>
<td>4</td>
<td>1</td>
<td>5</td>
<td>N</td>
<td>0</td>
<td>1</td>
<td>3</td>
<td>4</td>
<td>4</td>
<td>2</td>
</tr>
</tbody>
</table>

Column „Name“ (dictionary compressed)

Value-ID array
One element for each row in column

Dictionary

<table>
<thead>
<tr>
<th>Value</th>
<th>0</th>
<th>Baker</th>
<th>1</th>
<th>John</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>2</td>
<td>Johnson</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>Jones</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>Miller</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>Millman</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Value ID, implicitly given by sequence in which values are stored

Value

sorted

point into dictionary

© 2014 SAP AG or an SAP affiliate company. All rights reserved.
## Additional Compression Technologies

### 1) Prefix Coding
- **Uncompressed:**
  - Value
  - Number of occurrences
- **Prefix Coded:**
  - Value

### 2) Run Length Coding
- **Uncompressed:**
  - Run length encoded
  - Value start position
- **Compressed:**
  - Run length encoded
  - Value start position

### 3) Cluster Coding
- **Uncompressed:**
  - N=6, cluster size=4
- **Cluster coded:**
  - Bit vector 110101
  - Bit \(i\) if cluster \(i\) was replaced by single value
- **Compressed (conceptual):**
  - Block wise dictionary coding of Value IDs

### 4) Sparse Coding
- **Uncompressed:**
  - Value
- **Sparse Coded:**
  - Value
- **Bit vector:** 11100000011110

### 5) Indirect Coding
- **Block wise dictionary coding of Value IDs:**
  - Compressed (conceptual): 0 2 1 55 2 126 3 576
  - Dictionary for Block 1
  - Block 2 is not compressed
  - Dictionary for Block 3
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

Core A
- 1000032
- 67867868
- 2345
- 8986757
- 234123
- 2342343
- 78787
- 9999993
- 13427777
- 123123123
- 1212
- 2009
- ...
- 454544711

Core B
- 4545
- 6347264
- 3434
- 3333333
- ...
- 78787

Core C
- 2500
- 78675
- 3432423
- 89089
- 562356
- ...
- 3665364

Core 3
processed by Core 1
Core 4
processed by Core 2
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

Virtually no useful information in a single isolated event

Event window – e.g. 30 min

Sensor readings – 10’s of thousands per second

e.g. Compare variance of trends across multiple sensors against historical norms

Alert
### CCL: Continuous Computation Language – Some examples

<table>
<thead>
<tr>
<th>Input Stream:</th>
<th>Filter:</th>
<th>Aggregate:</th>
<th>Join:</th>
</tr>
</thead>
<tbody>
<tr>
<td>CREATE INPUT STREAM EventsIn</td>
<td>CREATE OUTPUT STREAM Filter1</td>
<td>CREATE OUTPUT WINDOW MovAvg</td>
<td>CREATE OUTPUT STREAM Join1</td>
</tr>
<tr>
<td>SCHEMA (ID INTEGER, Value integer, TS msdate)</td>
<td>AS SELECT * FROM EventsIn WHERE EventsIn.Value &gt; 28;</td>
<td>PRIMARY KEY DEDUCED AS SELECT EventsIn .ID ID, avg (EventsIn.Value) Value, EventsIn.TS TS FROM EventsIn KEEP 3 MIN GROUP BY EventsIn.ID;</td>
<td>AS SELECT Dev2.ID ID, Dev2.Temp Temp, MovAvg.Value Value, MovAvg.TS TS FROM Dev2 INNER JOIN MovAvg ON Dev2.ID = MovAvg.ID;</td>
</tr>
</tbody>
</table>
IoT – ESP Lite on the Edge & Scalable Cloud Platform

3rd Party Receiver/Controller

ESP/Lite

HTTP Adapter

HTTP/ MQTT Adapter

Custom Adapter/MQTT

ARM

ESP/HCP

Restful HTTPS

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 in Gateway**

- Sensors and Actuators

**MobiLink**

- Load Balancer
- ODBC

**MobiLink Server**

- Secure bi-directional, structured data synchronization across unreliable networks
- Synchronizes many remote databases with a single back-end server

**SQL Anywhere Server**

- Enterprise-caliber database for embedded and frontline environments
- Rich SQL support for OLTP and OLAP applications

**HANA, SQL Anywhere, ASE, IQ, DB2, Oracle, MySQL, SQL Server**
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

Machine Cloud (SAP)

- IoT Applications (SAP, Partner and Custom apps)
- HANA Cloud IoT Services
- HANA Cloud Integration
- HANA Cloud Platform

Machine Integration

- Device
- SAP Connector
- SQL Anywhere / MobiLink
- Partner/3rd party integration

Process Integration

- Business owner (SAP Customer)
- SAP Business Suite Systems (ERP, CRM, etc.)
- 3rd party apps and services
- External Data Sources

End Customer (On site)

- Device
- SAP Connector
- SQL Anywhere / MobiLink
- Partner/3rd party integration

Data Processing

- In-Memory Engines
- dynamic tiering
- Hadoop
- smart data streaming
- remote data sync

HANA Big Data Platform

© 2014 SAP AG or an SAP affiliate company. All rights reserved.
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 Business Suite

- Flexible Data Modeling
- Visual Rules Designer
- Predictive Analysis

Prediction Engine
- Monitor Asset
- Predict Health
- Create actionable insights

Unified Data Model

Data Model
- Product Data – BoM
- Serial Number – Warranties
- Telemetry – GPS – SLA – Maintenance History – Alarms – Events

Telemetry Data Import

Manufacturing
- Sales
- Operations
- Service

Telemetry M2M

Business Process Integration

Notification Server
- Publish/Subscribe
- Mail – Alerts – Notifications – Triggers – Reports

Notification / Report

User Interface
- Warranty Manager
- Spare Parts Manager
- Sales Representative
- Service Engineer
- Asset / Fleet Manager

Notification
- Server

Business Data Import
- in-memory computing

Data Import

SAP Business Suite

R&D Engineer
- Warranty Manager
- Spare Parts Manager
- Sales Representative
- Service Engineer
- Asset / Fleet Manager

Operator / Driver
- User Interface
SAP Predictive Maintenance and Service Process

Increase effectiveness
Effectiveness is the capability of producing a desired result.

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

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

Predictive Maintenance and Service

SAP ERP
Non-SAP Apps
Smart Green Roof

![Diagram of smart green roof system]

- **Node**: Contains sensors, actuators, and node receivers.
- **Node Adapter**: Connects nodes to the edge.
- **Node Agent**: Manages communication between nodes and business modules.
- **Business Modules**: Include database (DB) and other components.
- **Actuator**: Controls physical actions.
- **Cloud**: Hosts MobiLink and HANA for data management and analysis.

The system integrates edge devices with cloud services to optimize green roof operations.
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 Pipelne 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

<table>
<thead>
<tr>
<th>Test Case</th>
<th>SEEED</th>
<th>Plain</th>
<th>Impact</th>
</tr>
</thead>
<tbody>
<tr>
<td>Server-Side Only</td>
<td>Exact Search</td>
<td>2.0</td>
<td>1.7</td>
</tr>
<tr>
<td></td>
<td>Equi-Join</td>
<td>49.7</td>
<td>33.3</td>
</tr>
<tr>
<td></td>
<td>Grouping with Aggregation (Sum)</td>
<td>674.1</td>
<td>57.8</td>
</tr>
<tr>
<td>Incl. Client-Side</td>
<td>Order by Aggregate (Sum)</td>
<td>870.1</td>
<td>56.3</td>
</tr>
<tr>
<td>TPCH</td>
<td>Q4</td>
<td>2,402</td>
<td>235</td>
</tr>
<tr>
<td></td>
<td>Q5</td>
<td>1,373</td>
<td>207</td>
</tr>
</tbody>
</table>
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
@thedataneer

anil.goel@sap.com