SAP HANA Graph

Dmitry Buslov, SAP

Disclaimer

The information in this presentation is confidential and proprietary to SAP and may not be disclosed without the permission of SAP. Except for your obligation to protect confidential information, this presentation is not subject to your license agreement or any other service or subscription agreement with SAP. SAP has no obligation to pursue any course of business outlined in this presentation or any related document, or to develop or release any functionality mentioned therein.

This presentation, or any related document and SAP's strategy and possible future developments, products and or platforms directions and functionality are all subject to change and may be changed by SAP at any time for any reason without notice. The information in this presentation is not a commitment, promise or legal obligation to deliver any material, code or functionality. This presentation is provided without a warranty of any kind, either express or implied, including but not limited to, the implied warranties of merchantability, fitness for a particular purpose, or non-infringement. This presentation is for informational purposes and may not be incorporated into a contract. SAP assumes no responsibility for errors or omissions in this presentation, except if such damages were caused by SAP's intentional or gross negligence.

All forward-looking statements are subject to various risks and uncertainties that could cause actual results to differ materially from expectations. Readers are cautioned not to place undue reliance on these forward-looking statements, which speak only as of their dates, and they should not be relied upon in making purchasing decisions.

Scenarios

Bill of Material

Data source
- SAP ERP BOM (MARA, MAST, STKO, STPO)
- SAP SRM (EINA), non-SAP

Extended Bill of Material
- 250 k nodes (material, supplier, customer)
- 1 Mio. edges (supply, is-used-in, purchases)

Key use cases
- Supplier risk, cross-system "where used list"
- Margin, cost, price calculation
- Consistency checks

Challenges
- Data is distributed across multi systems
- Graph traversal + custom calculations

Graph processing
- BFS traversal with custom calculations

Utility Network

Data source
- SAP ERP (Asset Management /w custom extensions)

Electricity network
- 24 Mio. nodes (assets, cables)
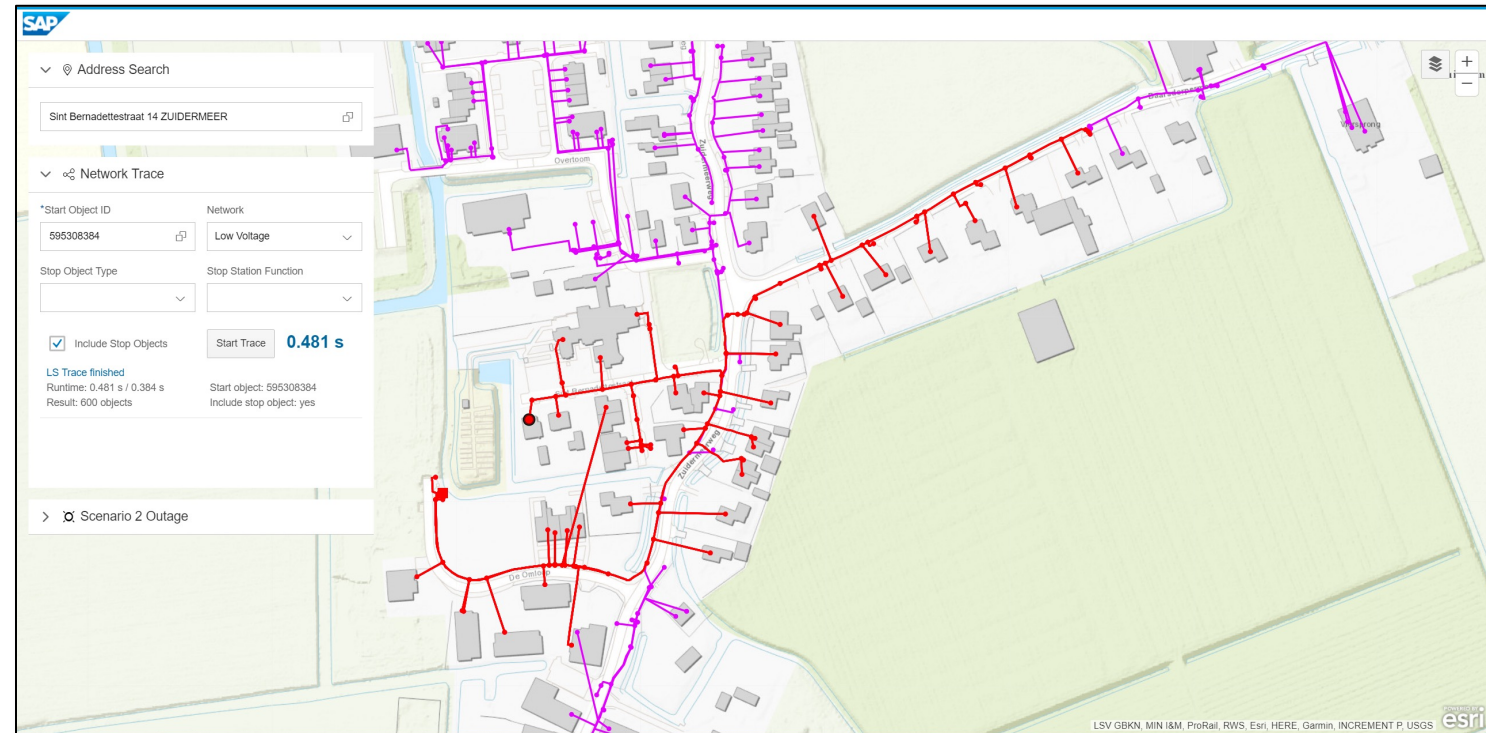- 47 Mio. edges (asset->cable->asset)

Key use cases
- Risk analysis for service desk
- Outage analysis
- Maintenance planning

Challenges
- Performance
- Spatial graph
- Spatial post-processing

Graph processing
- Neighborhood with stop conditions

Identifying ultimate beneficial owners (UBOs)

UBO simplified definition (weblink)

- All persons owning 25% or more of a company are UBOs
- All persons "controlling" a company which owns 25% or more are UBOs
- To assess "controlling ownership", you need to look at multiple paths and multiple levels of ownership

Challenges

- Efficiently traverse a large graph
- Evaluate logical conditions while traversing the graph
- Propagate percentages to the person nodes
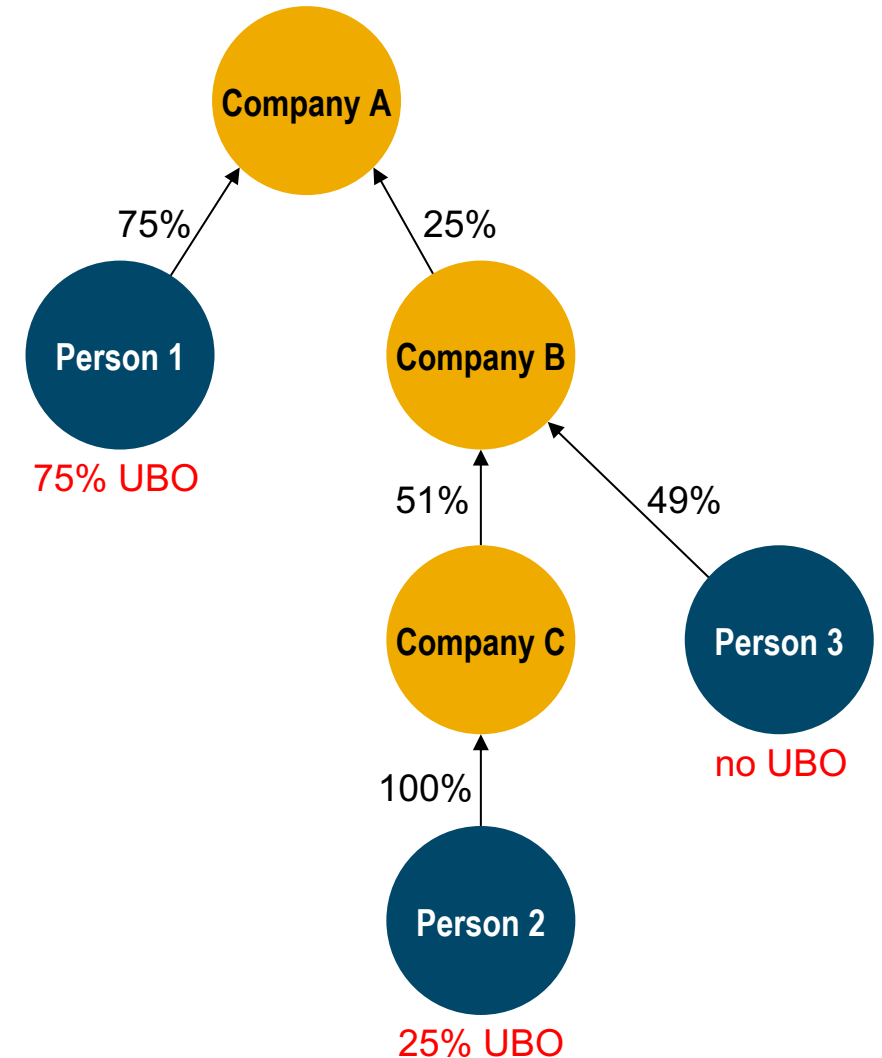
Cost-Based Paths

Data source
- Satellite images, spatial indices

Topological network
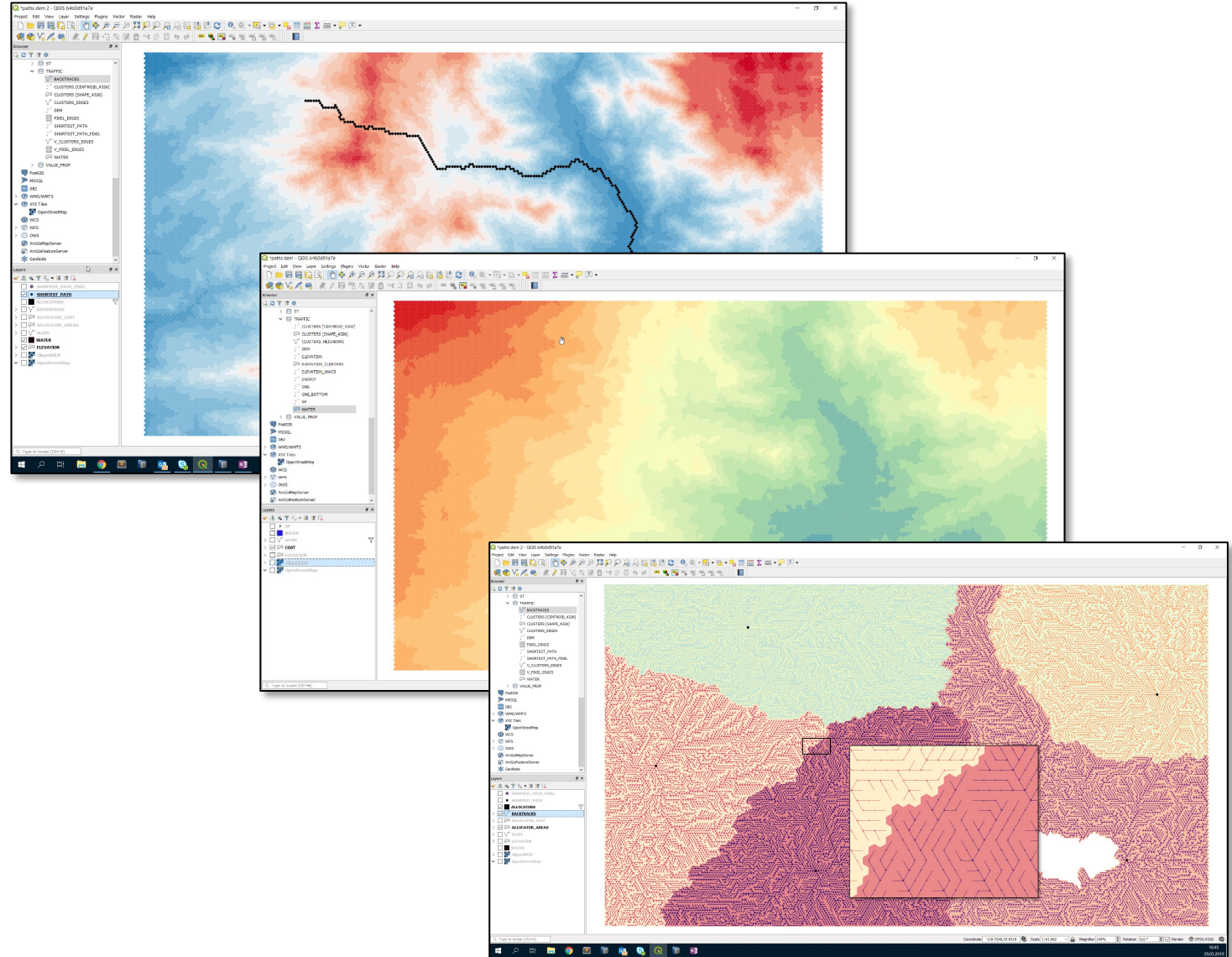- 40 k nodes (grid cells)
- 240 k edges

Key use cases
- Trafficability ("where can I go?")
- Evacuation paths, areas

Challenges
- Preprocessing (raster to vector)
- Spatial processing to generate a network
- Complex, dynamic cost function

Processing
- Raster to vector
- Spatial clustering
- Shortest paths (on-to-all)

Routing and Travel Time Prediction

Data source
- Road Network (static)
- Real-time traffic
- Predicted traffic (based on historic and real-time data)

Road Network with temporal traffic information
- 1.000 nodes (junctions of road segments)
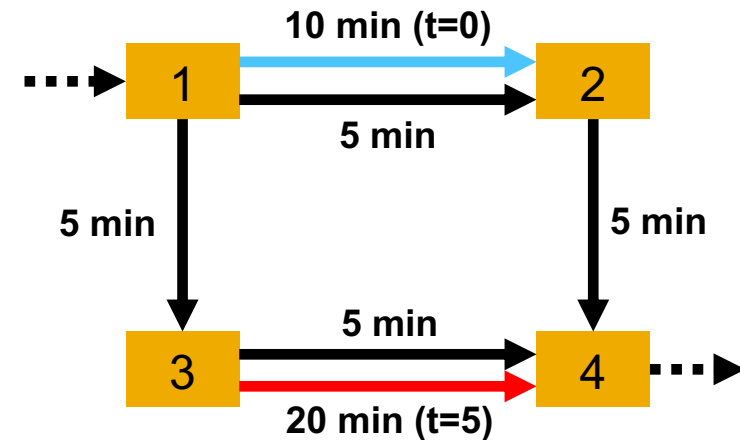- 50k edges (road segments with temporal traffic weights)

Key use cases
- Routing
- Travel time prediction
- Construction planning/traffic simulation

Challenges
- Temporal network, valid edges on a route depend on arrival time

Graph processing
- Custom GraphScript algorithm

Task: calculate best route from 1 to 4.

Edges are road segments.
Blue edge indicates current traffic (5 min slower than normal).
Red edge indicates predicted traffic at time t=5 (15 min slower).

Best route is 1->2->4, with and estimate travel time of 15 min.

SAP HANA Graph for Process Analysis

Project Evaluation and Review Techniques (PERT) is a method of analyzing tasks involved in completing a project/process.

- Calculating minimum time to complete the total process

The individual tasks can be modeled as events or edges in a graph and may have additional measures attached.
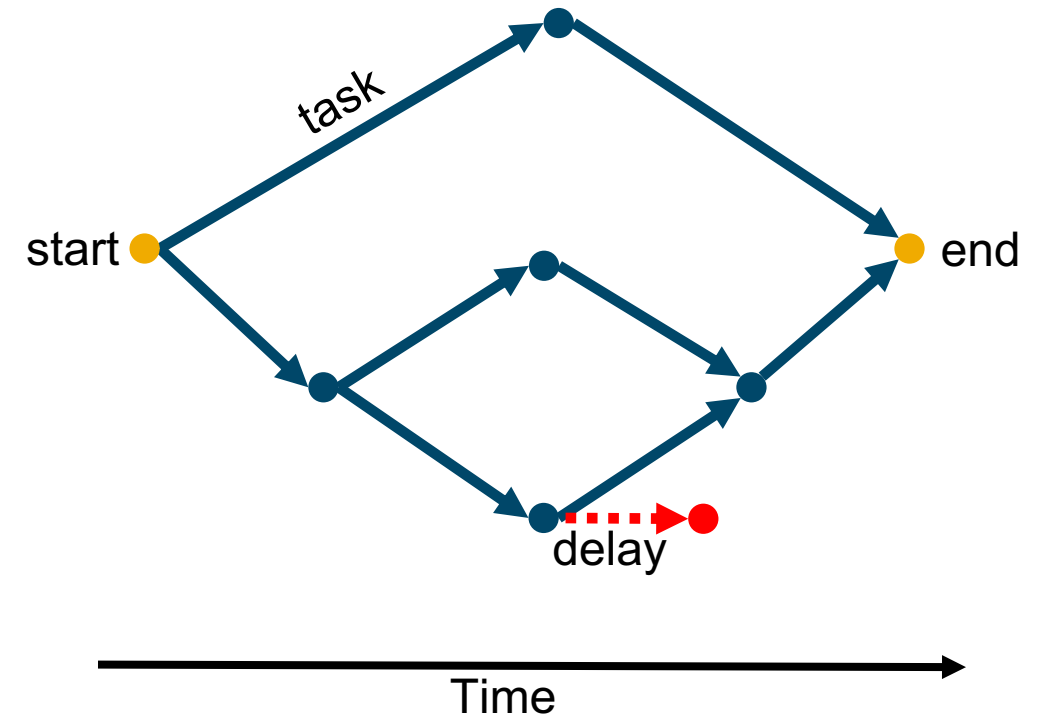
- Cost, resources, throughput, happy customers etc.

Prominent visualizations are Gantt charts or network diagrams.

Multiple instances of a process may run in parallel, competing for resources.

A larger business process is comprised of several smaller processes.

Goal: evaluate the impact of "shocks" in real-time.

Material Ledger/Actual Costing

Production chain analysis

Analyzing a production/value chain in order to understand the impact of price changes, e.g. in raw materials.

- Materials are sourced from a supplier, stock or production process

- Products are produced from materials

Other use cases

- Detect cycles in production process
- Understanding the BOM and structural changes over time
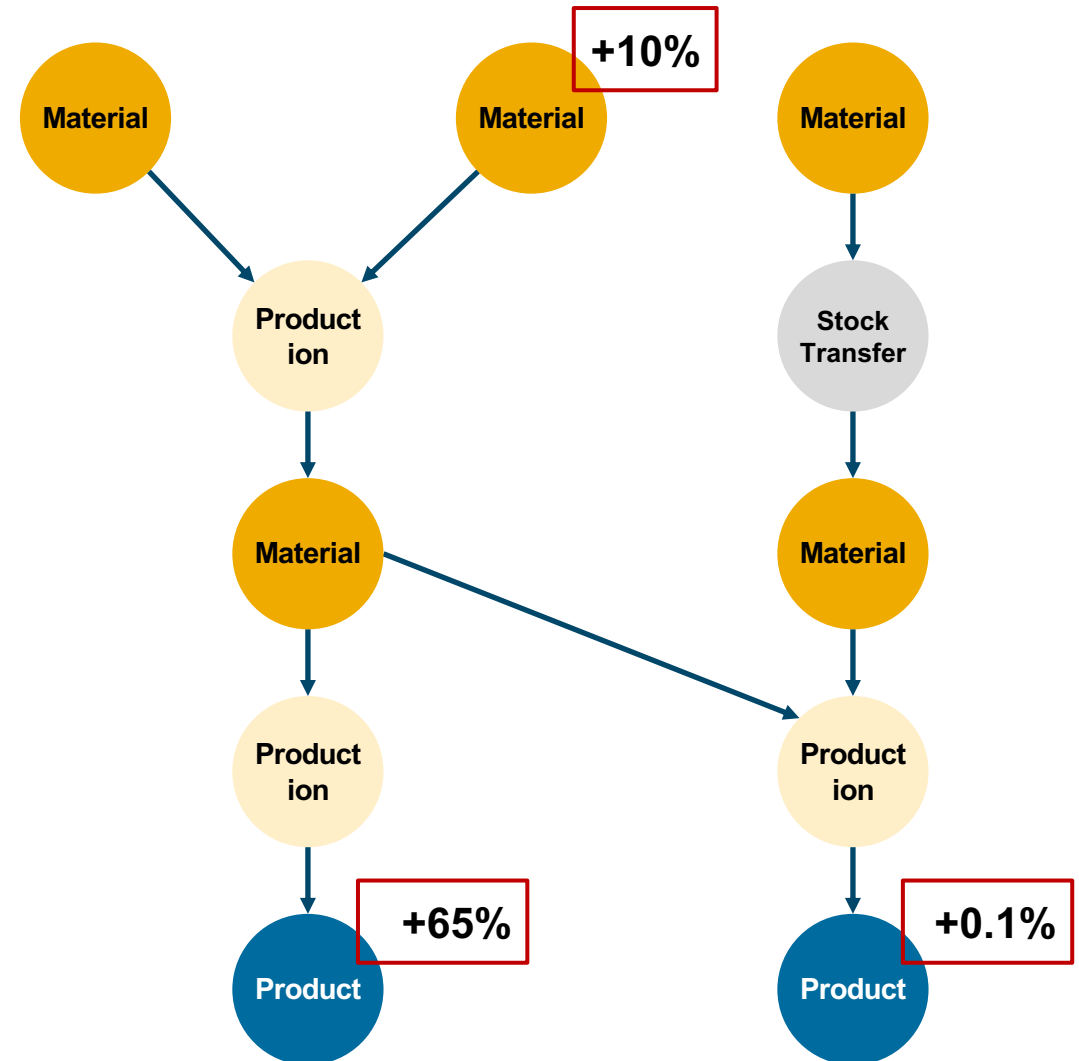
SAP HANA Info

SAP HANA: The business data platform for all applications

Accelerate with simplicity | Act with live intelligence | Innovate with confidence | Achieve cloud freedom

All Devices
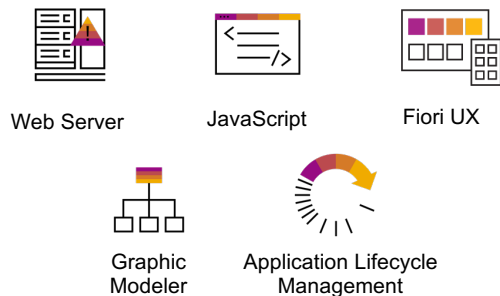
SAP, ISV and Custom Applications

SAP HANA Platform

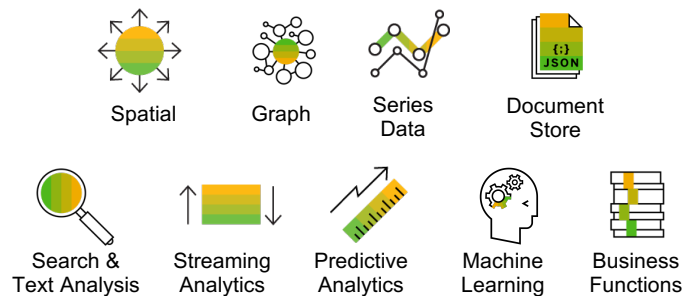**All-in-one, in-memory-first**

HTAP with broadest advanced analytics

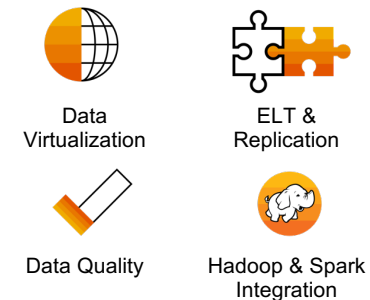**Secure & future-ready**

Hybrid and multicloud

Application Development

Web Server	JavaScript	Fiori UX

Graphic Modeler   Application Lifecycle Management

Advanced Analytical Processing

Spatial   Graph   Series Data   Document Store

Search & Text Analysis   Streaming Analytics   Predictive Analytics   Machine Learning   Business Functions

Data Integration and Quality

Data Virtualization   ELT & Replication

Data Quality   Hadoop & Spark Integration

Database Management

Columnar Store   Multi-Core & Parallelization   Advanced Compression   Multi-tenancy   Multi-Tier Storage   Data Modeling   Openness   Admin & Security   High Availability & Disaster Recovery
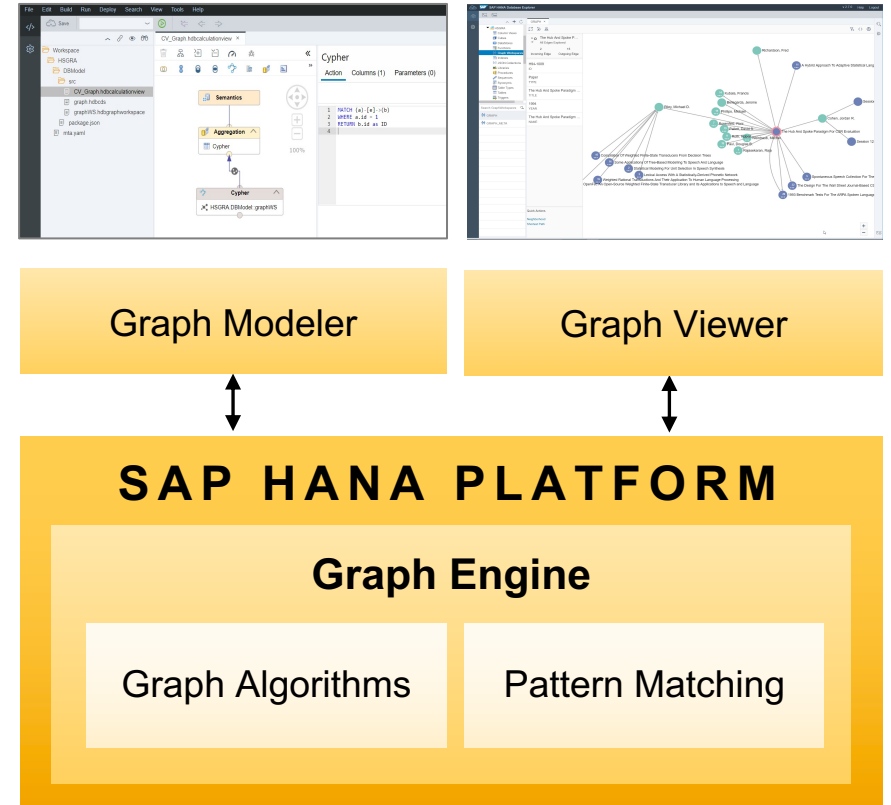
SAP HANA Graph

SAP HANA provides a native graph engine

- Property graph model embedded in relational/SQL, full transaction properties (ACID)
- Built-in functions like shortest path and strongly connected components
- Support for pattern matching using openCypher
- GraphScript to develop custom graph algorithms
- Graph viewer, SAP HANA Database Explorer

Benefits

- Store and analyze connected data
- Combine text, spatial, and advanced analytics with graph intelligence
- Tightly integrated in SAP HANA operations (security, backup/restore, scale-out, import/export etc.)

Graph Modeler

Graph Viewer

SAP HANA PLATFORM

Graph Engine

Graph Algorithms

Pattern Matching
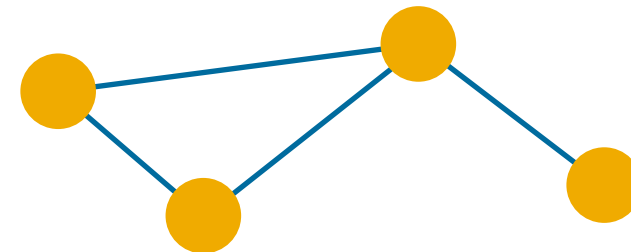
SAP HANA Graph
Relational and Graph Processing in ONE Platform
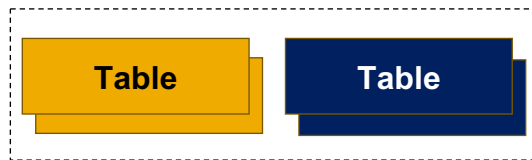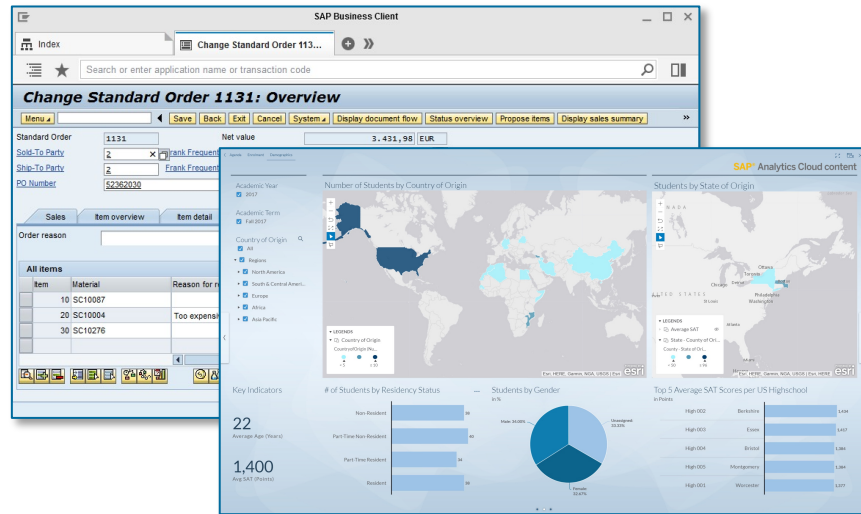
SQL integration
- text, spatial, predictive

Programming Model (GraphScript)
- database procedures

Built-in algorithms/functions
- paths, neighbors, BFS, DFS, degree etc.
- performance

Pattern Matching (openCypher)

graph construction

One operations concept, one copy of data

SAP HANA Graph

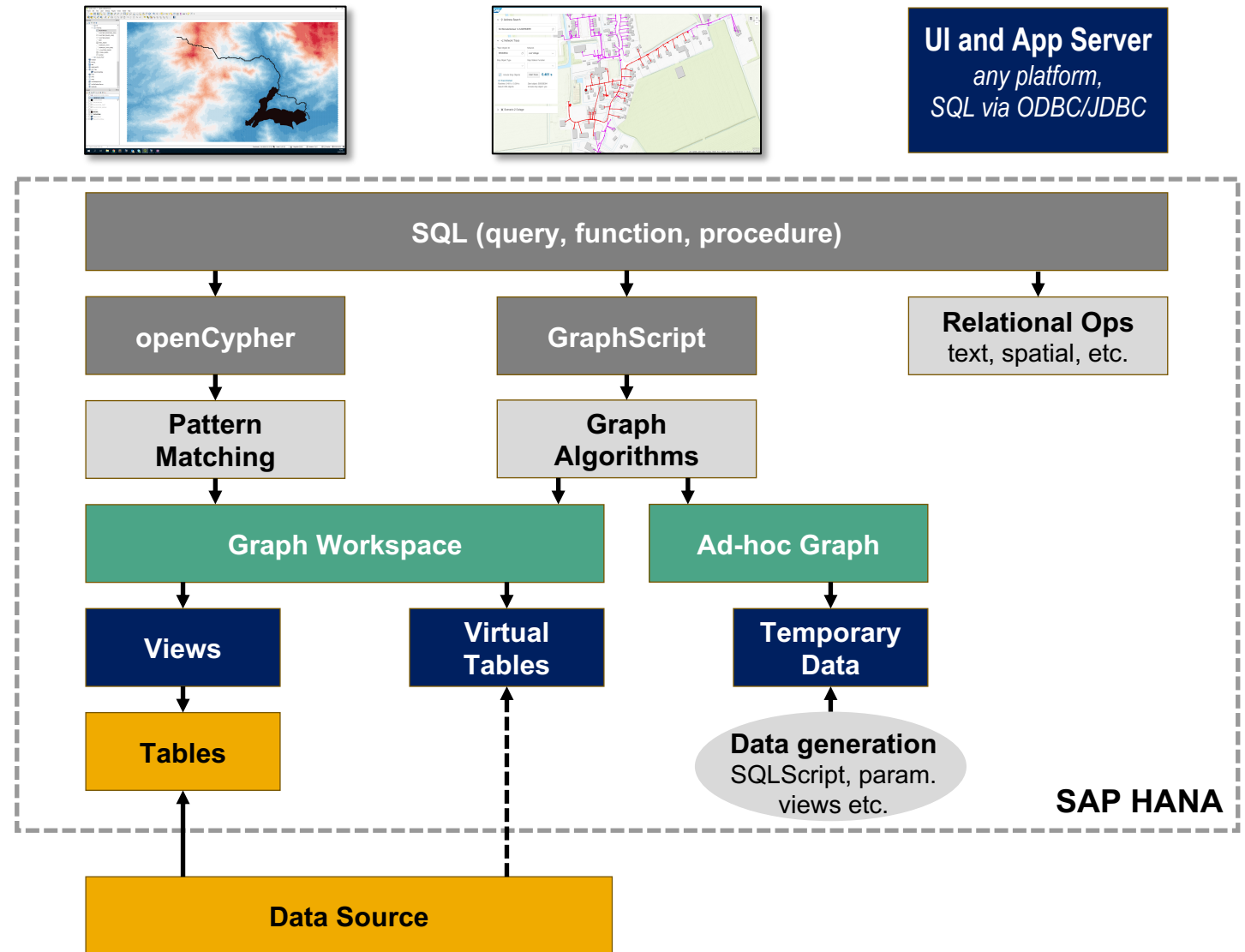Architecture

GraphScript DSL
- Database procedures
- Custom algorithms, like PageRank, Centrality

Built-in algorithms/functions
- Neighbors
- Breadth First Search (BFS)
- Depth First Search (DFS)
- Strongly Connected Components
- Shortest Paths (1-1, 1-all, top k)

openCypher
- "standard" for pattern matching
- Declarative language
- *MATCH (n1)-[e]->(n2)*

UI and App Server
any platform,
SQL via ODBC/JDBC

SQL (query, function, procedure)

openCypher **GraphScript** **Relational Ops**
text, spatial, etc.

Pattern Matching **Graph Algorithms**

Graph Workspace **Ad-hoc Graph**

Views **Virtual Tables** **Temporary Data**

Tables

Data generation
SQLScript, param.
views etc.

SAP HANA

Data Source

Nodes, Edges, and Workspaces

ID	TYPE	NAME	YEAR
AUT-6841	Author	Richardson, Fred	
H94-1009	Paper	The Hub and Spoke Paradigm for CSR Evaluation	1994
ORG-523	Organization	Boston University	
H92-1076	Paper	Spontaneous Speech Collection for the CSR Corpus	1992

ID	SOURCE	TARGET	TYPE
1	H94-1009	AUT-6841	isAuthoredBy
2	AUT-6841	ORG-523	isAffiliatedTo
3	H94-1009	H92-1076	citation

Nodes, Edges, and Workspaces

Node Identifier

Attributes

ID	TYPE	NAME	YEAR
AUT-6841	Author	Richardson, Fred	
H94-1009	Paper	The Hub and Spoke Paradigm for CSR Evaluation	1994
ORG-523	Organization	Boston University	
H92-1076	Paper	Spontaneous Speech Collection for the CSR Corpus	1992

Edge Identifier **Source (node)** **Target (node)**

Attributes

ID	SOURCE	TARGET	TYPE
1	H94-1009	AUT-6841	isAuthoredBy
2	AUT-6841	ORG-523	isAffiliatedTo
3	H94-1009	H92-1076	citation

Nodes, Edges, and Workspaces

A GRAPH
WORKSPACE
exposes the data
to the graph engine

```
CREATE GRAPH WORKSPACE [SCHEMA].[NAME]
  EDGE TABLE [SCHEMA].[EDGE TABLE/VIEW]
    SOURCE COLUMN source
    TARGET COLUMN target
    KEY COLUMN id
  VERTEX TABLE [SCHEMA].[NODE TABLE/VIEW]
    KEY COLUMN id;
```

ID	TYPE	NAME	YEAR
AUT-6841	Author	Richardson, Fred	
H94-1009	Paper	The Hub and Spoke Paradigm for CSR Evaluation	1994
ORG-523	Organization	Boston University	
H92-1076	Paper	Spontaneous Speech Collection for the CSR Corpus	1992

ID	SOURCE	TARGET	TYPE
1	H94-1009	AUT-6841	isAuthoredBy
2	AUT-6841	ORG-523	isAffiliatedTo
3	H94-1009	H92-1076	citation

GraphScript
Overview

GraphScript is a high-level, powerful domain-specific language for native stored procedures in SAP HANA. It is specifically designed to ease the development and integration of custom graph algorithms into the existing data management workflows.

```
CREATE PROCEDURE "[SCHEMA]"."[PROCEDURE NAME]" (OUT distance DOUBLE)
LANGUAGE GRAPH READS SQL DATA AS
BEGIN
    GRAPH g = GRAPH("[SCHEMA]"."[GRAPH WORKSPACE NAME]");
    VERTEX v1 = VERTEX(:g, 1);
    VERTEX v2 = VERTEX(:g, 2);
    WEIGHTEDPATH<BIGINT> p = SHORTEST_PATH(:g, :v1, :v2);
    distance = LENGTH(:p);
END;
```

Path Finding is a Fundamental Operation

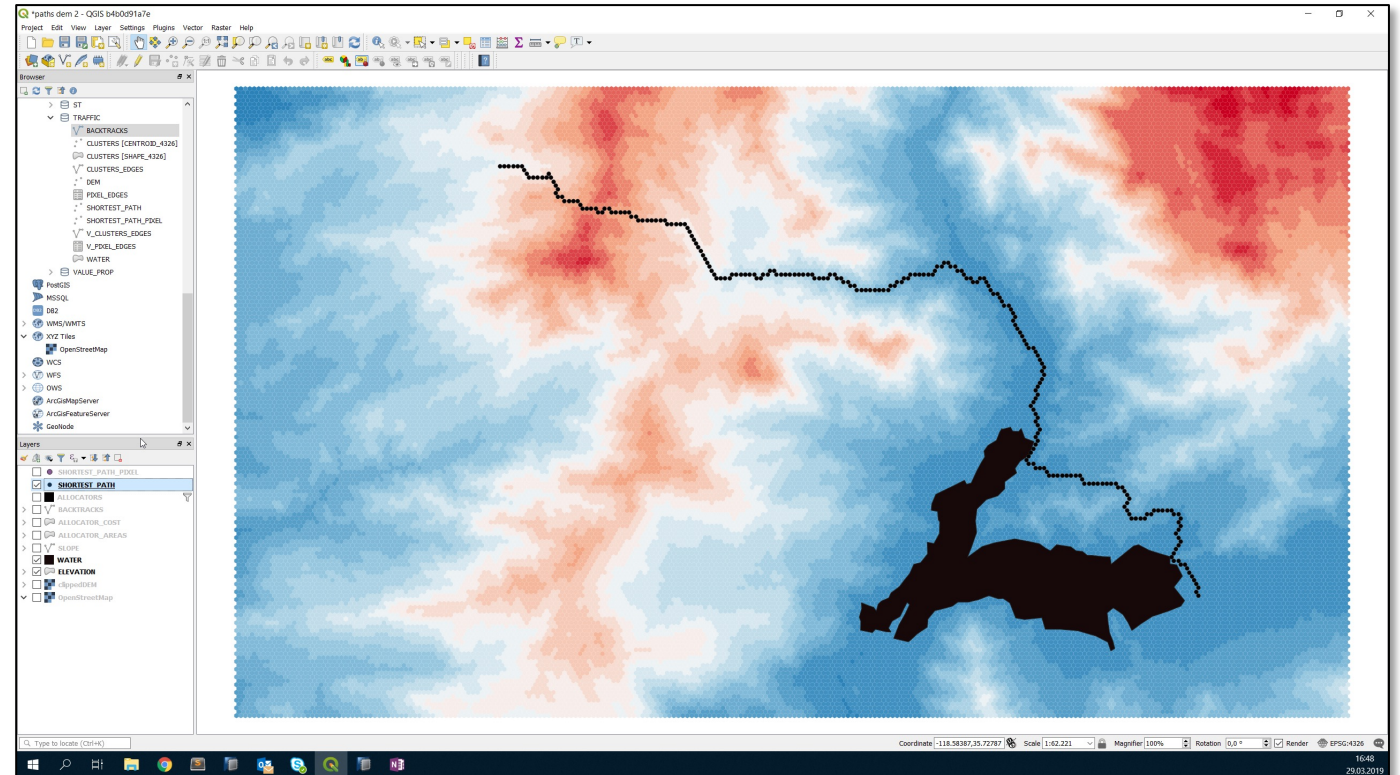Basic operations of graph analysis algorithms

- Neighbors, Breadth First Search, Depth First Search, Shortest Path

Shortest Path variants

- Shortest Path One-to-One
- Shortest Path One-to-All
- Top k Shortest Paths

Shortest Path cost functions

- Hop distance
- Cost function
  e.g. distance, price, attractivity
- Temporal validity
  e.g. congestion predicted for 3pm
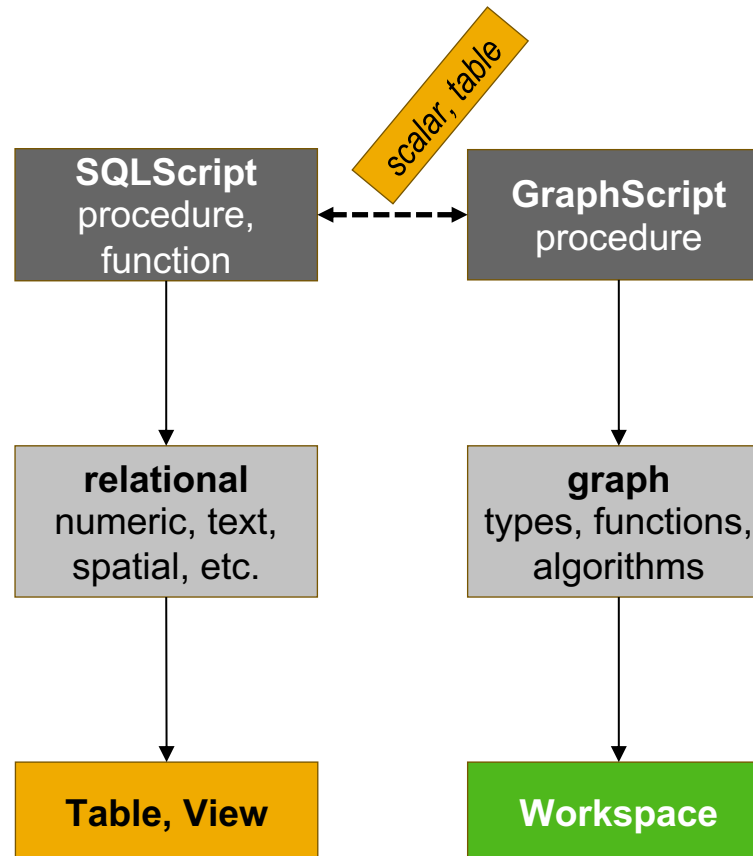
GraphScript
GraphScript and SQLScript

SQLScript

- SQL-like language for definition of stored procedures and functions

Relational

- SQL data types, functions, expressions, predicates

Base objects

- Tables and views

SQLScript procedure, function	scalar, table	GraphScript procedure

| relational
numeric, text,
spatial, etc. | | graph
types, functions,
algorithms |

| Table, View | | Workspace |

GraphScript

- Language for definition of graph-specific stored procedures

Graph

- Graph-specific data types, functions, expressions, algorithms

Base object

- Graph workspace

Further Information

SAP HANA Graph reference guide

- https://help.sap.com/viewer/f381aa9c4b99457fb3c6b53a2fd29c02/latest/en-US

openSAP – Analyzing Connected Data with SAP HANA Graph

- https://open.sap.com/courses/hsgra1

Developer Tutorial - Get Started with SAP HANA Graph

- https://developers.sap.com/group.hana-aa-graph-overview.html

SAP HANA Academy – Graph playlist

- https://www.youtube.com/playlist?list=PLkzo92owKnVwCuJeNPcC7J_v4eT5_s6-d

Code examples

- https://github.com/SAP-samples/hana-graph-examples

Geo+Graph+Search in HANA powered by QGIS

- https://github.com/SAP-samples/teched2021-DAT262

Homework

1) Create your own HANA Cloud trial instance or install HANA Express

https://developers.sap.com/tutorials/hana-cloud-deploying.html

2) Upload Moscow drive graph into HANA Cloud:

HINT:

import osmnx

G = ox.graph_from_place('Москва',network_type='drive')

3) Find all connected objects for ID 1143424870 (path min 1- max 5)

HINT:

The most simple way to do it - >