FEDERATED KPI AGGREGATION WITH LOCAL PRIVACY CONSTRAINTS AND COORDINATED OLAP ROLL-UPS

PD Dr. Felix Härer

Center for Digital Trust FHNW
University of Applied Sciences and Arts Northwestern Switzerland
Basel, Switzerland

University of Applied Sciences and Arts
Northwestern Switzerland



OUTLINE

1. **Introduction and Background**
2. Research Problem and Research Questions
3. Approach: Coordinated OLAP Roll-Ups
4. Prototype and Evaluation Results
5. Conclusion



1. Introduction and Background – Business Intelligence and Analytical Granularity

- **KPIs aggregate operational measures over dimensions [1–3]**
  - A data warehouse combines a measure such as revenue with time, geography, customer, and product dimensions.
  - OLAP hierarchy levels support **roll-up** to a coarser view, such as city to region or month to quarter.

- **Granularity determines which groups a query returns**
  - SUM(revenue) by city and exact age creates more detailed groups than SUM(revenue) by region and age decade.
  - Coarser groups combine contributions that were separate at the requested level.

[1] Chaudhuri and Dayal, “An Overview of Data Warehousing and OLAP Technology,” 1997.
[2] Gray et al., “Data Cube,” 1997. [3] Kimball and Ross, The Data Warehouse Toolkit, 2013.



1. Introduction and Background – Federated Analytics and Local Release Constraints

- **Federated sites may impose independently administered data protection constraints [4, 5]**
  - Each site retains its own warehouse or "data mart" and executes computations locally, e.g., within or across organizations.
  - The federation exchanges derived results while detailed records remain under local control.

- **Small local groups can expose sensitive information**
  - Statistical disclosure control [6] and similar approaches may be used.
  - A detailed query may cause aggregated data based on few local contributions, independent of federation-wide group size.
  - A site may require a **minimum cohort size** before releasing a KPI for a group.

[4] Wang et al., "A Survey on Federated Analytics," 2026. [5] Srinivas et al., "PAPAYA Federated Analytics Stack," NSDI 2025.
[6] Hundepool et al., Statistical Disclosure Control, 2012.



2. Problem and Research Questions

- **Problem: Local Contribution and Coordination**
  - **To prevent sensitive information exposure, local data contributions can be surpressed** according to a threshold or disclosure rule, which might cause **incomplete aggregate results**.
  - **Independently administered sites must coordinate the selection** while keeping capability, authorization, and execution decisions local.
- **RQ1: Common roll-up selection**
  - How can a federation select the least-generalized common OLAP roll-up that satisfies the local minimum-cohort threshold at every site?
- **RQ2: Coordination**
  - How can agents coordinate this selection across independently administered sites and still keep capability, authorization, and execution decisions local?

Related work covers privacy-preserving OLAP [9–11] and agent-built federated analytics workflows [14].
[9] Agrawal et al., 2005. [10] Zhang and Zhao, 2011. [11] Fugkeaw et al., 2024. [14] Ji et al., LAFA, 2025.



3. Coordinated OLAP-Rollups

- **Local Cohort Rules**
  - Each site evaluates its own effective threshold
  - The federation sends *k*. Site *i* applies the stricter of the federation rule and its local rule.
- **Common-Level Selection**
  - A candidate must be executable and safe at every site
  - Every non-empty local output group must contain at least κ_i contributing privacy units.
  - Order executable candidates by normalized hierarchy loss I(L). Select the first candidate that every site accepts.

$$\kappa_i = \max(k, k_i)$$

$$L^* = \arg\min_{L \in C(Q)} I(L), \text{ subject to } \mathrm{safe}_i(L) \text{ at every site } i$$



3. Coordinated OLAP-Rollups – Example

Suppose there are three sites and $k = 10$

- **Requested KPI:** SUM(revenue) by city and exact age
  - Minimum cohort size among non-empty groups at each site:

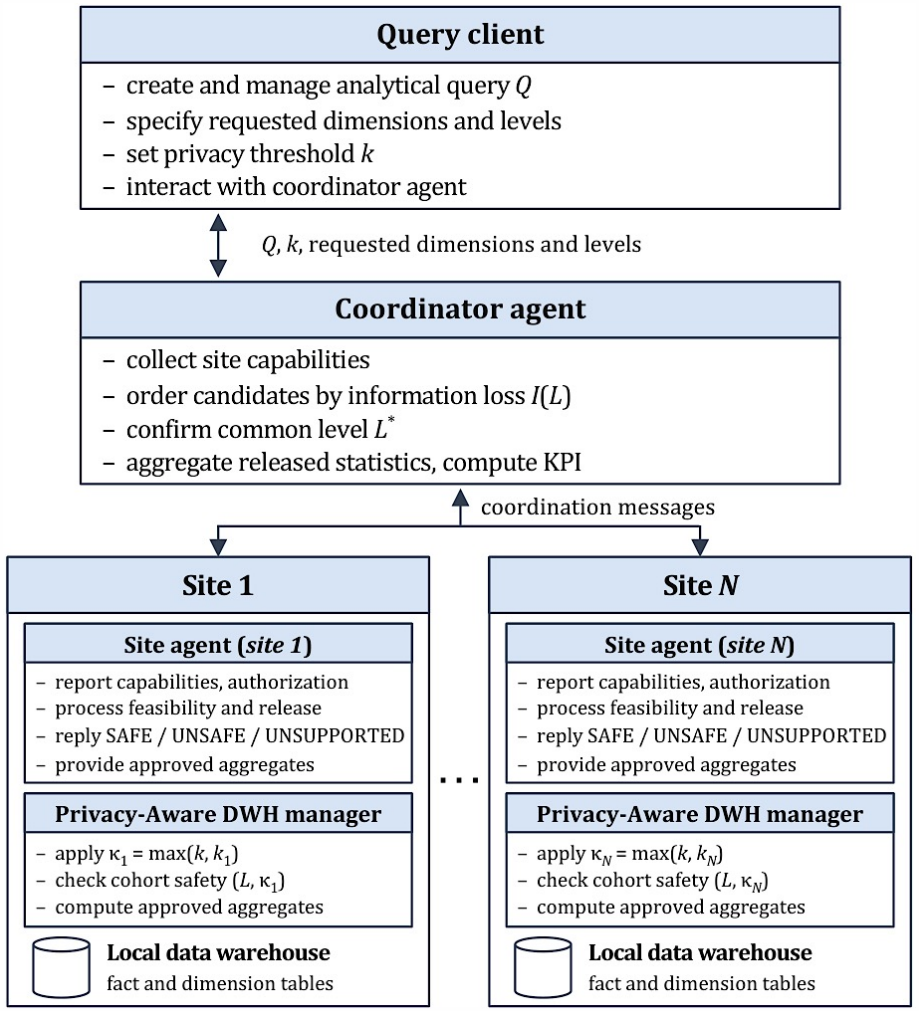
Candidate	OLAP Grouping Dimensions: Geography x Age	Site 1	Site 2	Site 3	Decision
L_0	City × exact age	4	12	6	Unsafe
L_1	Region × exact age	8	15	9	Unsafe
L_2	City × decade	7	13	8	Unsafe
L_3	**Region × decade**	**14**	**18**	**12**	**Select**

Geography: city → region → country. Age: exact → decade → age band.

- **Coordination is required to confirm L_3**



3. Coordinated OLAP-Rollups – Roll-Up Architecture

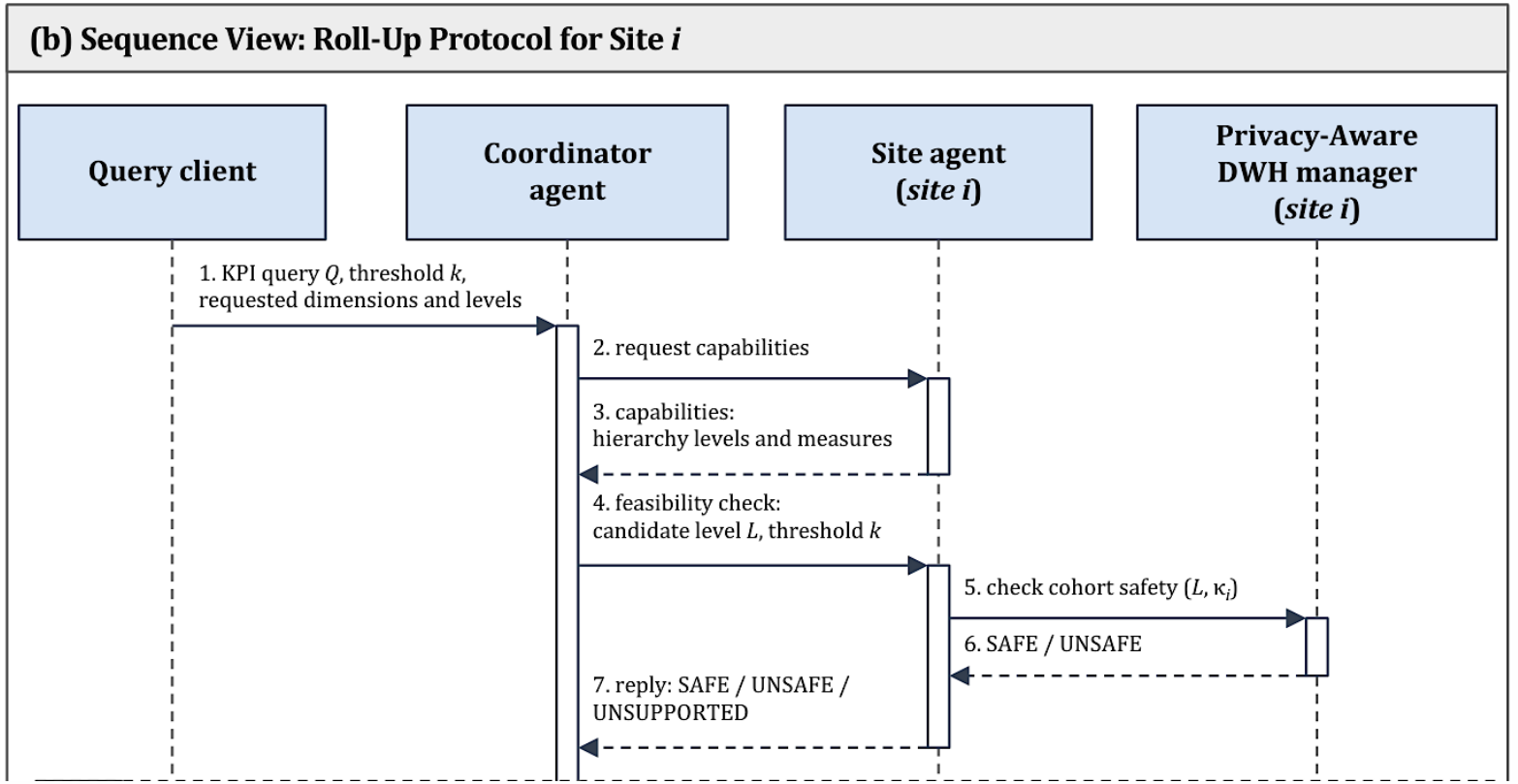


Components:

- **Query Client**
  - Creates and manages query, sets privacy theshold *k*
- **Coordinator agent**
  - Collects current site capabilities and orders candidate roll-ups.
  - Confirms a common level and combines released statistics.
- **Site agent and privacy-aware DWH manager**
  - Keep capability, authorization, and tool execution state local.
  - Check cohort safety and compute approved aggregates in deterministic functions.

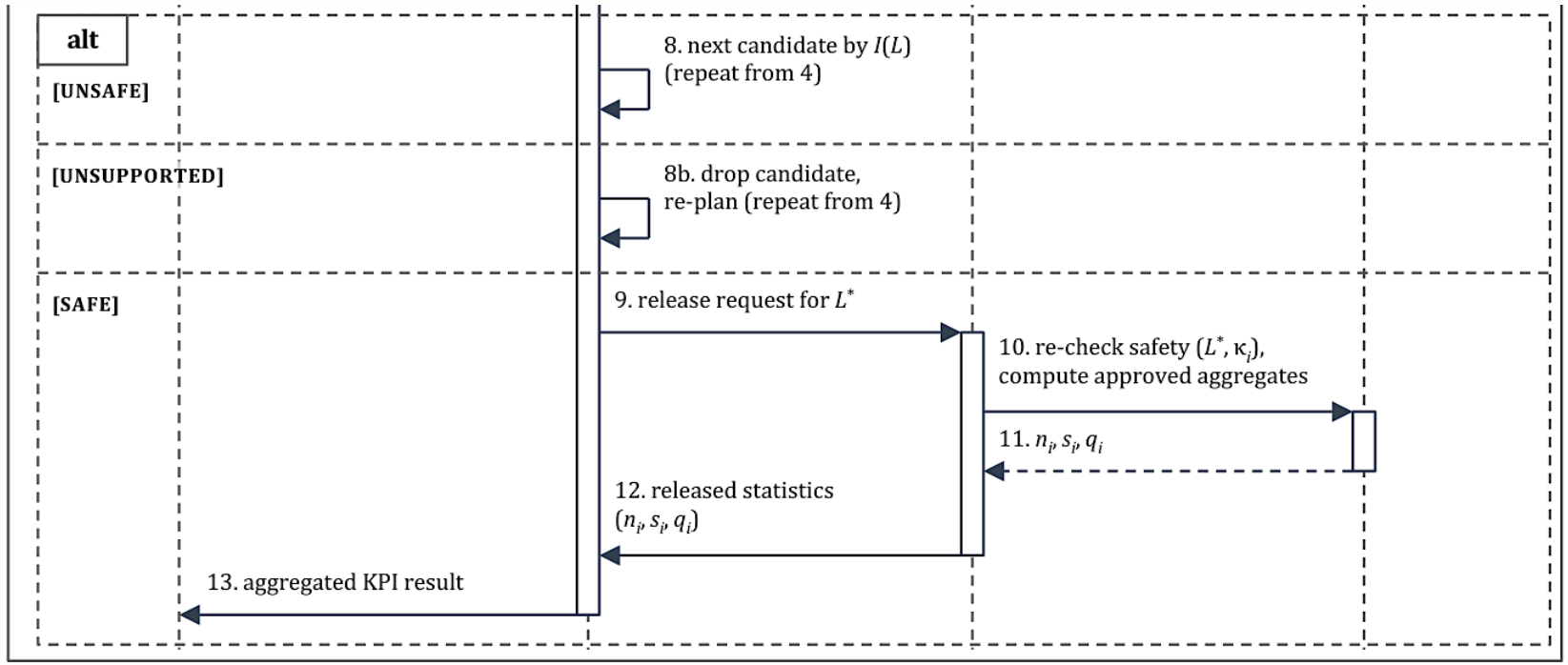
3. Coordinated OLAP-Rollups – Roll-Up Protocol

- **Capability exchange and candidate evaluation**



3. Coordinated OLAP-Rollups – Roll-Up Protocol

- **Candidate response and guarded release**
  - **UNSAFE or UNSUPPORTED:** The coordinator advances to another candidate or replans.
  - **SAFE at every site:** Confirm L*, recheck local safety, and release local statistics n_i, s_i, q_i.



3. Coordinated OLAP-Rollups – Roll-Up Protocol

- For each candidate evaluation with increasing l(L), site agents release additive local statistics for each group

  Example: $n_i = \text{COUNT}(x)$ $s_i = \text{SUM}(x)$ $q_i = \text{SUM}(x^2)$

- The coordinator agent aggregates the first common level accepted by each site

  Example: $n = \Sigma_i n_i$ $s = \Sigma_i s_i$

  Mean: $\mu = s / n$

  Population variance: $\sigma^2 = (\Sigma_i q_i) / n - \mu^2$



4. Prototype and Evaluation Results

- **Python prototype with four independent local warehouses**
  - **SQL Star schema** per site, with sales facts and customer geography, customer age, product, and time dimensions for hierarchies and roll-ups.
  - Adapter for the ***Hermes*** **agent** and the ***Buzz*** **agent communication** platform for sharing the local privacy and aggregation functions
- **Synthetic workload**
  - **4** sites, **450 customers** and **5,000 fact rows** per site, **six 2–3-dimensional queries**.
  - **18 query cases** per threshold, for k = 3, 5, 10, 20.
- **Measures and Results**
  - **Measures**: common-level selection results, candidate statistics, coordinator aggregation statistics. Test cases included normal, edge, and failure cases, including local capability changes.
  - **Results**: expected selection levels and no aggregation errors for candidate and coordinator statistics.

Prototype and results: https://github.com/fhaer/Federated-Aggregation-Coordination-and-Privacy



5. Conclusion

- **Results for common roll-up selection (RQ1) and agent-based coordination (RQ2)**
  - **Coordinated Roll-Up Architecture and Protocol**
    - Coordination agent requests candidate roll-ups at site agents for evaluation until a minimum accepted level is found.
    - Site agents retain capability and authorization decisions such that they can decide sufficient aggregation levels for privacy and report aggregate results accordingly
  - **Prototype implementation** showing feasibility
  - **Overall, the approach provides federated aggregation through coordinated granularity**

- **Outlook**
  - In future work, the current threshold-based candidate selection may be improved by further privacy metrics, anonymization approaches, or additional query controls

