

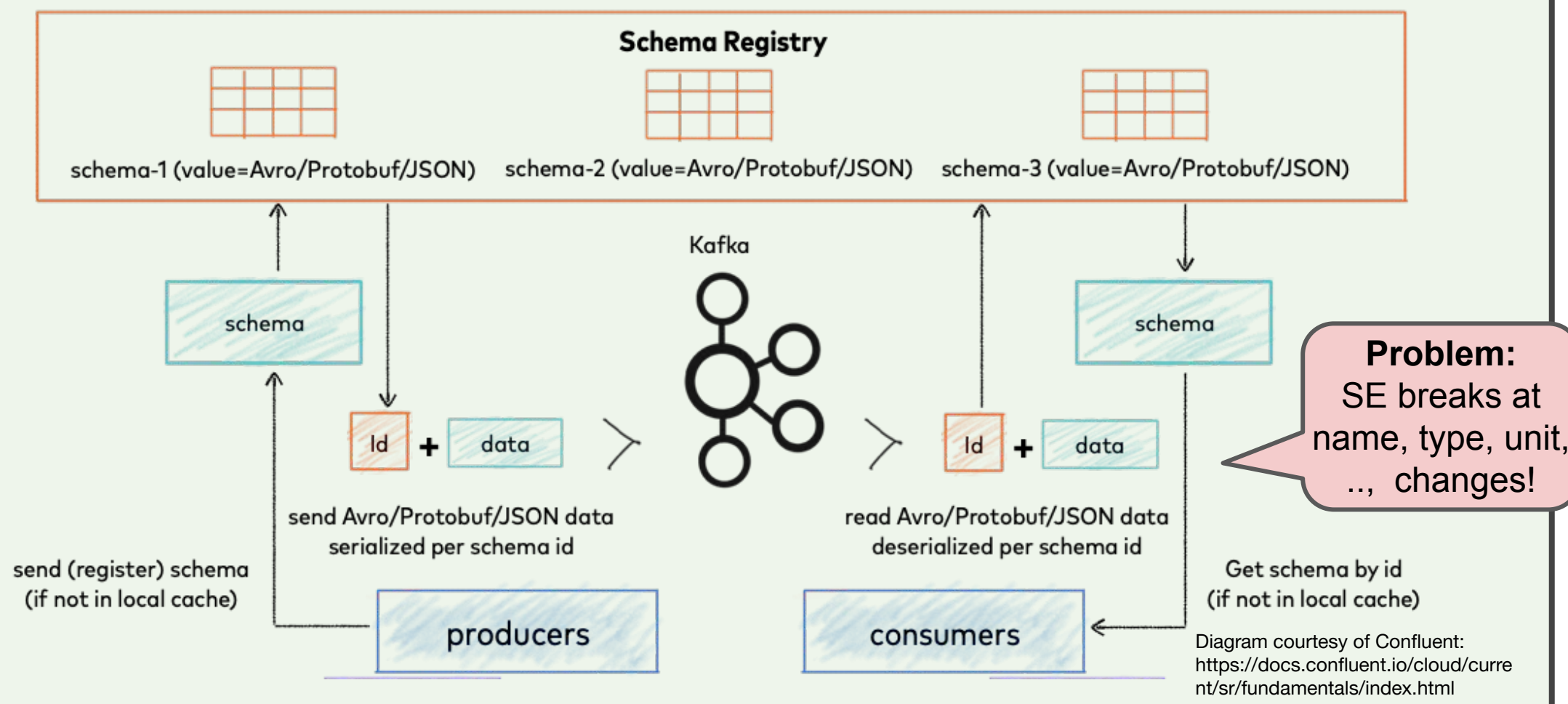
Compound Schema Registry

Silvery Fu^{1,2}, Xuewei (Sylvia) Chen¹

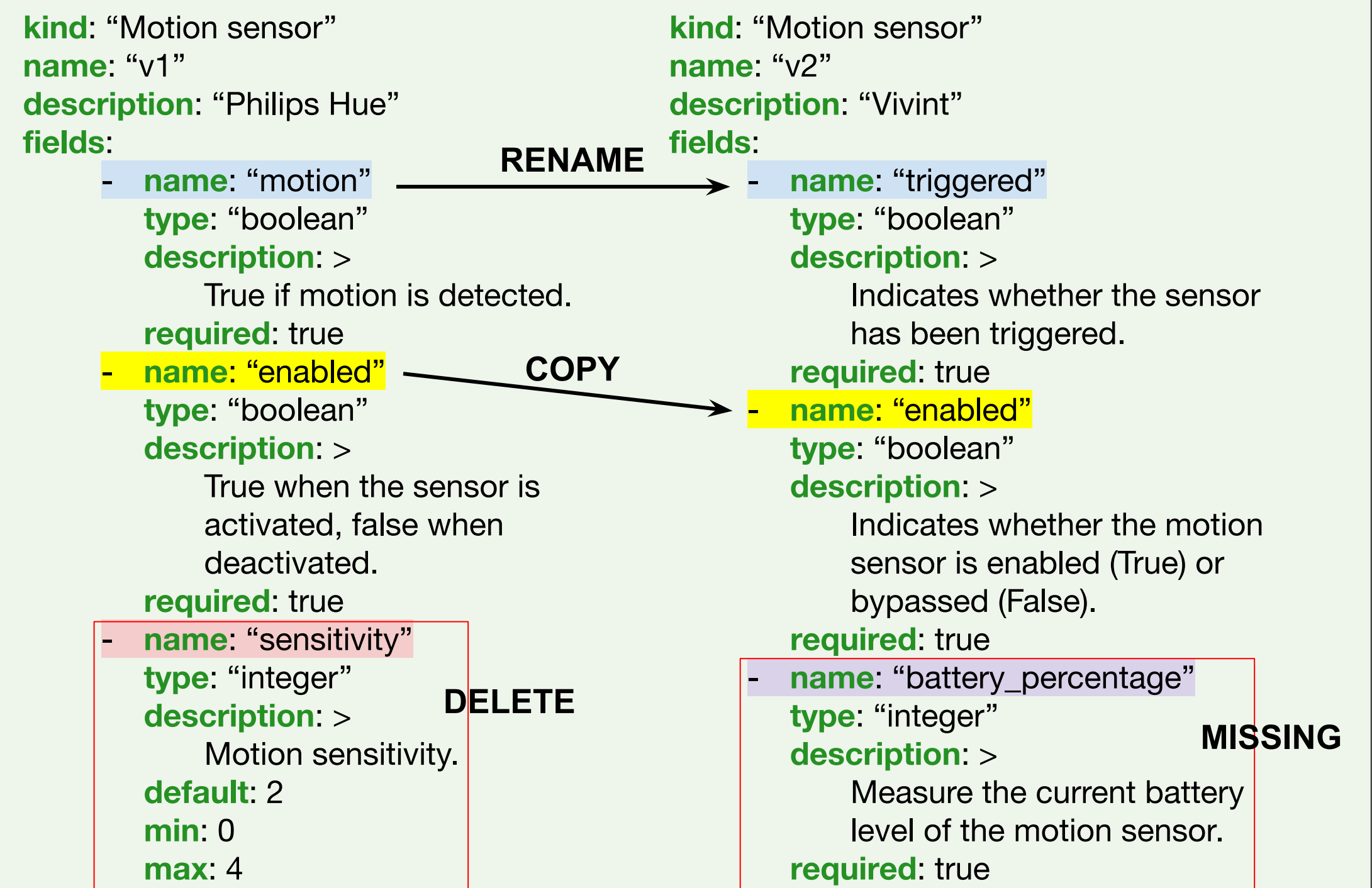
¹UC Berkeley, ²System Design Studio



Schema Registry: Today



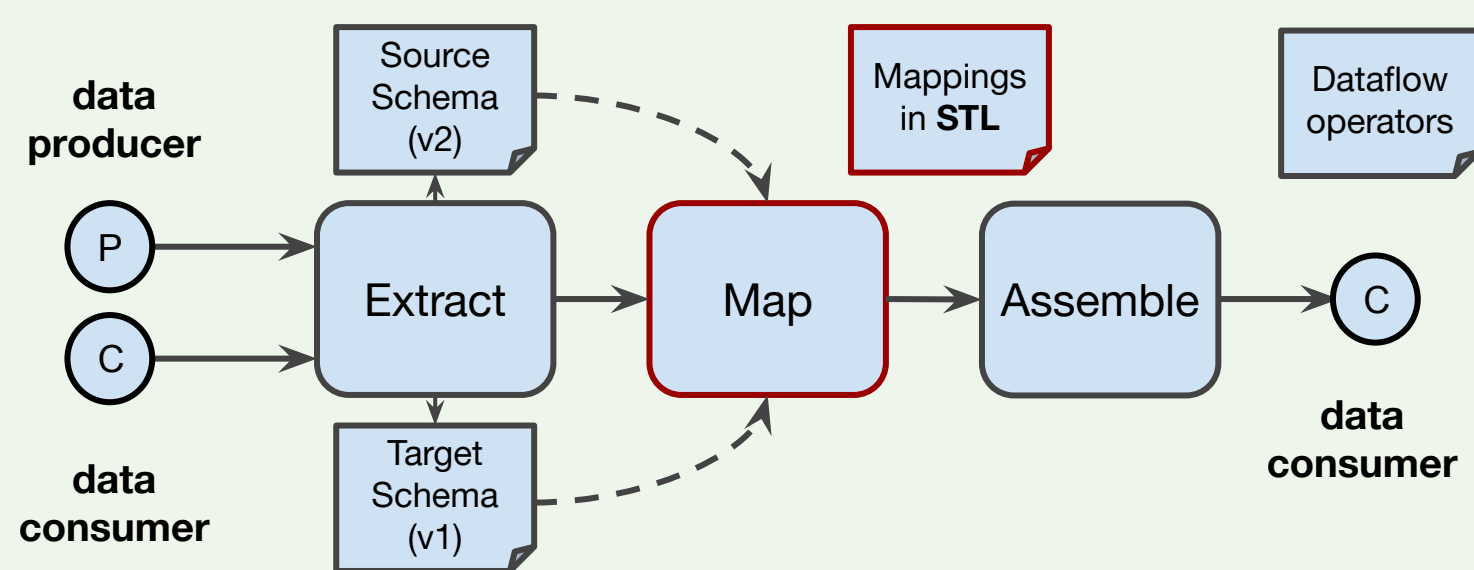
Task: Schema Mapping



Goal: Generalized Schema Evolution

- Enable data consumers to auto-adapt to schema changes by producers.
- **Challenge:** Must ensure accurate mapping of fields and values.
- **Challenge:** Must avoid model calls in the data path due to perf. and cost.

Design Pattern: Task-Specific Language and IR



- **STL defines a collection of schema mapping commands**
 - *Schema matching* commands assess compatibility between whole schemas
 - *Field transformation* commands directly match and modify schema fields
 - *Value transformation* commands convert field values to fit new schema specifications
- **STL streamlines schema evolution with fine-grained task decomposition**
 - STL decomposes schema mapping into unambiguous sub-tasks via task-specific commands
 - As an IR, STL separates *mapping gen* from *dataflow assembly*, simplifying both tasks
- **STL improves schema mapping accuracy from 78% to 94%**

Schema Transformation Language (STL)

Command class	Command name	Description
Schema matching	MATCH	Used to determine whether the source and target schemas correspond to the same entity; if they match , the schema mapping will continue; otherwise , it will abort .
	COPY	Directly copies data from the source field to the target field without any transformation .
Field transformation	ADD	Inserts a new field into the target schema that does not exist in the source schema.
	CAST	Converts the data type of the source field to match the expected type of the target field.
	DELETE	Removes the field from the source schema when it is not required in the target schema.
	RENAME	Changes the name of the source field to match the name of the target schema.
	DEFAULT	Assigns a predefined default value to a target field when data is unavailable or null.
	MISSING	Used when no appropriate mapping exists to map the source field to a target field, implying a schema mapping failure.
	SCALE	Adjusts the numerical values in the source field by a specified factor according to the value in the target field.
Value transformation	SHIFT	Modifies the values in the source field by adding or subtracting a constant value .
	LINK	Establishes a correspondence between values in the source field and defined values in the target field, used for fields with enum type .
	GEN	Generate a transformation function that defines how to convert values from the source field to fit the target field's requirements.
	APPLY	Applies a transformation function , either generated or predefined by the developer, to the value of a source field to derive the value of the target field.

Model call take STL specification and generates STL commands to invoke

```
{from: triggered, to: motion, transformation: RENAME triggered TO motion}
{from: battery_percentage, to: None, transformation: DELETE battery_percentage}
{from: None, to: sensitivity, transformation: ADD sensitivity TYPE integer}
{from: sensitivity, to: sensitivity, transformation: DEFAULT sensitivity TO 2}
{from: enabled, to: enabled, transformation: COPY}
```

Generated mappings are sent to assembly stage to get compiled to dataflow operators

Source schema	Target schema	Precision		Recall		F1	
		STL	Base	STL	Base	STL	Base
Philips Hue	Vivint	0.91	0.73	0.98	0.83	0.94	0.78
SimpliSafe	Vivint	1	0.2	0.8	0.2	0.89	0.2
SimpliSafe	Philips Hue	1	0.8	0.9	0.67	0.95	0.72

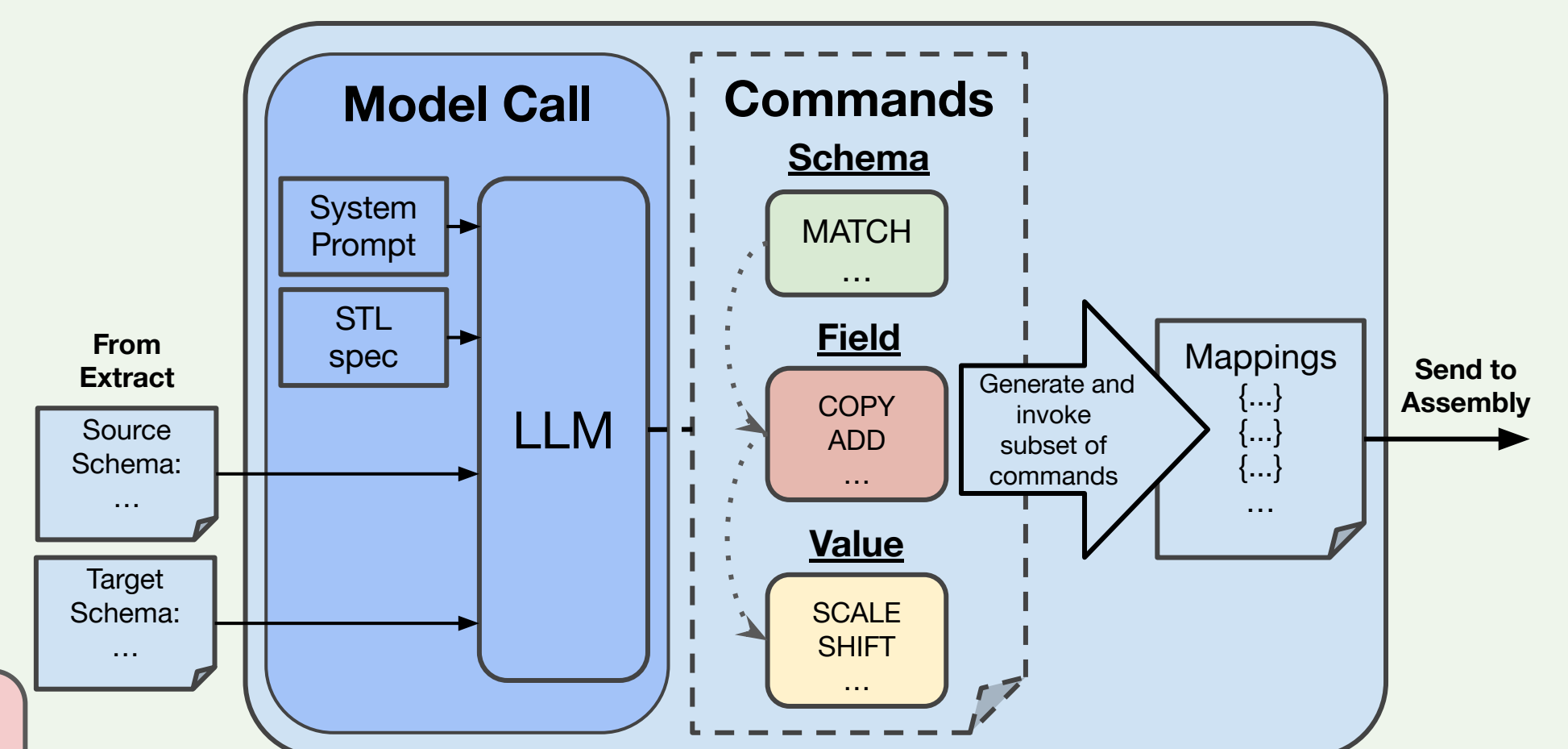
- Baseline: single model call with a high-level prompt
- ~20% to 70% higher mapping accuracy (20 runs, mapping granularity)

1. Schema Extraction

- Extract schema definitions from the data producer and data consumer and represent them in the STL.schema format.

2. Schema Mapping

- Perform schema matching, field transformation, and value transformation consecutively over extracted schema definitions.



3. Dataflow Assembly

- Compile schema mappings / STL commands to dataflow operators
- Update and patch the consumer's pipeline with the dataflow