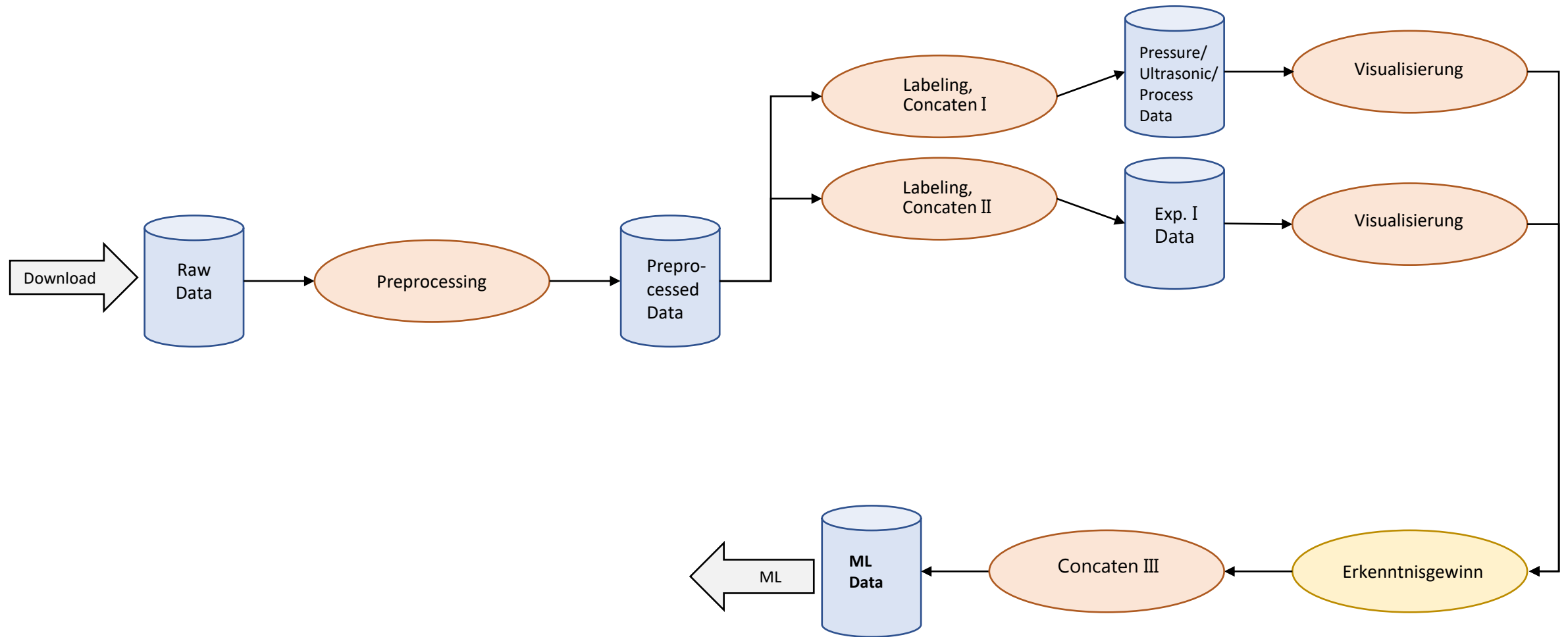
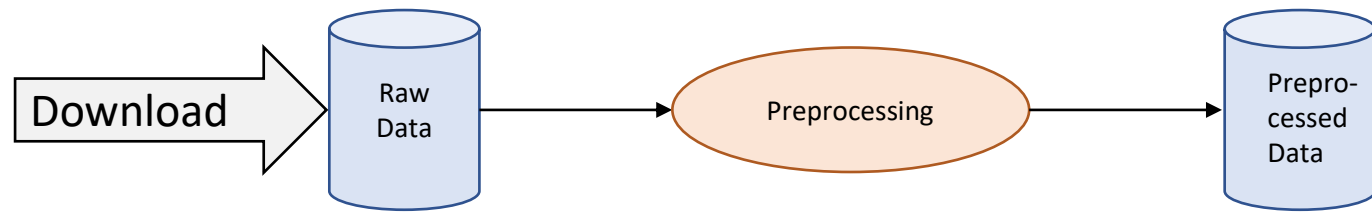


# SimOpt Data-Pipeline

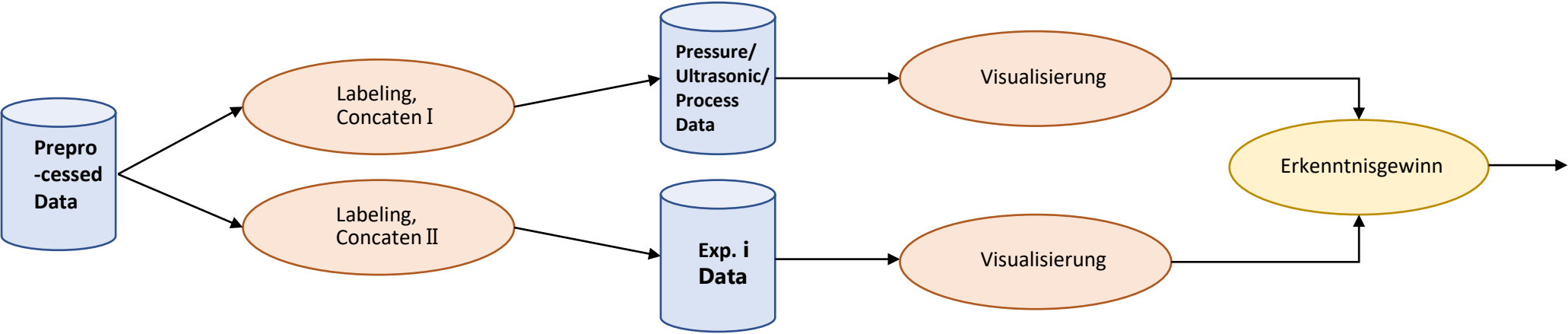




Raw Data	
Ordner Test9, Test10 aus gesamten Datensatz	
Pressure	<ul style="list-style-type: none"> <li>Unkalibriert</li> <li>unwichtige Spalten</li> <li>Kontextinfos</li> <li>➤ 18 Datensätze</li> <li>➤ a 300.000+ Zeilen</li> </ul>
Ultrasonic	<ul style="list-style-type: none"> <li>unwichtige Spalten</li> <li>➤ 18 Datensätze</li> <li>➤ a 600.000 Zeilen</li> </ul>
Process	<ul style="list-style-type: none"> <li>viele unwichtige Messreihen &amp; Spalten</li> <li>➤ 2 (gleiche?) Datensätze</li> </ul>

Preprocessing	
Aus Raw Data	
Pressure	<ul style="list-style-type: none"> <li>Auswahl relevanter Spalten</li> <li>Kalibrieren der Sensordaten</li> <li>Abspeichern Zeitstempel und Air- und Waterflow in sep. File</li> </ul>
Ultrasonic	<ul style="list-style-type: none"> <li>Dropfen der 3. Spalte</li> <li>Kürzen auf 600.000 Reihen</li> </ul>
Process	<ul style="list-style-type: none"> <li>Auswahl relevanter Spalten</li> <li>Auswahl relevanter Zeitreihen über Zeitstempel abgleich (mit Operational log, Anpassung Press. Data Zeitstempel nötig)</li> <li>Mergen des Wasser- und Luftzuflusses zu einer Spalte (nach Kriterium aus Anlagenbeschreibung)</li> </ul>

Preprocessed Data	
Nach Preprocessing	
Pressure	<ul style="list-style-type: none"> <li>➤ 18 Datensätze</li> <li>➤ 300.000x9 (außer 100_1)</li> <li>▪ Spalten: [B05,B8-14,B20]</li> </ul>
Ultrasonic	<ul style="list-style-type: none"> <li>➤ 18 Datensätze</li> <li>➤ 600.000x2</li> <li>▪ Spalten: [Timestamp, Voltage]</li> </ul>
Process	<ul style="list-style-type: none"> <li>➤ 18 Datensätze</li> <li>➤ 60x9</li> <li>➤ Spalten: siehe Table 2 aus Anlagenbeschreibung</li> </ul>



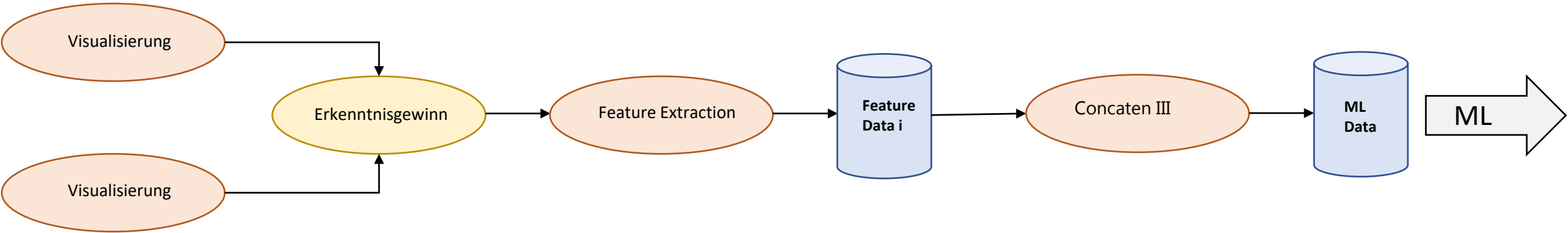
Preprocessed Data	
Nach Preprocessing	
Pressure	<ul style="list-style-type: none"><li>➤ 18 Datensätze</li><li>➤ 300.000x9 (außer 100_1)</li><li>▪ Spalten: [B05,B8-14,B20]</li></ul>
Ultrasonic	<ul style="list-style-type: none"><li>➤ 18 Datensätze</li><li>➤ 600.000x2</li><li>▪ Spalten: [Timestamp, Voltage]</li></ul>
Process	<ul style="list-style-type: none"><li>➤ 18 Datensätze</li><li>➤ 60x9</li><li>➤ Spalten: siehe Table 2 aus Anlagenbeschreibung</li></ul>

Labeling, Concaten I	
Aus Preprocessed Data	
Pressure	<ul style="list-style-type: none"><li>• Labeln der Datensätze nach Tabelle 5 aus Anlagenbeschreibung</li><li>• Zusammenhängen aller Druckmessungen zu einem Datensatz</li></ul>
Ultrasonic	<ul style="list-style-type: none"><li>• Labeln der Datensätze nach Tabelle 5 aus Anlagenbeschreibung</li><li>• Zusammenhängen aller Ultraschallmessungen zu einem Datensatz</li></ul>
Process	<ul style="list-style-type: none"><li>• Labeln der Datensätze nach Tabelle 5 aus Anlagenbeschreibung</li><li>• Zusammenhängen aller Prozessmessungen zu einem Datensatz</li></ul>

Labeling, Concaten II	
Aus Preprocessed Data	
Pressure	<ul style="list-style-type: none"><li>• Labeln der Datensätze nach Tabelle 5 aus Anlagenbeschreibung</li><li>• Zusammenhängen mit korrespondierender Ultraschallmessung und Prozessvariablen</li></ul>
Ultrasonic	<ul style="list-style-type: none"><li>• S.o.</li></ul>
Process	<ul style="list-style-type: none"><li>• S.o.</li></ul>

Pressure/Ultrasonic/Process Data	
Nach Labling, Concaten I	
Pressure	<ul style="list-style-type: none"><li>➤ 1 Datensatz</li><li>➤ ~5.400.000x9</li><li>▪ Spalten: [B05,B8-14,B20]</li></ul>
Ultrasonic	<ul style="list-style-type: none"><li>➤ 1 Datensätze</li><li>➤ 10.800.000x2</li><li>▪ Spalten: [Timestamp, Voltage]</li></ul>
Process	<ul style="list-style-type: none"><li>➤ 1 Datensätze</li><li>➤ 1.080x9</li><li>➤ Spalten: siehe Table 2 aus Anlagenbeschreibung</li></ul>

Exp. i Data	
Nach Labling, Concaten II	
Pressure	<ul style="list-style-type: none"><li>➤ In Exp-Datensätzen aufgeteilt</li></ul>
Ultrasonic	<ul style="list-style-type: none"><li>➤ In Exp-Datensätzen aufgeteilt</li></ul>
Process	<ul style="list-style-type: none"><li>➤ In Exp-Datensätzen aufgeteilt</li></ul>
Exp. i	<ul style="list-style-type: none"><li>• Druck, UT und Prozessdaten für jedes Experiment in einem Datensatz zusammengefasst</li><li>➤ 18 Datensätze</li><li>➤ 600.000x20(?)</li><li>– Größe abhängig vom Mergingverfahren</li></ul>



Feature Extraction		Feature Data i		Feature Extraction		ML Data	
Wahrscheinlich aus <b>Exp. I Data</b>		Nach Feature Extraction		Wahrscheinlich aus <b>Feature Data i</b>		Nach Concaten III	
Exp. i	<ul style="list-style-type: none"><li>Bearbeiten des Datensatzes in Betracht Länge und features(Spalten)</li></ul>	Exp. i	<ul style="list-style-type: none"><li>➤ 18 Datensätze</li><li>➤ Maße abhängig von features</li></ul>	Feature Data i (bzw. Whole Data)	<ul style="list-style-type: none"><li>Zusammenfügen der Datensätze zu einem großen finalen Datensatz</li></ul>	ML Data	<ul style="list-style-type: none"><li>➤ 1 Datensatz</li><li>➤ Gelabelt und fertig verarbeitet</li><li>➤ Bereit für ML-Algorithmen</li></ul>
		Whole Data	<ul style="list-style-type: none"><li>Feature Extraction prinzipiell auch aus Preprocessed Data möglich</li></ul>				