Computing Model

>10 PB of raw data per year
→ Distributed computing model
Share per fraction of PhDs

More than 10 PB of raw data per year

Regional Data Center

KEK Data Center (100%)

India Data Center (10%)

Korea Data Center (10%)

North America
PNNL Data Center (30%)

Canada Data Center (10%)

Germany Data Center (20%)

Italy Data Center (20%)

Europe

MC production site

Physics analysis skim

GRID site

Cloud site

Computer cluster site

Local resource
MC Campaigns

Computing system based on DIRAC
→ Many possibilities to contribute
Meeting with Takanori Hara tomorrow

Normalized CPU usage by Site
96 Weeks from Week 08 of 2015 to Week 52 of 2016

- Estimate: 138 kHS06
- PRE-MC5
- Now Almost Automated
- MC5 10 ab^-1
- Will Be Semi-automated
- MC5 Data Validation
- Grid Servers Upgrade
- MC6 1 ab^-1
- KEKCC Replacement
- MC7 2 ab^-1
- Coll. Tools @DESY
- BPAC Review
- Data Production Group
Migration svn → git

- Belle II decided last year to migrate collaborative services from KEK to DESY
  - We used that opportunity to switch from svn to git
    - Adjustment of procedures and tools required

Commits over Time

- Martin Ritter
- Thomas Kuhr
- Christian Pulvermacher

Journée Belle II @ LAL 2017-03-22
### Software Framework

- Dynamic loading of modules
- Data exchange via DataStore
- Relations
- Conditions data interface
- Root I/O
- Belle data input (b2bii)

- Parallel processing on event level
- Steering via python → meta-frameworks
Conditions Database

- Conditions data stored in objects in root files (payloads) → Provided via CVMFS or downloaded from server
- Global tag: Assignments of intervals of validity (IoV) to payloads → Database
- Distribution, redundancy
- Online integration
Tracking

- Cluster information in track finding
- GenFit for track fitting, ACTS as alternative?
inputMdst(...) # create "mu+:loose" ParticleList (and c.c.) stdLooseMu() # create Ks -> pi+ pi- list from V0 # keep only candidates with 0.4 < M(pipi) < 0.6 GeV fillParticleList('K_S0:pipi', '0.4 < M < 0.6')

# reconstruct J/psi -> mu+ mu- decay # keep only candidates with 3.0 < M(mumu) < 3.2 GeV reconstructDecay('J/psi:mumu -> mu+:loose mu-:loose', '3.0 < M < 3.2')

# reconstruct B0 -> J/psi Ks decay # keep only candidates with 5.2 < M(J/PsiKs) < 5.4 GeV reconstructDecay('B0:jspiks -> J/psi:mumu K_S0:pipi', '5.2 < M < 5.4')

# perform B0 kinematic vertex fit using only the mu+ mu- # keep candidates only passing C.L. value of the fit > 0.0 (no cut) vertexRave('B0:jspiks', 0.0, 'B0 -> [J/psi -> ^mu+ ^mu-] K_S0')

# build the rest of the event associated to the B0 buildRestOfEvent('B0:jspiks')

# perform MC matching (MC truth asociation) matchMCTruth('B0:jspiks')

# calculate the Tag Vertex and Delta t (in ps) # breco: type of MC association. TagV('B0:jspiks', 'breco')

# create and fill flat Ntuple with MCTruth, kinematic information and D0 FlightInfo toolsDST = ['EventMetaData', '^B0'] toolsDST += ['MCTruth', '^B0 -> [^J/psi -> ^mu+ ^mu-] [^K_S0 -> ^pi+ ^pi-]'] toolsDST += ['Vertex', '^B0 -> [^J/psi -> mu+ mu-] [^K_S0 -> pi+ pi-]'] toolsDST += ['DeltaT', '^B0'] toolsDST += ['MCDeltaT', '^B0']

# write out the flat ntuples ntupleFile('B2A410-TagVertex.root') ntupleTree('B0tree', 'B0:jspiks', toolsDST)
Event Display

➢ Virtual reality: https://vimeo.com/185549878
Roadmap

<table>
<thead>
<tr>
<th>Version</th>
<th>Status</th>
<th>Progress</th>
<th>Start date</th>
<th>Release date</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>release-00-09-00</td>
<td>UNRELEASED</td>
<td></td>
<td>01/Dec/16</td>
<td>01/May/17</td>
<td>Ready for Cosmic Data: raw data unpackers, cosmos simulation and reconstruction</td>
</tr>
<tr>
<td>release-01-00-00</td>
<td>UNRELEASED</td>
<td></td>
<td></td>
<td>01/Nov/17</td>
<td>Ready for BEAST Phase 2: alignment and calibration of outer detectors, generators for Upsilon(nS) verified, all software components ready for physics analysis</td>
</tr>
<tr>
<td>release-01-01-00</td>
<td>UNRELEASED</td>
<td></td>
<td></td>
<td>01/May/18</td>
<td>Ready for VXD cosmos</td>
</tr>
<tr>
<td>release-02-00-00</td>
<td>UNRELEASED</td>
<td></td>
<td>01/Oct/18</td>
<td></td>
<td>Ready for phase 3 data taking: full data processing chain working, procedures for systematic error determinations defined</td>
</tr>
<tr>
<td>release-02-01-00</td>
<td>UNRELEASED</td>
<td></td>
<td></td>
<td>01/May/19</td>
<td>Tuning to real data</td>
</tr>
<tr>
<td>release-03-00-00</td>
<td>UNRELEASED</td>
<td></td>
<td></td>
<td>01/Jan/21</td>
<td>High luminosity phase: optimized simulation, reconstruction, and alignment/calibration procedures</td>
</tr>
</tbody>
</table>

* Many exciting opportunities to contribute now!
## Institutional Commitments

<table>
<thead>
<tr>
<th>Institution</th>
<th>Commitment</th>
<th>Last Update</th>
<th>Timescale</th>
</tr>
</thead>
<tbody>
<tr>
<td>PSA</td>
<td>Development of the VXD Track Finder 2. In particular:</td>
<td>2017-02</td>
<td></td>
</tr>
<tr>
<td></td>
<td>- Development of the sector map, segment network and cellular automation parts</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>- Tuning, deployment and maintenance</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>- Monitoring of the VXD TF2 performances during the first years of the data taking</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>- Development of the algorithms to add PXD and SVD points to CDC-only tracks</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CMU</td>
<td>• CDC dE/dx reconstruction and calibration</td>
<td>2016-11</td>
<td>data production tasks associated with coordinator role</td>
</tr>
<tr>
<td></td>
<td>• CDC dE/dx track-level simulation</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>• charm related generators</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>• reconstruction package librarian</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>• data production: communication with physics, computing, detector, and daq/trigger groups</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>• data production: generation of release validation samples</td>
<td></td>
<td></td>
</tr>
<tr>
<td>KEK</td>
<td>• Background simulation (interface to machine)</td>
<td>2016-10</td>
<td></td>
</tr>
<tr>
<td></td>
<td>• Software related to the infrastructure of detector (TRG, DAQ)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ljubljana</td>
<td>• ARICH reconstruction software and data base</td>
<td>2016-10</td>
<td></td>
</tr>
<tr>
<td></td>
<td>• TOP reconstruction software</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>• database</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>• background mixing</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>• analysis package librarian</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>• Belle to Belle II Conversion task force</td>
<td></td>
<td></td>
</tr>
<tr>
<td>MPP Munich</td>
<td>• vertexing</td>
<td>2016-10</td>
<td></td>
</tr>
<tr>
<td></td>
<td>• flavor tagging</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Jülich</td>
<td>• genfit2</td>
<td>2016-10</td>
<td>3 years</td>
</tr>
<tr>
<td>Virginia Tech</td>
<td>• bkkm package</td>
<td>2016-10</td>
<td></td>
</tr>
<tr>
<td></td>
<td>• ext</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>• mud</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>• KLM trigger simulation</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>• KLM-related parts of b2bii</td>
<td></td>
<td></td>
</tr>
<tr>
<td>British Columbia</td>
<td>• ECL clustering</td>
<td>2016-09</td>
<td>2 years</td>
</tr>
<tr>
<td>LMU</td>
<td>• software tools</td>
<td>2016-09</td>
<td>2 years</td>
</tr>
<tr>
<td></td>
<td>• external software</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>• build system</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>• mdst package</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>• geometry package</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>• database interface in basf2</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>• index files</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>• software installations on cvmfs</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>• buildbot infrastructure, including automatic (monthly) builds, checks, tests, documentation generation</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>• commit hooks</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>• shift tools</td>
<td></td>
<td></td>
</tr>
<tr>
<td>PNNL</td>
<td>• database back-end</td>
<td>2017-02</td>
<td></td>
</tr>
<tr>
<td></td>
<td>• TOP analysis software</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Workshop @ Elba May 15-19 2017

- https://agenda.infn.it/conferenceDisplay.py?confld=1262
Backup
Software Development at Belle II

Aim:
- Reliable, sophisticated, and easy-to-use software for acquisition, simulation, reconstruction, and analysis of Belle II data

Challenge:
- Regional distribution, different (cultural) backgrounds and skills of developers
- ✔ State-of-the-art tools
- ✔ Commonly accepted rules and guidelines
- ✔ Well defined procedures
- ✔ Efficient communication channels
Software Group Structure

Belle II Organization

- Outreach Committee
  Chair: Toru Iijima
- Speakers Committee
  Chair: Ida Peruzzi
- Diversity Committee
  Chair: Kay Kinouchi

Management
  Spokesperson: Thomas Browder
  Project Manager: Yutaka Ushiroda
  Financial Officer: Yoshihide Sakai

Institutional Board
  Chair: Christopher Haerty
  Institutional representative

Executive Board
  Chair: Francesco Forti

Financial Board
  Chair: Yoshihide Sakai
  National representative

- Software Coordinator: Thomas Kuhr

- Generators
  T. Ferber

- Simulation
  D. Kim

- Background
  M. Staric

- Tracking
  M. Heck, E. Paoloni

- Alignment and calibration
  T. Bilka, C. Kleinwort

- Database
  M. Bracko, L. Wood

- Clustering and neutrals
  T. Ferber

- Software package librarians
- Detector contacts
Software Quality Control

Automated checks:

- code style
- gcc/clang/icc
- cppcheck, clang static analyzer
- unit/execution tests
- Doxygen
- geometry overlaps
- valgrind memcheck
- execution time and output size monitoring
- high level validation plots using simulated samples
Software Quality Shifts

- Monitoring of quality indicators
- Interaction with developers and librarians
- Detailed instructions on confluence
- One week duty
- Reports in weekly developers meeting

<table>
<thead>
<tr>
<th>Package</th>
<th>Librarian</th>
<th>Build Result</th>
<th>Intel Build Result</th>
<th>Clang Build Result</th>
<th>Cppcheck</th>
<th>Test Result</th>
<th>Geometry</th>
<th>Code Documentation</th>
<th>Dependencies</th>
</tr>
</thead>
<tbody>
<tr>
<td>alignmtn</td>
<td>Tadeas Bilka</td>
<td>OK</td>
<td>OK</td>
<td>OK</td>
<td></td>
<td>0/6, 0/0</td>
<td>OK</td>
<td>Missing: 1</td>
<td>Extra: 2, Missing: 2</td>
</tr>
<tr>
<td>analysis</td>
<td>Anze Zupanc</td>
<td>OK</td>
<td>OK</td>
<td>OK</td>
<td>Remarks: 1</td>
<td>0/69, 0/11</td>
<td>OK</td>
<td>Missing: 1084</td>
<td>OK</td>
</tr>
<tr>
<td>arch</td>
<td>Luka Santeij</td>
<td>Warnings: 1</td>
<td>OK</td>
<td>OK</td>
<td>Remarks: 1</td>
<td>0/0, 0/1</td>
<td>Overlaps:</td>
<td>Missing: 55</td>
<td>OK</td>
</tr>
<tr>
<td>b2bill</td>
<td>Anze Zupanc</td>
<td>OK</td>
<td>OK</td>
<td>OK</td>
<td>OK</td>
<td>OK</td>
<td>OK</td>
<td>Missing: 306</td>
<td>OK</td>
</tr>
<tr>
<td>background</td>
<td>Marko Staric</td>
<td>OK</td>
<td>OK</td>
<td>OK</td>
<td>None</td>
<td>OK</td>
<td>OK</td>
<td>OK</td>
<td>OK</td>
</tr>
<tr>
<td>beast</td>
<td>Igel Jajgle</td>
<td>Warnings: 3</td>
<td>OK</td>
<td>Warnings: 7</td>
<td>Remarks: 9</td>
<td>None</td>
<td>OK</td>
<td>Missing: 404</td>
<td>OK</td>
</tr>
<tr>
<td>bklm</td>
<td>Leo Pilonen</td>
<td>OK</td>
<td>OK</td>
<td>OK</td>
<td>OK</td>
<td>0/0, 0/1</td>
<td>OK</td>
<td>OK</td>
<td>OK</td>
</tr>
<tr>
<td>calib</td>
<td>Tadeas Bilka</td>
<td>OK</td>
<td>OK</td>
<td>OK</td>
<td>OK</td>
<td>0/0, 0/3</td>
<td>OK</td>
<td>Missing: 12</td>
<td>OK</td>
</tr>
<tr>
<td>cdc</td>
<td>Nakano Eichi</td>
<td>OK</td>
<td>OK</td>
<td>OK</td>
<td>OK</td>
<td>0/0, 0/1</td>
<td>OK</td>
<td>OK</td>
<td>OK</td>
</tr>
</tbody>
</table>
Simulation

➢ Detector geometry implemented in Geant4
➢ Parameters obtained from xml file/database
➢ Energy deposits stored as SimHits
➢ Digitization in modules
➢ Background mixing
➢ Background overlay
➢ Fast simulation?
Charged Particle Identification

- Neyman Pearson lemma
- Likelihood for each detector: \( L(\text{detector response} | \text{part. type}) \)
- Combination: product of likelihoods
- Probability can be calculated with analysis dependent priors
Full Event Interpretation

- Huge number of $B$ meson decay modes
  - Hierarchical reconstruction
  - Multivariate classifiers

→ Tools for analysis specific training of classifiers